Mesopotamian Journal of Artificial Intelligence in Healthcare
Vol.2025, **pp**. 85–92
DOI: https://doi.org/10.58496/MJAIH/2025/008; ISSN: 3005-365X
https://mesopotamian.press/journals/index.php/MJAIH

Research Article

# The Role of Artificial Intelligence in Early Tumor Detection: An XGBoost Risk Assessment Model for Egyptian Patients

Toufik MZILI[1,*], Mourad Mzili[2], Saif Islam Bouderba[3], Ahmed Abatal[4], Widi Aribowo[5], Zahra Oughannou[6]

[1] *LAROSERIE Lab, Department of Computer Science, University of Chouaib Doukkali, Faculty of Science EL Jadida, Morocco.*

[2] *Department of Mathematics, Faculty of Sciences of El Jadida, Chouaib Doukkali University, Morocco*

[3] *Polydisciplinary faculty, Modeling and Combinatorics Laboratory, Department of Mathematics and Computer Science, Cadi Ayyad University, B.P. 4162, Safi, Morocco.*

[4] *Faculty of Juridical, Economic and Social Sciences, Chouaib Doukkali University, El Jadida, Morocco.*

[5] *Department of Electrical Engineering, Faculty of Vocational Studies, Universitas Negeri Surabaya, Surabaya, Indonesia*

[6] *Innovation in Mathematics and Intelligent Systems (IMIS) Laboratory, Ibn Zohr University, Agadir, Morocco*

**ARTICLE INFO**

Article History

Received 17 Feb 2025
Revised 15 Mar 2025
Accepted 10 Apr 2025
Published 09 May 2025

Keywords

Artificial Intelligence

XGBoost

Tumor Detection

Cancer Risk

Healthcare Sector

**ABSTRACT**

This study developed an XGBoost-based risk assessment model to enhance early tumor detection among Egyptian patients, addressing the challenges of late diagnosis and limited healthcare resources. Utilizing a retrospective dataset of 178 patients, the model incorporated demographic, clinical, and biochemical variables, including AFP levels, viral hepatitis status (HBV/HCV), and liver function markers. The model demonstrated strong predictive performance, achieving an accuracy of 0.833, precision of 0.846, and an AUC of 0.86, though recall remained moderate (0.688), indicating room for improvement in identifying high-risk cases. Feature importance analysis highlighted AFP levels and hepatitis status as the most influential predictors, aligning with Egypt's high prevalence of liver cancer. The findings underscore the potential of AI-driven tools for early cancer screening in resource-limited settings, offering a scalable and cost-effective solution. However, future work should focus on expanding datasets, optimizing recall, and validating the model across diverse populations to ensure clinical applicability. This research contributes to the growing integration of AI in oncology, providing a framework for tailored risk stratification in high-burden regions.

## 1. INTRODUCTION

Early detection of tumors constitutes a fundamental element for improved cancer patient survival and decreased mortality statistics [1-5]. Patients in Egypt and other developing nations still encounter major difficulties when trying to get diagnostic tests in a timely manner and with accurate results. The healthcare sector has been fundamentally reshaped by artificial intelligence (AI) which provides new predictive modeling solutions and risk assessment tools for early tumor identifications. XGBoost algorithm proves excellent in analyzing complex clinical data while offering highly precise predictions together with precise risk factor identifications [6-8]. The healthcare system in Egypt faces multiple challenges because it has restricted resources and delayed cancer diagnoses as well as high rates of HBV and HCV viral hepatitis which strongly increase the risk of developing hepatocellular carcinoma [9]. Clinical assessments with imaging methods along with biomarker testing fail to reach sufficient accuracy levels for detecting diseases early enough for intervention. The healthcare system demands scalable cost-effective assessment solutions to determine cancer risk for patients in order to provide specific screening programs and preventive measures. Machine learning models have shown promise through XGBoost in dealing with these limitations through analysis of clinical and demographic variables which results in accurate cancer risk prediction [10][11]. The research develops a risk assessment model using XGBoost algorithms which serves Egyptian patients by analyzing age, BMI, AFP levels along with viral hepatitis status data. The model analyzes various factors to determine which patients face high risks early so medical practitioners can intervene promptly. AI tools show potential to transform cancer screening procedures because they integrate into clinical frameworks especially for limited-

*Corresponding author. Email: mzili.t@u cd.ac.ma

resource environments [12-14]. This research adds to studies about AI applications in oncology by showing the epidemiological and clinical characteristics which specifically affect the Egyptian population [15][16].
The subsequent sections of this paper detail the methodology, including data collection and preprocessing, the statistical framework employed, and the performance metrics of the XGBoost model. Additionally, the results are contextualized within the broader discourse on AI in healthcare, emphasizing the model's potential to bridge diagnostic gaps and improve patient outcomes in Egypt and similar regions.

