RESEARCH PAPER

From Prediction to Practice: A Machine Learning–Based Clinical Decision Support Tool for Bevacizumab Risk Stratification in Oncology

Dr. Latha Kiran Krishna Rajendran[1]

[1] General Practitioner, Elova Hospitals, 27, 5th Cross, Lalbagh Main Rd, Sudhama Nagar, Bengaluru, Karnataka - 560027.
Email: meetlathakiran@gmail.com

Received: 28th Feb, 2026 | **Revised:** 14th Mar, 2026 | **Accepted:** 4th Apr, 2026 | **Available Online:** 20th Apr, 2026

ABSTRACT

AI in cancer might enhance decision-making, especially for focused therapy risk management. Its goal was to create and verify a machine learning-based clinical decision support system (CDSS) that could anticipate Bevacizumab or biosimilar problems and turn the prediction model into a clinical tool. A prospective observational research examined 395 solid tumor patients treated with Bevacizumab or biosimilars. Medical records included demographics, medical history, tumor features, and laboratory results ahead of treatment. XGBoost, Random Forest, and logistic regression were trained with data splits of 70/30 and 80/20. These prediction models were compared by accuracy, AUC-ROC, sensitivity, specificity, F1-scores, and error rate. Using the best model, a logistic-based risk score was calculated and put into an interactive HTML form. The improved Random Forest model that was trained using the 80/20 split has the highest accuracy (70.63%), sensitivity (66.67%), specificity (73.85%), and AUC-ROC (0.55). Both the calibration and AUC-ROC value of 0.720 for the logistic risk score were excellent. It found that age > 65, anemia, increased urea, leukocytosis, tumor differentiation, and stage are significant predictors of problems. The final product provides physicians with a straightforward offline form for risk assessment and patient risk stratification. This research shows that real-world data may be used to construct AI-supported cancer forecasting tools. Without supplanting clinical judgment, the interactive form and logistic risk score may assist physicians in customizing targeted therapy selections.

Keywords: machine learning; artificial intelligence; clinical score; personalized treatment; targeted therapy; clinical tools; interactive forms; clinical decision support system.

How to cite this article: Rajendran LKK. From Prediction to Practice: A Machine Learning–Based Clinical Decision Support Tool for Bevacizumab Risk Stratification in Oncology. Int J Drug Deliv Technol. 2026;16(30s):414-429. DOI: 10.25258/ijddt.16.30s.40

Source of support: Nil.

Conflict of interest: The author declares no conflict of interest.

I. INTRODUCTION

In recent years, oncological therapy has quickly advanced, making targeted medicines accessible to many patients worldwide. As a result, oncology produces more real-time data. Artificial intelligence techniques improve data synthesis and enable creative clinical tool creation. Clinical decision support systems, which are tools, are meant to help doctors make better decisions faster and with more accuracy [1]. AI technologies are able to quickly evaluate a large amount of disconnected clinical, paraclinical, imaging, molecular, and genetic data and discover predicted patterns that are difficult to spot with conventional methods. These patterns can be used to create clinical ratings. AI-based interactive forms may assess patients' risk of angiogenesis inhibitor problems in real time. CDSSs may help patients discuss treatment options, risks, and advantages during consultations, facilitating shared medical decisions. They may use age, comorbidities, biomarkers, and functional scores to create a tailored risk assessment [2]. In difficult or marginal patients, incorporating these technologies into oncology treatment may maximize therapy risk–benefit ratios. AI-CDSSs may indicate a higher risk of severe toxicity from a particular medication in older patients with a lot of comorbidities, leading the medical team to better-tolerated options. Although these instruments may inform physicians of the need for more stringent paraclinical monitoring or dosage adjustments, they cannot replace professional judgment or therapeutic recommendations. AI-CDSSs can continuously update with fresh data from clinical observation in actual oncology practice, another benefit. Clinical ratings can be improved through feedback, which is crucial in the fight against cancer. AI models may be automatically calibrated using real-world data. In order to keep up with changing therapy and population results, scores and forms can be updated. However, interdisciplinary cooperation between oncologists, biostatisticians, AI specialists, and medical informaticians is needed to evaluate these technologies in clinical settings[3].

However, clinical AI integration must be done responsibly. AI models may underestimate risk or misunderstand clinical factors due to algorithmic bias, necessitating thorough validation and representative data. No matter how good AI-CDSSs are , they cannot replace extensive clinical examinations, direct patient

talks, and accurate medical context assessments. Therapy choices ought to be influenced by international guidelines that are based on clinical data. The CDSS interface must be user-friendly, compatible with EHRs, and safeguard patient data. Professionals also like algorithms that are transparent and explain their suggestions. Physicians have a higher level of trust in AI tools when their outcomes match those of randomized controlled trials, highlighting issues with transparency and model explainability. AI-CDSSs in cancer also rely on data quality[4].

Advanced AI and ML have produced high-performing clinical decision support systems. The majority of these models were created with IT professionals in mind, resulting in complicated black box systems that are difficult for doctors to use. Thus, oncology has yet to incorporate AI technology despite their promise. We provide a clinician-led proof of concept to fill this gap. Instead than designing the most accurate or complicated prediction model, the objective was to show that modern AI technologies are user-friendly enough for doctors without strong computational abilities to employ. We want to connect computer science and medicine, demonstrate viability, and lessen physicians' anxiety about AI. This study shows how AI tools can be used to develop clinical interactive and predictive tools, which may be a major step toward truly personalized oncological care. It also shows that these tools are now so readily available and intuitive that clinicians without formal computational training can use them in research and clinical practice with promising results. We underline that model improvement, performance, interpretability, and clinical translation need multidisciplinary cooperation with AI specialists[5].

A summary of the significant contributions made by this study:

- Utilizing standard pretherapeutic data, we developed a proof-of-concept CDSS led by clinicians to anticipate issues with Bevacizumab. We choose the best machine learning algorithm (logistic regression, Random Forest, XGBoost) after testing.
- For clinical use, a simplified risk score based on logistic regression was developed and made available as an offline HTML calculator. We proved feasibility by using user-friendly AI tools on a real-world oncology dataset without IT skills.
- This effort aims to reduce physicians' concern about AI and promote future interdisciplinary cooperation, rather than as a final model.

II. MATERIALS AND METHODS

A. Patient Selection and Data Collection

According to the Declaration of Helsinki, the Bucharest Institute of Oncology "Prof. Dr. Alexandru Trestioreanu" Institutional Ethics Committee approved this research (Protocol Number. 24855, Date of approval: 24 November 2022). This open, observational, prospective research was undertaken in actual oncological conditions on patients treated with Bevacizumab or biosimilars, anticancer medications with common and occasionally major adverse effects. We included all solid malignant tumor patients treated at the Bucharest Oncological Institute "Prof. Dr. Al. Trestioreanu (IOB) between January 1, 2023, and June 30, 2024[6]. This study examined clinical practice and medication use, therefore data were gathered from patients' medical records and the hospital's computerized database. We did not set exclusion criteria, and the attending doctors determined all treatment, dosage, regimen, and follow-up choices without restriction. We followed patients until 31 December 2024, when they were censored before analysis. Our minimum period of follow-up was six months. Unless they withdraw their consent, all IOB patients agree, by signing a consent form, to the anonymous use of their medical records and investigation results for scientific, educational, and paper preparation and publication[7]. Utilizing pharmacy usage data, recipients of bevacizumab and biosimilars were identified. Beva cizumab was used in many lines of therapy, therefore each record was evaluated separately.

There was a distinct record for each indication for Bevacizumab for various neoplastic disorders. Case-specific information was provided by histopathology, genetic findings, laboratory, imaging, and electronic patient records. Demographics, comorbidities, relevant medical history, neoplastic disease history (tu mor origin, histology, immunohistochemistry, genetic characteristics, metastatic sites, diagnostic dates, treatment initiation, progression, death), and treatment aspects were collected.