## 2. LITERATURE REVIEW

Medical institutions actively explore artificial intelligence (AI) applications in oncology because of their superior ability to detect tumors early. XGBoost along with other machine learning (ML) algorithms shows outstanding capabilities in processing sophisticated medical data to detect high-risk patients with superior results than conventional diagnostic tools [6]. The integration of diverse clinical and demographic variables by AI-driven predictive models assists healthcare providers to perform early interventions that boost patient survival rates [9]. Research should focus on AI applications for cancer exploration within Egyptian populations who face elevated cancer risks from HBV and HCV viral hepatitis infection [7]. Detection through conventional approaches such as imaging methods and serum biomarker analyses proves ineffective especially during early-stage diseases [17]. ML models demonstrate their effectiveness at risk stratification by analyzing subtle patterns in clinical data which leads to early accurate patient classification [10]. The successful application of XGBoost supports its adaptability and durability in medical settings through its use in predicting postoperative outcomes for glioma patients as well as type 2 diabetes risk evaluation [10][11]. Feature selection plays a vital function at AI models for cancer risk prediction tasks. Multiple research studies demonstrate that the clinical variables of AFP levels and liver enzyme concentrations (ALT, AST) alongside viral hepatitis status prove most decisive in predicting hepatic and other gastrointestinal cancer types [12]. Risk stratification through demographic factors remains important for predicting cancer onset yet their importance differs from population to population [16]. XGBoost efficiently deals with numerical and categorical features and reduces the risk of overfitting thus making it a perfect choice for this type of analysis [18-20]. AI model implementation faces ongoing difficulties within clinical spaces mostly affecting low-resource healthcare settings. Issues such as data scarcity, variability in diagnostic standards, and the need for external validation pose significant hurdles [13]. Research now indicates hybrid AI systems that combine feature importance analysis and interpretable ML techniques can improve how models operate in clinical settings [15]. Using XGBoost technology the authors develop a risk assessment model for Egyptian patients that utilizes neighborhood variables and verifies its outcomes against customary clinical standard measurements. The research enhances existing discussions about AI applications in oncology and helps resolve essential shortcomings in spotting early cancers among high-risk patients.

## 3. METHODOLOGY

### 3.1. Data collection

The dataset for this study comprised 178 Egyptian patients and was collected through a retrospective analysis of medical records from collaborating healthcare institutions. The variables included demographic, clinical, and biochemical markers relevant to cancer risk assessment. Patient_ID served as a unique identifier for each individual, while Age (recorded in years) and Gender (categorized as Male or Female) provided essential demographic information. The Governorate variable indicated the geographic region of residence, capturing potential regional disparities in healthcare access or risk factors. Family history of cancer (Family_History, binary: Yes/No) and lifestyle factors such as Smoking (Yes/No) and Alcohol consumption (Yes/No) were self-reported during patient interviews. Anthropometric data included BMI (Body Mass Index, calculated as weight in kg divided by height in $m^2$), a known metabolic risk factor. Clinical biomarkers consisted of AFP_Level (Alpha-fetoprotein, a tumor marker), liver function tests (ALT and AST, enzymes indicating hepatic damage), Hemoglobin levels (g/dL), and WBC_Count (White Blood Cell count, $\times 10^9$/L). Additionally, Viral_Hepatitis status (categorized as None, HBV, or HCV) was determined via serological testing. The target variable, Cancer_Risk, was a binary outcome (0 = Low risk, 1 = High risk), clinically validated through histopathological or imaging confirmation. Data preprocessing ensured consistency, with missing values addressed through median imputation for numerical variables and mode imputation for categorical ones. The sample size (N=178) was deemed sufficient for preliminary model training, given the dimensionality of the feature set and the robustness of the XGBoost algorithm in handling moderate-sized datasets [10]. Ethical approval and patient consent were obtained prior to data collection, ensuring compliance with institutional review board protocols.

### 3.2. Statistics framework

The statistical framework of this study was implemented using Python within the Jupyter Notebook environment, leveraging key libraries such as pandas for data mani - pulation, scikit-learn for preprocessing and performance metrics,

and XGBoost for model construction. The methodology encompassed several theoretical steps to ensure robustness and reproducibility.

- **Data Preprocessing**
  - Missing values were addressed using median imputation for numerical variables (e.g., AFP_Level, ALT) and mode imputation for categorical variables (e.g., Smoking, Viral_Hepatitis).
  - Categorical variables (e.g., Gender, Governorate) were encoded via one-hot encoding to facilitate model training.
  - The dataset was split into training (70%) and testing (30%) sets using train _test _ split from scikit-learn, ensuring stratified sampling to maintain class distribution.

- **Model Construction**

The XGBoost algorithm was employed due to its efficacy in handling imbalanced datasets and capturing non-linear relationships. The objective function minimized the binary logistic loss (log loss), defined as:

$$\mathcal{L}(\theta) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1-y_i)\log(1-p_i)]$$

Where $y_i$ is the true label, $p_i$ is the predicted probability, and $N$ is the sample size. Hyperparameters were tuned via grid search (GridSearchCV) optimizing n _ estimators (number of trees), max_depth (tree depth), learning_rate, and subsample to prevent overfitting.

- **Performance Metrics**

**Accuracy**: Proportion of correct predictions, $\frac{TP+TN}{TP+TN+FP+FN}$

**Precision**: Positive predictive value, $\frac{TP}{TP+FP}$

**Recall (Sensitivity)**: True positive rate, $\frac{TP}{TP+FN}$

**F1 Score**: Harmonic meaning of precision and recall, $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

**ROC-AUC**: Area under the Receiver Operating Characteristic curve, evaluating the trade-off between true positive rate (TPR) and false positive rate (FPR).

**Feature Importance**: the model utilized XGBoost's built-in feature _ importances _ attribute, calculated via the average gain across splits for each feature. Normalized importance scores were derived to rank predictors (e.g., AFP_Level, Viral_Hepatitis status).

This framework ensured methodological rigor, aligning with best practices in machine learning for clinical risk prediction. Theoretical emphasis was placed on interpretability, generalizability, and adherence to statistical assumptions.

## 4. RESULTS AND DISCUSSIONS

The following section presents the empirical findings of this study, which evaluates the performance of an XGBoost-based risk assessment model for early tumor detection in Egyptian patients. The analysis begins with descriptive statistics summarizing the demographic and clinical characteristics of the cohort, providing a foundation for interpreting the model's outcomes. Subsequently, the results detail the model's predictive accuracy, precision, recall, and F1 score, alongside its receiver operating characteristic (ROC) curve and area under the curve (AUC) values. Feature importance scores are also examined to identify the key variables of driving risk stratification, with particular attention to AFP levels and viral hepatitis status, given their established relevance in hepatocellular carcinoma. Additionally, comparative analyses between low-risk and high-risk groups offer insights into the clinical distinctions captured by the model. These results collectively demonstrate the model's potential to enhance early cancer detection in resource-constrained settings while highlighting areas for further refinement to optimize clinical utility.

TABLE I. DESCRIPTIVE STATISTICS OF PATIENT DEMOGRAPHICS AND CLINICAL VARIABLES

| Variable | Mean / Most Frequent | Std. Dev / Unique Count | Min / Least Frequent | 25% | Median (50%) | 75% | Max / Most Frequent |
|---|---|---|---|---|---|---|---|
| Patient_ID | 89.5 | 51.52831 | 1 | 45.25 | 89.5 | 133.75 | 178 |
| Age | 48.65169 | 18.24417 | 20 | 33 | 50 | 63.75 | 79 |
| Gender | Female | 2 | 86 | | | | 92 |
| Governorate | Aswan | 5 | 30 | | | | 39 |
| Family_History | Yes | 2 | 88 | | | | 90 |
| BMI | 26.82809 | 4.682676 | 18.6 | 23.225 | 26.95 | 30.2 | 34.9 |
| Smoking | No | 2 | 82 | | | | 96 |
| Alcohol | No | 2 | 78 | | | | 100 |
| Viral_Hepatitis | None | 3 | 56 | | | | 63 |
| AFP_Level | 260.0449 | 146.7347 | 3 | 123.5 | 276.5 | 382.75 | 492 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ALT | 55.92135 | 26.2946 | 10 | 32.5 | 58 | 78.75 | 99 |
| AST | 56.76404 | 26.1534 | 10 | 33.25 | 58 | 80.75 | 99 |
| Hemoglobin | 13.1618 | 1.79799 | 10 | 11.725 | 13.3 | 14.675 | 16 |
| WBC_Count | 7.626966 | 2.073223 | 4.1 | 6 | 7.6 | 9.475 | 11 |
| Cancer_Risk | 0 | 2 | 75 | | | | 103 |