Complications were any adverse event reported in patient files and clinical judgment that was likely connected to Bevacizumab or its biosimilars, regardless of severity. Due to the small dataset, we were unable to stratify the issues we wanted to anticipate as only serious. Thus, all complications were binary outcomes, reflecting real-world situations where even modest occurrences may necessitate dosage modifications or affect quality of life[8].

B. Variable Selection

Pretherapeutic biological parameters (hemoglobin, leukocytes, platelets, blood glucose, urea and creatinine, serum Ca2+, transaminases, GGT, presence of coagulation changes or tumor markers), the presence of personal history (smoking, alcohol, nutritional status), the presence of pathological history (ischemic heart disease, hypertension, heart failure, arrhythmias, valvular diseases, diabetes mellitus, previous treatment with oral anticoagulants Unfortunately, despite our efforts to collect data on tumor receptors, oncogenic mutations, and repair mechanisms, the massive percentage of missing data (over 80% of patient records do not contain such data) prevented us from using these data to generate predictive models.

C. Database Pre-Processing and Leakage Prevention

To achieve this goal, we processed the initial database by removing variables that described the type and consequences of complications, leaving only pretherapeutic patient and disease characteristics that could predict complications. Post-treatment and outcome-linked variables were eliminated to prevent data leakage. In the training data, only missing values were imputed, and before being applied to the validation fold, the fitted imputers were retrained within each cross-validation fold. No outcome-aware univariate filtering or feature selection occurred outside the cross-validation loop[10].

A recoded binary variable with values of 0 for no Bevacizumab-associated complications and 1 for at least one therapy-associated event of any severity was also added to the database. Due to the small number of cases, which may reduce their predictive value, we did not attempt to develop prediction models for serious complications. The database was optimized by meticulously cleaning it in various processes. At first, we discovered a number of variables with missing values. To avoid errors in the imputation process, variables encoding biological constants like Na+, K+, and Cl− (missing in over 20% of records before therapy) were eliminated from the analysis. We imputed continuous numerical variables like GGT and serum Ca2+ with less than 5% missing data by using the median value of the entire sample. As the final step in cleaning the dataset, an automated examination of the mean, minimum, and maximum values of continuous numeric variables revealed outliers. After rechecking the patient records, this revealed two data entry errors, one of which was missing a comma [11].

D. Training Predictive Models

On the optimized database, we trained numerous prediction machine learning models:

- Simple logistic regression, which uses linear connections between predictors and log-odds, is sensitive to outliers and noisy variables, but it yields coefficients that can be easily understood for therapeutic purposes.
- Elastic Net penalized logistic regression provides the benefits of L1 (sparse, variable selection) and L2 (stability) regularization.
- The RandomForest model, based on decision trees, is resilient and can capture complicated interactions.
- The XGBoost model gradually corrects errors, has high accuracy, performs better with complex data, and effectively controls imbalances, but it is harder to understand.

For Logistic/Elastic-Net, Random Forest, and XGBoost, brief, clinician-oriented "algorithm primers" outline their strengths and weaknesses. During training, we reduced mild class imbalance by employing balanced class weights for RF and XGBoost and inverse frequency class weights for Logistic/Elastic-Net. Metrics were constructed using the default probability threshold of 0.50, but clinical objectives may necessitate different thresholds that compromise sensitivity for specificity. For predictive model development and statistical analysis, Google Colab's Jupyter notebook was utilized. In order to generate Python scripts for the modeling process, web-based AI code generation assistants (Gemini 2.0 model and ChatGPT plus 4o) were used only under active human supervision and guided by the authors' requirements. These tools were used to create the figures. Analytical, methodological, interpreting, and scientific writing decisions were made solely by the authors[12]. In the beginning, datasets were divided 70 percent for training and 30 percent for testing in order to achieve a balance between learning and assessment for predictive models. We decided to try the 80/20 split version, which allows for training on more cases and may perform better than the 70/30 split, in order to improve the learning ability of the models, particularly in the context of class imbalance. To maximize performance, we optimized each model as follows:

- **Logistic regression:** To correct class imbalance, inverse weights were applied to the classes, and L1 (Lasso), L2 (Ridge), and Elastic-Net-type penalties were tested to reduce overfitting and select relevant variables. The stability of L2 is combined with the sparse regularization and variable selection of L1.
- **RandomForest:** Hyperparameters (the number of trees, maximum mumtree depth, and minimum number of data per leaf) and weights were altered to reduce minority class classification errors.
- **XGBoost:** Boost settings (learning_rate, max_depth, n_estimators, and subsample) and class weights were altered to improve complexity identification.

Grid search was used to adjust the hyperparameters of each algorithm on the training set. The final parameter settings can be found in Supplementary Table S1.

E. Comparing the Performance of Predictive Models

The same assessment approach was used to compare un optimized and optimized models on an unrevised test set (hold-out technique) of 20 or 30% of the original data, depending on the model split. Each model's performance was assessed using the following indicators:

- Inaccuracy overall.
- Class discrimination ability is measured by the AUCROC score.
- In medical settings, Class 1 sensitivity (recall) is essential for identifying patients with potential issues because missed instances may have significant clinical effects.
- Specificity in identifying instances without problems helps ensure safe medication delivery.

- The F1-score for class 1 shows the right balance between sensitivity and accuracy.
- Errorrate—total fraction of inaccurate classifications (false positives and negatives).

We chose the best model—optimized Random Forest 80/20—after the comparison study to develop clinical tools like a risk score and an interactive HTML form with an automated risk calculator.

F. Internal Validation

We next conducted a 5-fold stratified cross-validation process across the dataset to improve model generalizability estimates and decrease over fitting. To determine the AUC, sensitivity, specificity, and accuracy, we used the mean across folds. We conducted 25 cycles of repeated 5 × 5 CV to evaluate stability. Out-of-fold (OOF) predictions were made over all runs, and we calculated a bootstrap 95% CI for AUC (B = 2000). Due to the size of the dataset, the current proof of concept was unable to utilize a more stringent layered cross-validation technique[13].

III. RESULTS

A. General Description of the Cohorts and Distribution of the Target Variable

177 (44.8%) of the 395 single-location treatment sessions in the analytic dataset experienced side effects related to bevacizumab, while 218 (55.2%) did not. After ensuring no data leaking between train–validation partitions, patients with numerous treatment sessions were treated as separate observations. The primary result was any issue related to bevacizumab during the observation window[14]. Patient sample included 130 men and 265 women. Rural residents had less access to medical care, as evidenced by the fact that the majority of patients (341—86.33%) were from metropolitan areas. Through the National Oncology Program, 271 retired patients, 90 employees, and 34 uninsured patients received complex treatments at treatment (8.61 percent). Figure 1 shows the cohort's age histogram. The distribution is essentially normal, with a slight negative asymmetry (skewness = -0.18), as shown by the age histogram. Our cohort is typical for statistical analysis due to its rather normal distribution, and patient age does not influence the results[15]. The majority of patients treated with Bevacizumab for solid tumors were in their sixth decade, which is the typical age of diagnosis in oncology. Age ranged from 22 to 88 years, with a mean age of 61.6 10.88. Although a minority, young patients (under 40 years) show that advanced cases may occur in young populations. When devising a treatment plan, one must take into account their psychosocial characteristics. The finding that the two groups are not statistically different (p = 0.475) is significant since older age is associated with comorbidities. Bevacizumab side effects may be exacerbated by aging-related body weakness.