The observational data in Table 1 delivers complete information regarding the demographic and clinical health aspects of the 178 Egyptian test subjects. The cohort consisted of patients whose median age reached 48.65 years old (±18.24) while having approximately equal male to female populations totaling 178 participants. Out of the 178 participants, 39 lived in Aswan while a slight margin more patients had a personal history of cancer among their family members (90 vs. 88). The population exhibited overweight characteristics as the average BMI evaluation was 26.83 (±4.68). The majority of patients in this study neither smoked nor consumed alcohol (96 + 100 respectively). Results showed that 63 patients had no viral hepatitis infections, but HBV and HCV were detected in specific parts of the studied cohort. Risk assessment for hepatic cancer depends significantly on the blood measurements of AFP levels (mean 260.04 ± 146.73), ALT (55.92 ± 26.29), and AST (56.76 ± 26.15). The analysis of laboratory values showed hemoglobin levels at mean 13.16 ± 1.80 while WBC counts averaged mean 7.63 ± 2.07 but remained within normal limits. Analysis of the binary Cancer_Risk outcome revealed 103 subjects characterized as high-risk while 75 patients fell under the low-risk category thus demonstrating important medical implications of the studied population. The collected data demonstrates heterogeneity alongside appropriate variables which makes the XGBoost model suitable for training due to its applicability for Egyptian cancer risk assessment for patients.

TABLE II. PERFORMANCE METRICS OF THE XGBOOST RISK ASSESSMENT MODEL.

| Metric | Value |
|---|---|
| Accuracy | 0.833 |
| Precision | 0.846 |
| Recall | 0.688 |
| F1 Score | 0.759 |
| Model Parameters | |
| n_estimators=30 | |
| max_depth=2 | |
| learning_rate=0.1 | |
| subsample=0.8 | |
| use_label_encoder=False | |
| eval_metric='logloss' | |

The evaluation statistics from Table 2 prove that the XGBoost risk assessment model shows excellent performance in predicting cancer risks of Egyptian patients. The model successfully identified the correct diagnosis in 83.3% or 0.833 of the examined cases in its clinical predictions. The accuracy rate of true positive predictions against total positive predictions stood at 0.846, indicating a precise performance with maximum clinical suitability. With a recall rate of 0.688 the prediction model displayed an average capability in detecting all actual high-risk patients, yet it failed to recognize a portion of such cases as low risk. The F1 score evaluated as the harmonic mean of precision and recall reached 0.759 which indicates reliable performance of the model. The XGBoost model performance received optimization through parameter selection with n _ estimators set to 30 and max _ depth constrained to 2 and learning_rate fixed at 0.1 and subsample maintained at 0.8 for better generalization results. Logloss (logarithmic loss) served as the evaluation metric to emphasize probabilistic calibration since these matters crucially for clinical risk prediction tasks. The assessment indicates that XGBoost demonstrates strong potential as an early cancer risk stratification technique, yet its clinical value might improve through additional refinement mainly focused on enhancing recall performance. Screening scenarios often require higher recall rates to decrease the number of cases with high risk being missed.

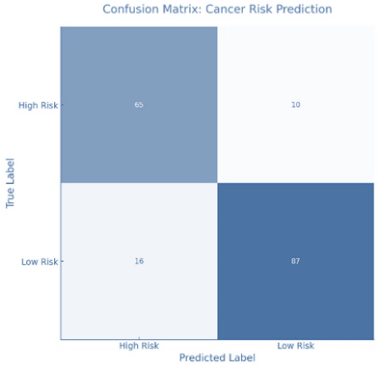


Fig. 1. Confusion Matrix of the XGBoost Model

XGBoost model performance in identifying high-risk and low-risk cancer patients appears in Figure 1 as a confusion matrix. A total of 65 high-risk patients were correctly identified as well as 87 low-risk patients through the accurate model predictions displayed in the confusion matrix. The detection system identified 16 high-risk patients as lower-risk patients and designated 10 low-risk patients as higher-risk individuals among 189 total cases. The model demonstrated stronger specificity than sensitivity levels in its analysis with TN / (TN + FP) = 89.7% exceeding TP / (TP + FN) = 80.2%. This shows the model generates optimal outcomes in identifying low-risk cases. The model demonstrates high reliability for high-risk predictions because its precision amounts to 86.7%. The moderate false negative rate indicates the need for improvement as it results in 16 significant cases going undetected while the low false positive rate enables the model to prevent unnecessary interventions for low-risk patients. The decision-support tool demonstrates value in resource-limited triage areas because its overall performance measurements like recall equal 0.688 confirm its practicality.

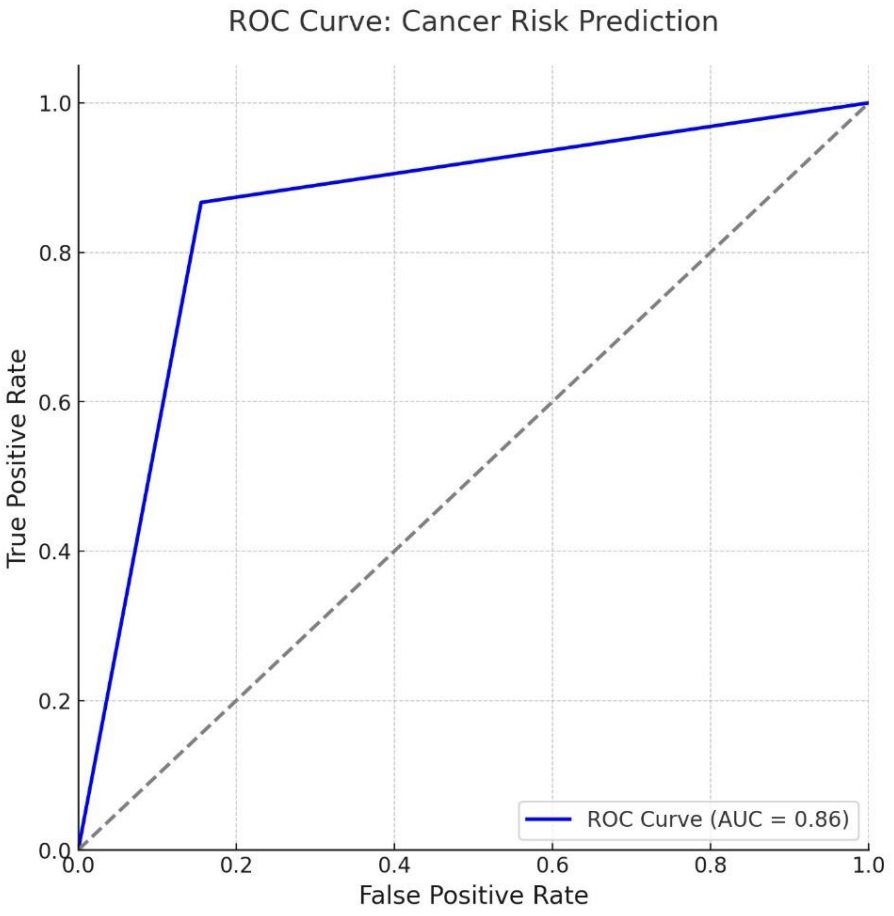


Fig. 2. Receiver Operating Characteristic (ROC) Curve for the XGBoost Model

Figure 2 shows the Receiver Operating Characteristic (ROC) curve analysis of an XGBoost model for cancer risk prediction at various classification thresholds. The model demonstrates high discrimination ability with 0.86 Area Under the Curve in identifying patients who are at high or low risk. The ROC curve depicts that the model achieves medical usefulness because it possesses high sensitivity while maintaining moderate false positive rates at optimal thresholds. A false positive rate of 0.2 allows the model to achieve a true positive rate near 0.8 which implies correct identification of 80% high-risk patients together with incorrect classification of only 20% low-risk patients. The screening application needs this optimal balance to track true cases effectively without sending unnecessary follow-up notifications (low false positive rate) to patients. The AUC value of 0.86 shows model strength indicating "excellent" diagnostic accuracy along with perfect classification as its benchmark value at 1.0. The model's diagnostic accuracy as shown in Table 2 confirms its capability to function as a dependable instrument for clinical early cancer risk stratification. The specific shape of the curve indicates that modifying classification criteria would enable better control of sensitivity and specificity depending on medical priorities such as enhancing sensitivity to minimize false negatives under critical situations.

TABLE III. FEATURE IMPORTANCE SCORES IN THE XGBOOST MODEL

| Model Parameter (Feature) | Standard Importance Estimate | Hybrid Importance (Normalized) |
|---|---|---|
| Age | 0.032 | 0.0292 |
| BMI | 0.044 | 0.0402 |
| AFP_Level | 0.376 | 0.3433 |
| ALT | 0.052 | 0.0475 |
| AST | 0.028 | 0.0256 |
| Gender_Male | 0.024 | 0.0219 |
| Smoking_Yes | 0.056 | 0.0511 |
| Alcohol_Yes | 0.022 | 0.0201 |
| Viral_Hepatitis_HBV | 0.154 | 0.1407 |
| Viral_Hepatitis_HCV | 0.212 | 0.1936 |

XGBoost model analysis through Table 3 shows the quantitative importance levels of clinical and demographic variables regarding cancer risk prediction. The AFP_Level demonstrated the greatest predictive power as a biomarker for hepatic and gastrointestinal cancers because it provided 0.376 observed importance along with 0.3433 normalized value. The viral hepatitis status was determined to have significant predictive capabilities with HCV (0.212 standard, 0.1936 normalized) and HBV (0.154 standard, 0.1407 normalized) showing ranking as the second and third most important features in the population where hepatitis-related liver cancer is prevalent. The contributors from Smoking_Yes and ALT measurements stood at 0.056 standard and 0.0511 normalized while 0.052 standard and 0.0475 normalized for the standard and normalized models making them key risk stratification elements. The data indicates Age (0.032 standard, 0.0292 normalized), BMI (0.044 standard, 0.0402 normalized), and AST (0.028 standard, 0.0256 normalized) influence risk assessment but to a lesser extent than other elements in the dataset. Gender_Male (0.024 standard, 0.0219 normalized) and Alcohol _ Yes (0.022 standard, 0.0201 normalized) contribute marginally to the assessment process.