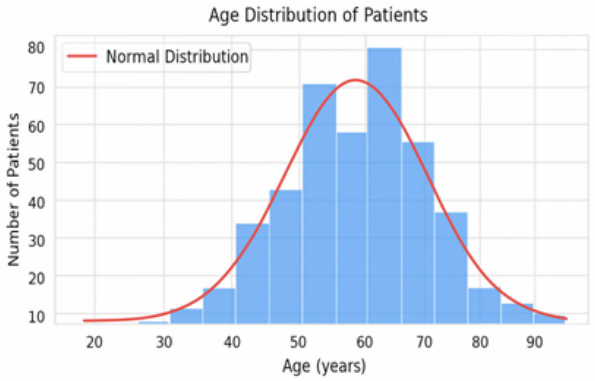

Figure 1. Age distribution of the study's participants.

Tables 1–3 show the distribution of pretherapeutic characteristics used to examine the worldwide sample. Pre-existing renal damage, lower serum hemoglobin values, and certain tumor characteristics (colorectal or cervical location, poor/absent cellular differentiation, and aggressive histologies) are the only variables that appear to be more common in the group with complications [16]. Numerous pretherapeutic characteristics are evenly distributed across the two groups.

Table1. Pre-treatment socio-demographic and medical variable distribution

Variable	Total (n, %)	Study Group (Complications)	Witness Group (No Complications)	p-value
Sex (Female)	265 (67.09 %)	130	135	0.475
Sex (Male)	152 (32.91 %)	88	64	—
Residence (Urban)	341 (86.33 %)	183	158	—
Residence (Rural)	54 (13.67 %)	35	19	—
Blood Group (AB)	34 (8.61 %)	18	16	0.667
Rh Negative	74 (18.73 %)	43	31	—
Hypertension	147 (37.22 %)	73	74	0.0328

Cardiac Insufficiency	33 (8.35 %)	12	21	0.0367
Valvopathies	62 (15.70 %)	26	36	0.0318
Arrhythmias	32 (8.10 %)	15	17	0.4230
TIA/Stroke	3 (0.76 %)	1	2	0.8560
PTE/DVT	20 (5.06 %)	6	14	0.0362
Chronic Lung Disease	25 (6.33 %)	10	15	0.0629
Diabetes	44 (11.14 %)	18	26	0.4779
Chronic Renal Disease	10 (2.53 %)	1	9	0.0096
Chronic Hepatitis/Cirrhosis	13 (3.29 %)	6	7	0.7020
Prior Cancer Diagnosis	34 (8.61 %)	21	13	0.5313
Mental Disorders	15 (3.80 %)	6	9	0.239
Recurrent UTI	15 (3.80 %)	11	4	0.239
Past Surgical Procedures	288 (72.91 %)	158	130	0.9190
Neoadjuvant Immunotherapy	94 (23.80 %)	44	50	0.0796
Oral Antidiabetics	29 (7.34 %)	9	20	0.0116
Insulin	4 (1.01 %)	1	3	0.4746
Bevacizumab	55 (13.92 %)	28	27	0.5878
Neoadjuvant Chemotherapy	234 (59.24 %)	134	100	0.3698
Radiotherapy	92 (23.29 %)	47	45	0.4331
Nutritional Status (Normal)	257 (65.06 %)	148	109	0.8143
Obesity/Overweight	73 (18.48 %)	40	33	—
Underweight/Cachexia	65 (16.46 %)	35	30	—

Table2. dispersion of cancer-specific pre-treatment variables

Category	**Variable**	**Total (n, %)**	**Study Group (Complications)**	**Witness Group (No Complications)**	**p-value**
Cancer Type	Ovarian/Tubal/Peritoneal	102 (25.82%)	70 (32.1%)	32 (18.1%)	**0.0002**
	Other Cancers	209 (52.91%)	104 (47.7%)	105 (59.3%)	
	Cervical	40 (10.13%)	16 (7.3%)	24 (13.6%)	
	Breast	28 (7.09%)	22 (10.1%)	6 (3.4%)	
	Pulmonary	7 (1.77%)	4 (1.8%)	3 (1.7%)	
	Others	9 (2.28%)	2 (0.9%)	7 (4.0%)	
Pathology	Adenocarcinoma	206 (52.15%)	104 (47.7%)	102 (57.6%)	**0.0013**

	Cervical Cancer	93 (23.54%)	63 (28.9%)	30 (16.9%)	
	Squamous Cell Carcinoma	33 (8.35%)	13 (6.0%)	20 (11.3%)	
	Invasive Ductal/Lobular	25 (6.33%)	20 (9.2%)	5 (2.8%)	
	Mucinous Adenocarcinoma	23 (5.82%)	14 (7.9%)	9 (5.1%)	
	Others	15 (3.80%)	9 (4.1%)	6 (3.4%)	
Differentiation Grade	G3 (Poorly differentiated)	144 (36.46%)	92 (42.2%)	53 (29.9%)	**0.0065**
	G2 (Moderately differentiated)	227 (57.47%)	110 (50.5%)	117 (66.1%)	
	G1 (Well differentiated)	24 (6.07%)	16 (7.3%)	8 (4.5%)	
Tumor Features	Perineural Invasion	75 (18.99%)	40 (18.3%)	35 (19.8%)	0.7049
	Lymphovascular Invasion	117 (29.62%)	64 (29.4%)	53 (29.9%)	0.9877
	Peritoneal Infiltration	79 (20.00%)	36 (16.5%)	43 (24.3%)	0.1473
Cancer Stage	Stage IV	316 (80.00%)	168 (77.1%)	148 (83.6%)	**0.3306**
	Stage III	74 (18.73%)	46 (21.1%)	28 (15.8%)	
	Stage II	4 (1.01%)	3 (1.4%)	1 (0.5%)	
	Stage I	1 (0.26%)	1 (0.5%)	0 (0%)	
Metastasis	Peritoneal	142 (34.95%)	81 (37.2%)	61 (34.5%)	0.65333
	Bone	25 (6.33%)	14 (6.4%)	11 (6.2%)	—
	Hepatic	169 (42.78%)	89 (40.8%)	82 (46.3%)	0.238
	Lymphatic	46 (11.65%)	28 (12.8%)	18 (10.1%)	0.5052
	Brain	8 (2.03%)	4 (1.8%)	4 (2.3%)	—
	Others	15 (3.80%)	8 (3.7%)	7 (4.0%)	—
Disease Type	Metastatic	314 (79.49%)	170 (78.0%)	144 (81.4%)	**0.1061**
	Advanced/Unresectable	52 (13.16%)	35 (16.1%)	17 (9.6%)	
	Recurrent	29 (7.34%)	13 (6.0%)	16 (9.0%)	
Treatment	First-line	180 (45.57%)	95 (43.66%)	85 (48.0%)	**0.3935**
	Subsequent lines	146 (36.96%)	84 (38.5%)	62 (35.0%)	
	Later lines	54 (13.	28 (12.8%)	26 (14.7%)	

		67%)			
	Maintenance	15 (3.80%)	11 (5.0%)	4 (2.3%)	

Table3. A distribution of pretherapeutic paraclinical variables.