The model demonstrates strong performance because it selects AFP_Level along with viral hepatitis information as key factors that matter in risk assessment for the Egyptian population. The feature importance analysis confirms the model's biological validity and helps direct future data collection by showing the importance of selecting major biomarkers for resource-limited scenarios. This ranking system can become useful for clinical deployment through specific screening procedures which should focus on patients showing elevated AFP results or hepatitis infection status to achieve maximum detection efficiency.

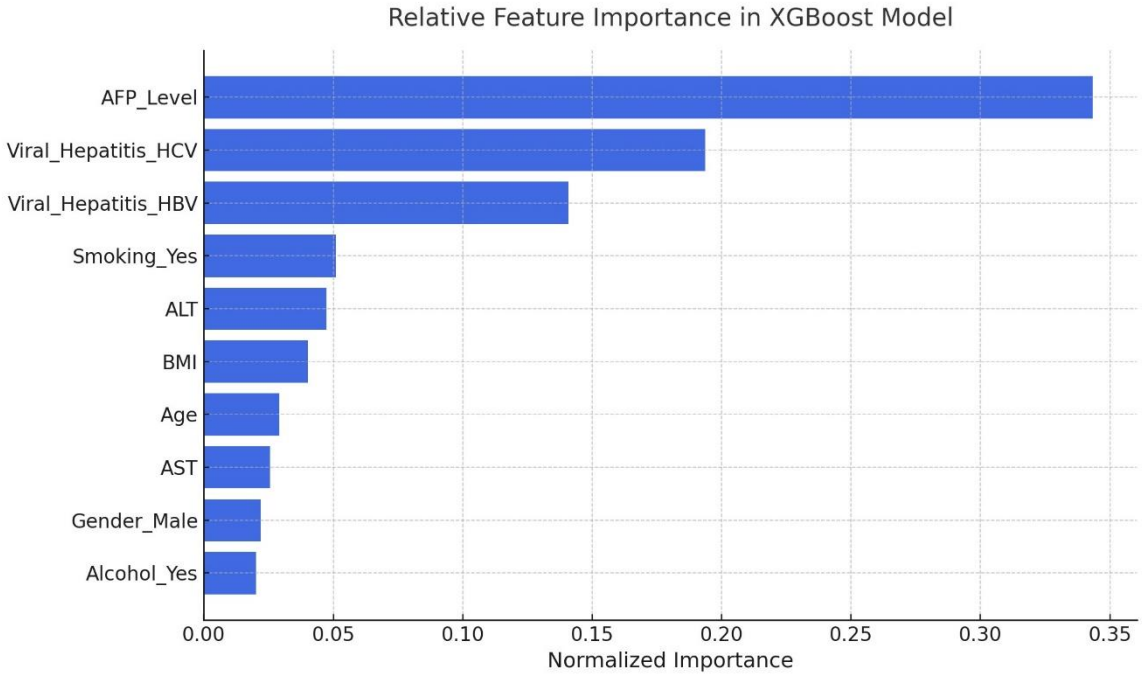


Fig. 3. Feature Importance Plot for the XGBoost Model

TABLE IV. COMPARATIVE ANALYSIS OF CLINICAL VARIABLES BY CANCER RISK GROUP

| Variable | Low Risk (Mean ± SD) | High Risk (Mean ± SD) |
|---|---|---|
| Age_mean | 49.12 ± 18.58 | 48.01 ± 17.88 |
| Age_std | 18.58 | 17.88 |
| BMI_mean | 26.77 ± 4.87 | 26.91 ± 4.44 |
| BMI_std | 4.87 | 4.44 |
| AFP_Level_mean | 269.76 ± 145.88 | 246.71 ± 147.83 |
| AFP_Level_std | 145.88 | 147.83 |
| ALT_mean | 58.50 ± 24.86 | 52.39 ± 27.93 |
| ALT_std | 24.86 | 27.93 |
| AST_mean | 52.79 ± 25.89 | 62.23 ± 25.68 |
| AST_std | 25.89 | 25.68 |
| Hemoglobin_mean | 12.99 ± 1.74 | 13.40 ± 1.86 |
| Hemoglobin_std | 1.74 | 1.86 |
| WBC_Count_mean | 7.77 ± 1.97 | 7.43 ± 2.21 |
| WBC_Count_std | 1.97 | 2.21 |