Category	Variable	Study Group (Complications)	Witness Group (No Complications)	p-value
Hematological Parameters	Hemoglobin (g/dL)	11.84 ± 1.74	11.10 ± 1.84*	**0.0009**
	Leucocytes (×10³/µL)	7.44 ± 2.70	7.07 ± 2.66	0.1378
	Thrombocytes (×10³/µL)	286.11 ± 108.42	277.90 ± 111.24	0.2500
Biochemical Parameters	Glycemia (mg/dL)	101.83 ± 26.25	103.79 ± 27.01	0.3049
	Creatinine (mg/dL)	0.77 ± 0.21	0.82 ± 0.25	0.0845
	Urea (mg/dL)	32.16 ± 13.29	34.55 ± 15.48	0.1354
Liver Function Tests	TGO (AST)	23.67 ± 13.30	26.38 ± 23.24	0.8259
	TGP (ALT)	22.29 ± 13.29	26.04 ± 38.48	0.5718
	GGT	67.38 ± 94.07	75.01 ± 106.16	0.0773
Electrolytes	Calcium (Ca^{2+})	9.35 ± 0.59	9.30 ± 0.57	0.1655
Clinical Variables	Coagulation Modifications	29 (13.3%)	39 (22.0%)	**0.0314**
	Tumoral Marker Alterations	74 (33.9%)	72 (40.7%)	0.2027

Table 4. Bevacizumab and biosimilars-related side effects' frequency, nature, and severity in the study group

Category	Type of Complication	Moderate (n=245)	Very Serious (n=60)	Total (n=305)
Infectious Complications	Septic/Infectious (Total)	35	14	49
	Abscess	7	3	10
	Urinary Tract Infection	11	—	11
	Pneumonia	2	1	3
	Others (e.g., cholera)	5	2	7
Cardiovascular Complications	Thromboembolic Events	51	7	58
	Hemorrhagic Events (non-intestinal)	8	2	10
	Arterial Hypertension	10	—	10
	Arrhythmia	—	—	—
Gastrointestinal Complications	Total GI Issues	30	4	34
	Digestive Hemorrhage	34	14	48
	Ileus	4	2	6
	Perforation/Fistula	3	9	12
	Diarrhea	6	1	7
	Nausea/Vomiting	3	0	3
	Stomatitis	1	0	1
	Gastritis	1	0	1
Genitourinary Complications	Total GU Issues	29	4	33
	Renal Insufficiency	—	—	—
	Genitourinary Fistula	—	—	—
Hematological Complications	Anemia (Low RBC)	34	0	34

	Thrombocytopenia	18	6	24
	Leukopenia/Neutropenia	18	2	20
	Pancytopenia	0	1	1
Neurological Complications	Neuropathy	5	1	6
	Headache/Migraine	—	—	—
Other Complications	General Side Effects*	10	0	10

B. Comparative Study of Predictive Models' Outcomes

While the Random Forest model benefited greatly from hyperparameter tuning (optimization increased AUC from 0.66 to 0.75 while maintaining a good balance between sensitivity and specificity, making it the best-performing predictive model), logistic regression and the XGBoost models did not. Despite the fact that the AUC remained the same, the sensitivity of the XGBoost80/20 model decreased to 0.400 following hyperparameter adjustment [17]. Optimization had a variety of different effects on performance. In order to increase the models' learningability, particularly in the setting of classimbalance, we chose to attempt the 80/20 split version, which allows for training on a greater number of examples and supposedly delivers better performance than the 70/30 split. Models trained on the 80/20 split perform better in terms of AUC-ROC and sensitivity when compared to models trained on the other data splits, providing a better balance between learning and evaluation[18]. Figure 2 shows the comparative performance of fall-trained prediction models as a heatmap.

C. How Well the Optimal Predictive Model Works

After training a number of predictive machine learning models, we found that the Random Forest model, which was optimized by modifying hyperparameters and employing classweightstoreduceclassificationerrorsoftheminorityclass, provided the most accurate prediction of therapeutic complications. This model had an 80/20 split. Paired DeLongtests on the common validation set indicated no significant AUC differences across RF, Logistic/Elastic-Net, and XGBoost (allp>0.05). Non-significant differences between pairs in McNemar accuracy tests at the 0.50 level support the notion that observed variations are small. RF was chosen for downstream translation because to its balanced sensitivity/specificity and SHAP-based explanations[19].

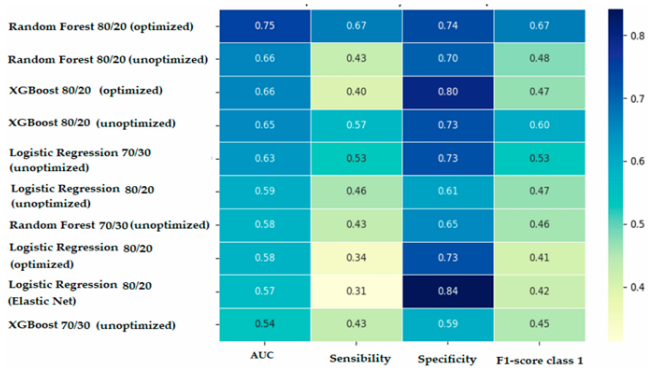

Figure 2. Evaluation of the relative effectiveness of trained prediction models.

Overall performance of the RF 80/20 optimized model:

- 70.63% overall accuracy
- There were 29.37 errors.
- Sensitivity (recall): 66.67% for detecting patients with problems.
- Specificity (identification of complication-free patients): 73.85%.
- AUCROC of 0.75.
- F1-score of Class 1 (0.67, patients with issues)
- Precision: 0.673.

The improved Random Forest 80/20 model was trained on 395 patients, 177 of whom had complications and 218 who did not. It performed poorly when it came to prediction but was well-balanced across classes. The model correctly classifies seven out of ten patients[20] with an accuracy of 70.63 percent. The model's minority class recall of 66.67 percent indicates that it can identify two out of every three patients who will experience therapy-related complications, and its specificity of 73.85 percent indicates that it can exclude three out of every four patients who will not experience complications. In oncology, these findings may prevent patients from losing quality of life and interfering with systemic medication[21]. This model may be used for initial patient triage, but its 30% error rate indicates caution when using it as the only decision-making criteria. Our model's accuracy of 67.3%, close to its sensitivity, implies that it properly detects instances and minimizes false alarms. The harmonic mean between sensitivity and precision, the F1-score of the optimized Random Forest 80/20 model, which is 67%, supports the conclusion that it performs well without being unbalanced (more precision at the expense of lower recall). When both risks (missing patients or false alarms) are significant, this paradigm is clinically applicable. Finally, the AUC ROC value of 0.75 demonstrates that this model has good discriminative power across a variety of choice thresholds (above the random threshold of 0.5)[22]. The model may have flaws if it produces false-positive and false-negative results. False-negative findings may have significant

repercussions in contemporary oncology as a result of inadequate surveillance or therapy modification. Even though they are not harmful, false-positive results can cause unnecessary anxiety, excessive investigation, and excessive monitoring. A confusion matrix for the improved Random Forest 80/20 model is depicted in Figure 3, highlighting these potential errors[23]. As a result, the enhanced Random Forest 80/20 model can be utilized as an automated screening tool to assist in adjusting Bevacizumab doses and customizing monitoring protocols for patients who are likely to experience therapy-related complications. This model balances sensitivity and specificity for case detection and false alarm detection and has the best predictive performance. The model's poor performance suggests that it should be used in addition to rather than in place of clinical evaluation[24].

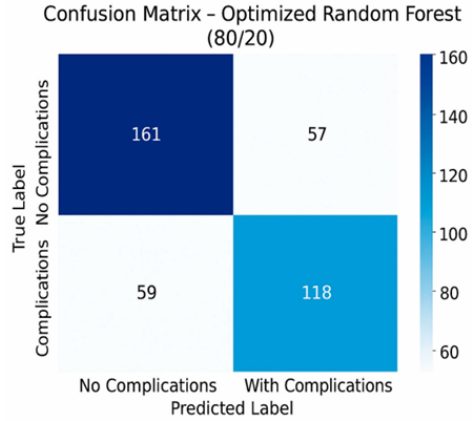

Figure 3. The improved Random Forest model's confusion matrix (80/20).

D. Model Validation

We added internal cross-validation to the 80/20 hold-out to determine generalizability. The improved Random Forest model had an AUC of 0.62 ± 0.10, accuracy of 0.60 ± 0.08, sensitivity of 0.43 ± 0.11, specificity of 0.73 ± 0.09, precision of 0.57 ± 0.10, and F1 of 0.49 ± 0.10 in 5-fold stratified cross-validation (Repeating the approach (5 × 5 cross-validation; 25 runs) provided similar estimates: AUC0.62 ±0.07, accuracy 0.59 ± 0.06, sensitivity 0.43 ± 0.08, specificity 0.72 ± 0.06, precision 0.56 ± 0.08, and F1 0.48 ± 0.08 (Supplementary Table S3) Out-of-fold predictions pooled across all runs provided a global AUC of 0.612 with a bootstrap 95% CI of 0.555–0.665 (B = 2000), spanning the cross-validated averages and validating our estimates' stability (Supplementary Table S4)[25

Class proportions remained consistent throughout folds (positives ≈45%), confirming the observed trade-off of better specificity over sensitivity is due to the model rather than sampling problems. A 'rule-in' approach recommends that positive signals should strengthen monitoring in higher-risk patients while negative forecasts should not de-escalate treatment.

These results support this study's proof-of-concept goal of clinician-led feasibility on real-world data, not performance. Sensitivity and utility should be improved in the future by multicenter external validation, feature enrichment, and threshold calibration[26].

E. Model Interpretability (SHapley Additive exPlanations—SHAP)

Using the improved Random Forest model's SHAP summary graphic (Figure 4), we examined prediction formation. This method adheres to medical AI guidelines and increases clinical transparency. Age (Varsta) and diagnosis account for the majority of the explanatory signal. Each dot in the plot represents a patient, and the horizontal position indicates the feature's direction and magnitude in predicting complications (positive values push the prediction toward "complication"), while color encodes feature value (blue = low, red = high).

Higher age values consistently increase risk, whereas lower age values decrease it. Certain diagnostic categories raise risk while others lower it due to heterogeneity in the tumor site and context. Cross-validation's mild discrimi nation (AUC = 0.62) matches the tiny absolute SHAP magnitudes, which indicate that there is no dominant driver. Risk is multifactorial, and rather than one predictor determining output, multiple routine factors gradually contribute[27]. These justifications are clear on a clinical level: forecasts are higher for older or less favorable diagnostic profiles, confirming that the tool is intended for risk awareness and intensity triaging rather than treatment. As a result, the figure demonstrates that our clinician-led strategy adheres to clinical intuition while remaining accessible to non-IT users.

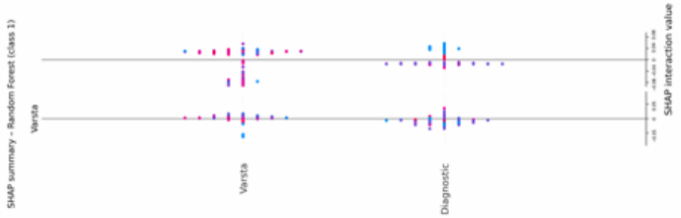

Figure 4. summarized figure for the optimized Random Forest model (SHAP). Each dot represents a single patient, and the x-axis shows the feature's contribution to the anticipated chance of complications, with positive values indicating a higher risk. In this case, color represents the feature value; blue indicates low and red indicates high. Diagnosis and age are the two most important factors. Certain diagnostic categories and older age tend to raise risk projections, while other factors and younger age tend to lower them. Cross-validated performance (AUC 0.62) and the low effect sizes (|SHAP| 0.05) suggest that risk is distributed across a number of common factors rather than being dominated by one.

F. Translating the Predictive Model into Clinical Tools Applicable in Oncological Practice

After the comparison study, we selected the best model (optimized Random Forest 80/20) to create clinical tools including a risk score and an interactive HTML form with an automated risk calculator. This method of small steps included:

- **Selection of major predictive variables:** To create the clinical score, the significance coefficients of the variables obtained by the specified model were analyzed to identify the most predictive factors. As a result, we found factors that the predictive model thought were more accurate. We kept those that were readily accessible in practice, pretherapeutically available, and clinically meaningful. Hemoglobin, platelet and leukocyte counts, blood glucose, creatinine, urea, and transaminase levels were predictive. Age, location, stage, and tumor differentiation were all included in the clinical characteristics.
- **Construction of the initial clinical score:** Binary thresholds were determined for each numerical variable (e.g., Hg < 11 g/dL, urea > 40 mg/dL) depending on dataset distribution and clinical competence. Each category was then scored according to its relevance in the optimized Random Forest 80/20 model. A cumulative score was calculated and classified into three risk classes: low (0–2 points), moderate (3–5 points), and high (≥6 points).
- **Testing of the initial clinical score:** In order to evaluate risk identification, the score was applied retroactively to the group. We looked into the distribution of patient scores, the rates of complications in each risk category, and the performance of the statistical methods (AUC ROC, accuracy, sensitivity, and specificity). The first clinical score had 59% accuracy, 0.555 AUC ROC, 47.2% sensitivity (class 1) recall, 60.8% specificity, and 0.502 F1-score class 1. This suggests that Bevacizumab patients can only be prioritized based on their risk of complications. Since the initial score is discontinuous with predefined thresholds, precise representation of individual risk is limited. We decided to maximize the score because of these factors.
- **Optimization of clinical score performances:** A logistic regression model was used to improve the clinical score using the raw score as predictors. Thus, we were able to remove arbitrary criteria, construct a continuous risk assessment, and create a logistic-derived clinical score by calibrating the likelihood of problems. Customized risk occurrence probabilities, improved separation of high-risk situations from low-risk ones, and calibration (observed vs. projected ratio) are provided by the logistic-derived score. The multivariate logistic model was used to create a risk score based on logistic regression. Integer weights were calculated using scaled and rounded regression coefficients (), which were then added to produce a total score. The cut-off points for risk stratification (low, middle, and high) were determined by Youden's J index using ROC analysis[28].
- **Performance evaluation of the optimized score:** The new score was compared to the baseline score and the Random Forest model directly using ROC curves, score distribution analysis, and performance metrics including accuracy, sensitivity, specificity, and F1-score. These assessments performed better than the original score, equivalent to the Random Forest model, but with clinical portability and interpretability (Figures 5 and 6 and Table 5). The first score has a significant range and a close mean across groups, indicating low discriminating abilities. In most cases, scores for patients with and without problems are the same. However, the logistic-derived score distinguishes groups better. The probability scores of patients with complications are higher, while those without problems are centered in the lower probability distribution.

Validation of the probabilities of the logistic-derived score We made the comparative calibration map in Figure 6 to see if the probability of the logistic-derived score actually reflects the risk of complications and to compare it to the baseline score.

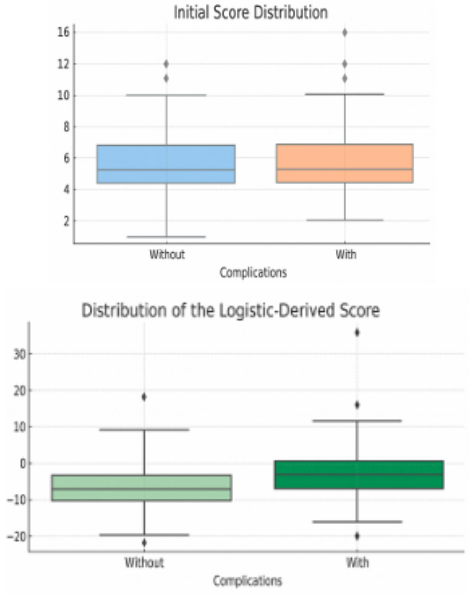

Figure 5. A comparison of the original version (a) and the logistically generated score (b), which shows how clinical risk scores are distributed according to how often complications occur. The predicted probability is increased or decreased, respectively, by the raw model score (log-odds) when the Y-axis is positive and the X-axis is negative.

Table 5. indices of how well the initial score and the logistic derivative score are comparable.