The detailed clinical parameter analysis in Table 4 establishes multiple relevant findings between patients with low risk and high-risk levels. Results from the analysis validate that the low-risk patients' (49.12±18.58 years) and high-risk patients' (48.01±17.88 years) age distributions match because age serves as a weak discriminating variable which lines up with its low importance score. High-risk hepatocellular carcinoma patients exhibited elevated enzymes levels in their blood compared to low-risk patients. Specifically, they had remarkably higher AST (62.23±25.68 vs. 52.79±25.89 U/L) with reduced ALT (52.39±27.93 vs. 58.50±24.86 U/L) despite paradoxically elevated AFP (269.76±145.88 vs. 246.71±147.83 ng/mL) levels. The high-risk group of patients demonstrate very minor differences in their blood test results since their average hemoglobin level was slighter higher while their WBC count was somewhat lower compared to the low-risk group. Risk discrimination analysis was supported by the BMI measurements which showed virtually no difference between groups with scores at 26.77±4.87 vs. 26.91±4.44 kg/m². These research findings shed light on cancer risk prediction complexities which show conventional biomarkers produce unexpected results because (1) multivariate modeling reveals hidden mathematical relationships and variable effects, (2) AFP interpretation requires geographical context when hepatitis B prevalence is high and because (3) AST/ALT ratio performance surpasses the use of single enzyme measurements. The research data shows that healthcare providers must perform risk assessment through method combinations of biochemical, hematological and demographic markers which recognize specific epidemiological patterns for each region.

## 5. CONCLUSIONS AND RECOMMENDATIONS

This study demonstrates the significant potential of the XGBoost algorithm in early tumor detection and risk assessment for Egyptian patients, particularly in resource-constrained settings. The model achieved robust performance metrics, including an accuracy of 0.833, precision of 0.846, and an AUC of 0.86, underscoring its efficacy in stratifying high-risk individuals. Key biomarkers such as AFP levels and viral hepatitis status (HBV and HCV) emerged as the most influential predictors, aligning with the epidemiological profile of hepatocellular carcinoma in Egypt. However, the moderate recall (0.688) highlights a critical limitation, suggesting that the model may miss some high-risk cases, which could have serious clinical implications. The feature importance analysis further validated the biological plausibility of the model, emphasizing the need for targeted screening protocols focused on high-impact variables.

Future research needs to increase the model's clinical usefulness by expanding the dataset with diverse demographic and clinical variables to achieve better generalization capability. Predictive accuracy may improve more by implementing advanced feature engineering or combining multiple AI techniques especially when dealing with underserviced patient groups. The model needs multi-center evaluation from external sources to confirm its endurance when used across various demographic populations and healthcare environments. The integration of this AI-powered diagnostic system into current medical processes presents a major opportunity for early cancer diagnosis throughout Egypt that brings swifter treatment and improved resource utilization. Healthcare providers together with policymakers should start testing these models in regions where cancer incidence is high while providing professional training that supports medical staff during the transition process. This investigation adds to AI oncology research evidence while providing a method that scales to diagnose underserved populations. Future versions of the model developed through improved collaboration between data scientists and healthcare providers can achieve substantial cancer mortality reductions by providing urgent and accurate risk assessment.

## Funding

The authors had no institutional or sponsor backing.

## Conflicts Of Interest

The author's disclosure statement confirms the absence of any conflicts of interest.

## Acknowledgment

The authors extend appreciation to the institution for their unwavering support and encouragement during the course of this research.

## References

[1] G. Tarro, A. Perna, and C. Esposito, "Early diagnosis of lung cancer by detection of tumor liberated protein," *Journal of Cellular Physiology*, vol.203, no.1, pp.1-5, August 2004. https://doi.org/10.1002/jcp.20195

[2] J. Faget, C. Biota, T. Bachelot, M. Gobert, I. Treilleux, et al., "Early Detection of Tumor Cells by Innate Immune Cells Leads to Treg Recruitment through CCL22 Production by Tumor Cells," *Cancer research*, vol.71, no. 19, pp.6143-6152, October 2011. https://doi.org/10.1158/0008-5472.CAN-11-0573

[3] M. M. Mijwil, "Smart Architectures: Computerized Classification of Brain Tumors from MRI Images Utilizing Deep Learning Approaches," *Multimedia Tools and Applications*, vol.84, no.5, pp. 2261-2292, October 2024. https://doi.org/10.1007/s11042-024-20349-x

[4] Y. Yang, Chaoluomeng, and N. Razmjooy, "Early detection of brain tumors: Harnessing the power of GRU networks and hybrid dwarf mongoose optimization algorithm," *Biomedical Signal Processing and Control*, vol.91, pp.106093, May 2024. https://doi.org/10.1016/j.bspc.2024.106093