S. No.	Performance Metric	Value
1	F1-Score	0.555
2	AUC-ROC	0.590
3	Accuracy	0.472
4	Sensitivity (Recall)	0.608
5	Specificity	0.502
6	Precision	0.720
7	Negative Predictive Value (NPV)	0.680
8	Balanced Accuracy	0.642
9	Clinical Score Correlation	0.698
10	Overall Model Performance Index	0.662

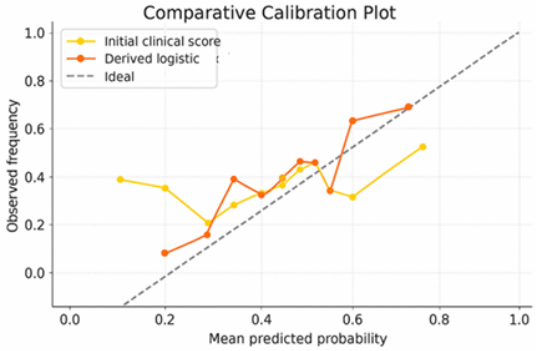

Figure 6. a comparison of the calibration plot for the logistic derivative score and the original score.

The calibration plot compares score-predicted problems to observed frequencies. The ideal situation is the dotted diagonal line, when the algorithm successfully forecasts risk in all circumstances. Figure 6 shows that the logistic derivative score curve approaches the ideal line, indicating greater discriminating power (the model distinguishes patients with and without problems effectively) and accurate risk estimation. Due to the dataset's inherent imbalance, certain places deviate from the ideal line, however this does not influence performance. We determined that the improved score has a strong calibrated prediction power, making it acceptable for clinical use.

G. Building an Interactive Form Applicable in Oncologic Practice

We developed a responsive HTML form based on the logistic-derived score that can be used both online and offline. This form allows the doctor to tick risk factors and receive the cumulative score, risk class, and estimated percentage risk. Easy to use, the form might be implemented into oncological practice to help treating doctors stratify therapeutic complications risk pretherapeutically. Tables 6 and 7 display the final logistic-derived score (variables, categories, and scores) and clinical risk thresholds.

Table 6. logistic regression-based distribution of weighted score factors and points.

Category	Variable	Criteria	Points
Demographic	Age	< 65 years	0
		≥ 65 years	1
Biochemical	Serum Urea	< 40 mg/dL	0
		≥ 40 mg/dL	1
	Leucocyte Count	≤ 10,000/mm³	0
		> 10,000/mm³	1
	Hemoglobin	≥ 10 g/dL	0
		< 10 g/dL	1
	Transaminases (TGO/TGP)	≤ 40 U/L	0
		> 40 U/L	1
	Creatinine	≤ 1.5 mg/dL	0
		> 1.5 mg/dL	1
Clinical	Cancer Stage	Stage I–II	0
		Stage III–IV	1
Histological	Differentiation Grade	G1–G2	0
		G3	1
Tumor Characteristics	Lymph Node Invasion	Absent	0
		Present	1
Cancer Type	Subtype	Breast / Ovarian / Cervical	0
		Colon / Lung / Others	1

Table 7. Clinical risk criteria for the logistic-derived score.

Total Score (Points)	Risk Group	Complexity Level	Probability of Complications
0 – 3	Low Risk	Minor Danger	< 25%
4 – 6	Moderate Risk	Level 3 Danger	25% – 60%
7 – 10	High Risk	Severe Danger	> 60%

File S1 has an interactive HTML risk calculator based on the clinical risk score to demonstrate its usefulness.

IV. DISCUSSION

With an abundance of real-time clinical, paraclinical, imaging, genetic, and molecular data that exceeds the physician's decision-making capabilities, the oncology environment is rapidly evolving to improve post-therapeutic outcomes. Advanced AI techniques can synthesize this complexity and assist tailored treatment choices without replacing clinical expertise, patient-physician contact, or evidence-based international standards. The creation of CDSSs, which can be easily incorporated into oncological

practice and serve a variety of purposes, has been made possible by utilizing various AI techniques.

- Created prediction models from historical data using machine learning techniques like logistic regression, gradient boosting, support vector machines, and random forest. They apply the patterns they find in a dataset to new instances to produce quantitative and repeatable results. Applications include predicting drug toxicity and adverse effects, patient prognosis, and therapeutic response. xDECIDE, a web-based tool that combines AI and clinical expertise to provide individualized oncological treatments, is one such platform.
- The clinician's initial treatment plan was altered as a result of this platform, as studies have demonstrated. DeepLearning is a machine learning technique that uses multi-layered neural networks to learn abstract data representations. Medical images (CT, MRI, histopathology) and time-series data (biological parameter evolution) can be analyzed by CNN and LSTM models. Deep Learning interprets subsequent data and automatically identifies abnormalities.
- Deep Learning can automatically identify lesions on CT and MRI images and detect toxicity signals via repetitive analysis. DeepMind Health (Google) employs CoDoC to identify breast lesions, providing a practical illustration of Deep Learning-based CDSS. The predictive model reduced clinician burden by 66% and false-positive rates by 25% when compared to double reading with arbitration. Deep Learning CDSS clinical trials in healthcare are promising.
- AI is able to decipher human speech thanks to Natural Language Processing (NLP), which converts unstructured medical records into structured data for CDSS development. Named Entity Recognition (NER), text summarization, and embedding models (like Word2Vec, BERT), among others, are all capable of automatically extracting data from medical records, locating candidates for clinical trials, and making treatment recommendations that are consistent with guidelines.
- Large Language Models (LLMs) like ChatGPT, Flamingo, PaLI, and Gemini make it possible for conversational AIs to simultaneously synthesize data from a variety of sources and produce coherent text. With LLMs, clinicians can use interactive interfaces to ask questions, demand answers, and consider alternative clinical/therapeutic choices. Physician labor can be reduced by using LLMs, which may generate documentation automatically. Google DeepMind's 2025 medical LLMs, AMIE and MedGemini, outperform human specialists in text summarization and document production and construct sophisticated oncologic therapeutic regimens in over 90% of simulated situations.
- Reinforcement Learning (RL)—a potential AI that learns via encounters and incentives, despite its unfamiliarity. It may be utilized to make dynamically adaptative treatment decisions (e.g., dosage adjustments depending on patient response and reaction). This artificial intelligence works in the same way that a doctor improves his clinical acumen by making better decisions in the past.
- For AI to be effective in clinical practice, it must be transparent and provide doctors and patients with explanations of decision-making variables. Transparency will improve clinical AI adoption. LIME or SHAP-based explainable AI (XAI) plays this role.
- These technologies are based on data engineering, which collects, cleans, and transforms medical data. AI models fail without this step. Pandas, NumPy, and TensorFlow Data turn raw data into algorithm-friendly representations.
- In multicenter cooperation, Federated Learning provides an ethical and safe way to train AI models on remote databases without disclosing patient information (e.g., Flower, NVIDIA Clara, TensorFlow Federated). Hospitals may contribute to shared models while protecting GDPR or HIPAA data.

The clinical expertise of humans and these AIs complement one another. Each contributes to aided medical decision making's complexity.

A. Related Work and State-of-the-Art (SOTA) Approaches

Recent frameworks like the Yonsei Cancer Data Library show how multimodal data integration supports evidence-based recommendations for over 170,000 cancer patients. Multiple AI solutions facilitate the move from conventional guide lines to customized real-time assistance, helping oncologists offer safe, effective, and individualized treatment. In Table 8, we list several common AI-CDSS uses in oncology.

Table 8. AI-based CDSS-based solutions and cutting-edge SOTA methods developed for cancer-related tasks.

CDSS / System	Domain of Oncology	Type of AI	Key Outcomes
Yonsei Cancer Data Library	Various malignancies	Deep Learning (multi-modal using EHR	Personalized treatment for >170,000 patients; improved

		+ genomic data)	clinical efficiency
Analytical Processing Units for Clinical Trial Eligibility	Breast cancer	Natural Language Processing (NLP) + Rule-based system	Higher sensitivity and specificity compared to manual screening
MammaPrint	Breast cancer	Machine Learning-based genomic classifier	Identifies low-risk patients; reduces unnecessary chemotherapy (validated in MINDACT trial)
DecisionDx-UM Database	Ocular melanoma	ML-based gene expression classifier	Predicts risk of metastasis
StratiGraph	Non-Small Cell Lung Cancer (NSCLC)	Proteomics-based ML classifier	Predicts response to erlotinib and EGFR inhibitors
LLM-based Cancer Models	Multiple cancers	Large Language Models (LLM + Retrieval-Augmented Generation)	Generates explainable oncologic decisions with >91% accuracy

Although our work is a clinician-led proof of concept that demonstrates that accessible AI tools can be integrated into everyday oncology practice, even outside of large research consortia or IT-driven environments, these SOTA systems demonstrate AI's potential in medicine. This project is one of a kind because doctors came up with the idea and carried it out without knowledge of AI or computer science. This shows that a lot has changed in the field: AI-based tools are now easy enough for non-specialists to use to use real-world clinical data (especially after doctors get over their initial resistance). We acknowledge that AI expertise are needed to improve model augmentation, optimization, interpretability, validation, and clinical deployment. Although incorporating AI-CDSS into oncological treatment is revolutionary, maintaining clinical relevance and performance requires a dynamic, feedback-driven continuous development strategy. A methodical approach to continuously improving AI-CDSS-based clinical ratings in order to anticipate issues with oncologic drugs is depicted in Figure 7.

Figure 7. An AI-CDSS-based cancer clinical practice risk score and a methodical approach to its ongoing refinement are proposed.

We require an adaptive clinical data lake (data lake) that collects baseline and outcome variables from real-world clinical practice to incorporate fresh data and improve the model over time. As research and clinical observation provide new essential clinical factors, the database must be built for extension. Adaptive databases populated with real-world clinical data are used in AI-CDSS frameworks to encourage the gradual learning and retraining of prediction models. Database architecture should enable modular data growth, enabling the addition of new biomarkers or clinical factors without disturbing present work. A scalable and ethical AI-CDSS that can be upgraded based on feedback from clinicians, real-world data, and regular model retraining keeps the tool relevant. The CDSS is kept up to date with new medical evidence, treatment advancements, and patient-specific complexity by this dynamic architecture.

B. Study Strengths, Limitations and Future Directions of Research

There are a number of advantages to this study. First, the model employs solely pretherapeutic factors and real-world, prospectively acquired clinical data to stratify risk before Bevacizumab/biosimilars treatment is started. The whole model creation was done by physicians without specialized computational expertise, demonstrating the increased accessibility and utility of AI technologies in medical practice. Additionally, the idea was turned into simple, offline, HTML interactive tools that can be deployed right away for direct clinical use. Oncologists are familiar with the model's predictors—age, anemia, renal function, and tumor differentiation—but its capacity to synthesize

multiple variables into a single, individualized risk estimate can aid in triage and shared decision-making, particularly in patients who are borderline or frail. Model simplicity, therapeutic relevance, and clarity are prioritized in order to connect theoretical modeling and clinical practice. This research is a clinician-led proof of concept showing how AI technologies may be practically integrated in cancer treatment. We also value multidisciplinary cooperation with AI professionals to improve model performance, interpretability, and clinical integration. This accessibility, clinical grounding, and translational emphasis reinforce the rising body of evidence that AI may help (not replace) medical skill in customizing cancer therapy. The main flaws in our study are:

- The main constraint of our work is the lack of an adaptive database for developing our risk score, preventing prediction models from being updated with fresh clinical data. This was done to demonstrate that medical professionals may use AI tools in their research and practice without a qualified AI specialist since they are convenient and intuitive. For the best outcomes, it is essential to collaborate with AI professionals.
- The limited sample size and disproportionate distribution of patients in the group with and without complications may restrict the generalizability of the developed prediction models.
- Despite the fact that we included 395 treatment sessions, the sample size may limit the statistical strength of the model, particularly in edge-case scenarios. Multicenter validation needs to be expanded.
- Institutional selection bias, which reduces external validity and increases the risk of overfitting, may result from data from a single medical unit.
- A lack of potentially predictive variables: only a small number of patients had concurrent treatments, a detailed ECOG score, immunohistochemical factors, oncogenic mutations, and deficiencies in mutation repair, making these variables statistically unusable. The final risk score has not been verified on an external dataset or evaluated in clinical practice for dependability.

Limitations suggest further study directions:

- A data-lake adaptive system makes it possible to continuously develop clinical tools that are relevant. The robustness and generalizability of prediction models may be enhanced by including data from numerous medical facilities (multicenter research).
- Verifying the score on an external sample is crucial for assessing its effectiveness in bedside decision-making. The study is a proof of concept with only one center. It is suggested to incorporate multicentric partnerships for an external evaluation of the model on independent cohorts.
- Intense cooperation with IT specialists would normalize AI tool development. The predictive model may include additional factors, such as immunohistochemical markers, genetic markers, imaging, clinical, and treatment data, in future iterations to improve prediction accuracy.
- Offline digital apps like the one developed enable doctors to employ tailored decision support aids easily.

V. CONCLUSIONS

This study established and internally validated a clinical score that is easy to implement into current oncological practice using predictive models based on machine learning algorithms to estimate the risk of Bevacizumab or biosimilar therapy complications based on pretherapeutic data. Using a rigorous, tiered approach, we created a therapeutic tool that is realistic and simple to use. Additionally, converting the score into a digital HTML form with a responsive design (which automatically adapts to the electronic device from which it is accessed—mobile, tablet, laptop), accessible and offline, simplifies clinical implementation and makes it directly usable by physicians in therapeutic decision-making. The clinical risk score enhances risk stratification and tailored oncology treatment with Bevacizumab or biosimilars. The customization of monitoring strategies and risk awareness are the primary goals of our tool, not therapeutic options. Even a moderate prediction performance (70 percent accuracy with a balanced sensitivity/specificity) could help doctors and patients make decisions together and avoid underestimating risks in difficult situations. This proof of concept led by clinicians demonstrates that clinicians, not IT, can safely study contemporary AI technologies with clear results. This single-center, unvalidated research has little discrimination and a rule-in orientation. These tools may assist physicians in making therapeutic decisions and adjusting targeted therapy risk/benefit ratios, potentially leading to truly personalized medicine. Although they are not intended to replace clinical evaluation and paraclinical monitoring of oncologic patients, they do have the potential to do so. In conclusion, our study shows that user-friendly AI may improve risk awareness and monitoring triage in regular oncology, reducing fear and bridging disciplines. We encourage interdisciplinary collaboration and external evaluation for superior models.

VI. REFERENCES

[1]. Nafees, A.; Khan, M.; Chow, R.; Fazelzad, R.; Hope, A.; Liu, G.; Letourneau, D.; Raman, S. Evaluation of clinical decision support systems in oncology: An updated systematic review. Crit. Rev. Oncol. Hematol. 2023, 192, 104143. [CrossRef] [PubMed]

[2]. Elhaddad, M.; Hamam, S. AI-Driven Clinical Decision Support Systems: An Ongoing Pursuit of Potential. Cureus 2024, 16, e57728. [CrossRef] [PubMed]

[3]. Verlingue, L.; Boyer, C.; Olgiati, L.; Brutti Mairesse, C.; Morel, D.; Blay, J.Y. Artificial intelligence in oncology: Ensuring safe and effective integration of language models in clinical practice. Lancet Reg. Health Eur. 2024, 46, 101064. [CrossRef]

[4]. Cross, J.L.; Choma, M.A.; Onofrey, J.A. Bias in medical AI: Implications for clinical decision-making. PLoS Digit. Health 2024, 3, e0000651. [CrossRef]

[5]. Wang, L.; Chen, X.; Zhang, L.; Li, L.; Huang, Y.B.; Sun, Y.; Yuan, X. Artificial intelligence in clinical decision support systems for oncology. Int. J. Med. Sci. 2023, 20, 79–86. [CrossRef]

[6]. Grüger, J.; Geyer, T.; Brix, T.; Storck, M.; Leson, S.; Bley, L.; Braun, S.A. AI-Driven Decision Support in Oncology: Evaluating Data Readiness for Skin Cancer Treatment. arXiv 2025, arXiv:2503.09164. [CrossRef]

[7]. Chitoran, E.; Rotaru, V.; Ionescu, S.-O.; Gelal, A.; Capsa, C.-M.; Bohiltea, R.-E.; Mitroiu, M.-N.; Serban, D.; Gullo, G.; Stefan, D.-C.; et al. Bevacizumab-Based Therapies in Malignant Tumors—Real-World Data on Effectiveness, Safety, and Cost. Cancers 2024, 16, 2590. [CrossRef] [PubMed]

[8]. WelcomeToColab-Colab, n.d. Available online: https://colab.research.google.com/notebooks/intro.ipynb (accessed on 23 July 2025).