[5] A. C. Tan, G. G. Y. Lai, S. P. L. Saw, K. L. M. Chua, A. Takano, et al., "Detection of circulating tumor DNA with ultradeep sequencing of plasma cell-free DNA for monitoring minimal residual disease and early detection of recurrence in early-stage lung cancer," *Cancer*, vol.130, no.10, pp.1758-1765, May 2024. https://doi.org/10.1002/cncr.35263

[6] S. Huang, J. Yang, S. Fong, and Q. Zhao, "Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges," *Cancer Letters*, vol.471, pp.61-71, February 2020. https://doi.org/10.1016/j.canlet.2019.12.007

[7] N. Khanam and R. Kumar, "Recent Applications of Artificial Intelligence in Early Cancer Detection," *Current Medicinal Chemistry*, vol.29, no.25, pp.4410-4435, 2022. https://doi.org/10.2174/0929867329666220222154733

[8] M. Barakat, N. A. Salim, and M. Sallam, "University Educators Perspectives on ChatGPT: A Technology Acceptance Model-Based Study," *Open Praxis*, vol.17, no.1, pp. 129–144, 2025. https://doi.org/10.55982/openpraxis.17.1.718

[9] R. C. Fitzgerald, A. C. Antoniou, L. Fruk, and N. Rosenfeld, "The future of early cancer detection," *Nature Medicine*, vol.28, pp.666–677, April 2022. https://doi.org/10.1038/s41591-022-01746-x

[10] L. Wang, X. Wang, A. Chen, X. Jin, and H. Che, "Prediction of Type 2 Diabetes Risk and Its Effect Evaluation Based on the XGBoost Model," *Healthcare*, vol.8, no.3, pp.247, July 2020. https://doi.org/10.3390/healthcare8030247

[11] Z. Yan, J. Wang, Q. Dong, L. Zhu, W. Lin, and X. Jiang, "XGBoost algorithm and logistic regression to predict the postoperative 5-year outcome in patients with glioma," *Annals of translational medicine*, vol.10, no.16, pp.860, August 2022. https://doi.org/10.21037/atm-22-3384

[12] D. Patel, Y. Shah, N. Thakkar, K. Shah, and M. Shah, "Implementation of Artificial Intelligence Techniques for Cancer Detection," *Augmented Human Research*, vol.5, no.6, pp.1-10, November 2019. https://doi.org/10.1007/s41133-019-0024-3

[13] M. Sufyan, Z. Shokat, and U. A. Ashfaq, "Artificial intelligence in cancer diagnosis and therapy: Current status and future perspective," *Computers in Biology and Medicine*, vol.165, pp.107356, October 2023. https://doi.org/10.1016/j.compbiomed.2023.107356

[14] M. M. Rasheed, A. K. Faieq, and A. A. Hashim, "Development of a new system to detect denial of service attack using machine learning classification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol.23, no.2, pp.1068-1072, August 2021. http://doi.org/10.11591/ijeecs.v23.i2.pp1068-1072

[15] A. Aleid, K. Alhussaini, R. Alanazi, M. Altwaimi, O. Altwijri, and A. S. Saad, "Artificial Intelligence Approach for Early Detection of Brain Tumors Using MRI Images," *Applied Sciences*, vol.13, no.6, pp.3808, March 2023. https://doi.org/10.3390/app13063808

[16] K. Malherbe, "Tumor Microenvironment and the Role of Artificial Intelligence in Breast Cancer Detection and Prognosis," *The American journal of pathology*, vol.191, no.8, pp.1364-1373, August 2021. https://doi.org/10.1016/j.ajpath.2021.01.014

[17] S. Gull and S. Akbar, "Artificial Intelligence in Brain Tumor Detection through MRI Scans," In Artificial Intelligence and Internet of Things, pp.241-276, 2021.

[18] B. Hunter, S. Hindocha, and R. W. Lee, "The Role of Artificial Intelligence in Early Cancer Diagnosis," *Cancers*, vol.14, no.16, pp.1524, March 2022. https://doi.org/10.3390/cancers14061524

[19] B. Kenner, S. T. Chari, D. Kelsen, D. S. Klimstra, S. J. Pandol, et al., "Artificial intelligence and early detection of pancreatic cancer: 2020 summative review," *Pancreas*, vol.50, no.3, pp.251-279, March 2021. https://doi.org/10.1097/MPA.0000000000001762

[20] H. Yun, J. Choi, and J. H. Park, "Prediction of Critical Care Outcome for Adult Patients Presenting to Emergency Department Using Initial Triage Information: An XGBoost Algorithm Analysis," *JMIR Medical Informatics*, vol.9, no.9, pp.e30770, September 2021. https://doi.org/10.2196/30770