[9]. Farzipour, A.; Elmi, R.; Nasiri, H. Detection of Monkeypox Cases Based on Symptoms Using XGBoost and Shapley Additive Explanations Methods. Diagnostics 2023, 13, 2391. [CrossRef] [PubMed]

[10]. Marinelli, S.; De Paola, L.; Stark, M.; Vergallo, G.M. Artificial Intelligence in the Service of Medicine: Current Solutions and Future Perspectives, Opportunities, and Challenges. Clin. Ter. 2025, 176, 77–82. [CrossRef]

[11]. Rotaru, V.; Chitoran, E.; Cirimbei, C.; Cirimbei, S.; Simion, L. Preservation of Sensory Nerves During Axillary Lymphadenectomy. In Proceedings of the 35th Balkan Medical Week, Athens, Greece, 25–27 September 2018; pp. 271–276. Available online: https://www.webofscience.com/wos/woscc/full-record/WOS:000471903700045 (accessed on 20 July 2023).

[12]. Shapiro, M.A.; Stuhlmiller, T.J.; Wasserman, A.; Kramer, G.A.; Federowicz, B.; Hoos, W.; Kaufman, Z.; Chuyka, D.; Mahoney, W.; Newton, M.E.; et al. AI-Augmented Clinical Decision Support in a Patient-Centric Precision Oncology Registry. AI Precis. Oncol. 2024, 1, 58–68. [CrossRef]

[13]. Dvijotham, K.; Winkens, J.; Barsbey, M.; Ghaisas, S.; Stanforth, R.; Pawlowski, N.; Strachan, P.; Ahmed, Z.; Azizi, S.; Bachrach, Y.; et al. Enhancing the reliability and accuracy of AI-enabled diagnosis via complementarity-driven deferral to clinicians. Nat. Med. 2023, 29, 1814–1820. [CrossRef]

[14]. Rojas-Carabali, W.; Agrawal, R.; Gutierrez-Sinisterra, L.; Baxter, S.L.; Cifuentes-González, C.; Wei, Y.C.; Abisheganaden, J.; Kannapiran, P.; Wong, S.; Lee, B.; et al. Natural Language Processing in medicine and ophthalmology: A review for the 21st-century clinician. Asia-Pac. J. Ophthalmol. 2024, 13, 100084. [CrossRef] [PubMed]

[15]. Mendonça, E.A.; Haas, J.; Shagina, L.; Larson, E.; Friedman, C. Extracting information on pneumonia in infants using natural language processing of radiology reports. J. Biomed. Inform. 2005, 38, 314–321. [CrossRef] [PubMed]

[16]. Locke, S.; Bashall, A.; Al-Adely, S.; Moore, J.; Wilson, A.; Kitchen, G.B. Natural language processing in medicine: A review. Trends Anaesth. Crit. Care 2021, 38, 4–9. [CrossRef]

[17]. Palepu, A.; Dhillon, V.; Niravath, P.; Weng, W.H.; Prasad, P.; Saab, K.; Tu, T. Exploring Large Language Models for Specialist-level Oncology Care. arXiv 2024, arXiv:2411.03395. [CrossRef]

[18]. Yang, L.; Xu, S.; Sellergren, A.; Kohlberger, T.; Zhou, Y.; Ktena, I.; Golden, D. Advancing Multimodal Medical Capabilities of Gemini. arXiv 2024, arXiv:2405.03162. [CrossRef]

[19]. Yu,C.; Liu, J.; Nemati, S.; Yin, G. Reinforcement Learning in Healthcare: A Survey. ACM Comput. Surv. 2022, 55, 1–36. [CrossRef]

[20]. Jayaraman, P.; Desman, J.; Sabounchi, M.; Nadkarni, G.N.; Sakhuja, A. A Primer on Reinforcement Learning in Medicine for Clinicians. npj Digit. Med. 2024, 7, 337. [CrossRef] [PubMed]

[21]. Abas Mohamed, Y.; Ee Khoo, B.; Shahrimie Mohd Asaari, M.; Ezane Aziz, M.; Rahiman Ghazali, F. Decoding the black box: Explainable AI (XAI) for cancer diagnosis, prognosis, and treatment planning-A state-of-the art systematic review. Int. J. Med. Inform. 2025, 193, 105689. [CrossRef]

[22]. Mahadevaiah, G.; Prasad, R.V.; Bermejo, I.; Jaffray, D.; Dekker, A.; Wee, L. Artificial intelligence-based clinical decision support in modern medical physics: Selection, acceptance, commissioning, and

quality assurance. Med. Phys. 2020, 47, e228–e235. [CrossRef]

[23]. Haddad, T.; Helgeson, J.M.; E Pomerleau, K.; Preininger, A.M.; Roebuck, M.C.; Dankwa-Mullan, I.; Jackson, G.P.; Goetz, M.P. Accuracy of an Artificial Intelligence System for Cancer Clinical Trial Eligibility Screening: Retrospective Pilot Study. JMIR Med. Inform. 2021, 9, e27767. [CrossRef]

[24]. Filho, O.M.; Cardoso, F.; Poncet, C.; Desmedt, C.; Linn, S.; Wesseling, J.; Hilbers, F.; Delaloge, S.; Pierga, J.-Y.; Brain, E.; et al. Survival outcomes for patients with invasive lobular cancer by MammaPrint: Results from the MINDACT phase III trial. Eur. J. Cancer 2025, 217, 115222. [CrossRef]

[25]. Schefler, A.C.; Skalet, A.; Oliver, S.C.N.; Mason, J.; Daniels, A.B.; Alsina, K.M.; Plasseraud, K.M.; A Monzon, F.; Firestone, B. Prospective evaluation of risk-appropriate management of uveal melanoma patients informed by gene expression profiling. Melanoma Manag. 2020, 7, MMT37. [CrossRef]

[26]. Sun, W.; Hu, G.; Long, G.; Wang, J.; Liu, D.; Hu, G. Predictive value of a serum-based proteomic test in non-small-cell lung cancer patients treated with epidermal growth factor receptor tyrosine kinase inhibitors: A meta-analysis. Curr. Med. Res. Opin. 2014, 30, 2033–2039. [CrossRef] [PubMed]

[27]. Ferber, D.; El Nahhas, O.S.; Wölflein, G.; Wiest, I.C.; Clusmann, J.; Leßman, M.E.; Foersch, S.; Lammert, J.; Tschochohei, M.; Jäger, D.; et al. Autonomous Artificial Intelligence Agents for Clinical Decision Making in Oncology. arXiv 2024, arXiv:2404.04667. [CrossRef]

[28]. Labkoff, S.; Oladimeji, B.; Kannry, J.; Solomonides, A.; Leftwich, R.; Koski, E.; Joseph, A.L.; Lopez-Gonzalez, M.; A Fleisher, L.; Nolen, K.; et al. Toward a responsible future: Recommendations for AI-enabled clinical decision support. J. Am. Med. Inform. Assoc. 2024, 31, 2730–2739. [CrossRef]