Cureus Journal of Computer Science

Open Access Original Article

Machine Learning for Predicting Financial Toxicity in Patients Receiving Complex Gastrointestinal Surgery

Received 04/30/2025
Review began 05/07/2025
Review ended 09/03/2025
Published 09/19/2025

DOI:
https://doi.org/10.7759/s44389-025-05997-w

Ruth Bahre [1] , Alex Masterson [2] , Gitonga Munene [3] , Zachary D. Asher [2] , Sahar Abedzadeh [4] , A C Fong [4]

1. Department of Electrical and Computer Engineering, Western Michigan University, Kalamazoo, USA 2. Department of Mechanical and Aerospace Engineering, Western Michigan University, Kalamazoo, USA 3. School of Medicine, Western Michigan University, Kalamazoo, USA 4. Department of Computer Science, Western Michigan University, Kalamazoo, USA

Corresponding author: A C Fong, acmfong@gmail.com

Abstract

Financial toxicity (FT), defined as the economic burden associated with medical care, has emerged as a critical factor affecting patient outcomes. However, this topic remains understudied in the context of complex gastrointestinal surgery for cancer patients. This study aimed to identify predictors of FT and develop a machine learning model capable of preoperatively stratifying patient risk. A dataset of 188 patients who underwent gastrointestinal surgery at Western Michigan Cancer Center during 2016-2023 was analyzed using a range of machine learning techniques. The Functional Assessment of Chronic Illness Therapy Comprehensive Score for Financial Toxicity instrument was used to assess FT, with a score of ≤26 indicating significant financial distress. Multivariable logistic regression was used for baseline analysis, while several classification models including Random Forest, Gradient Boosted Trees, Support Vector Machine, Naïve Bayes, K-Nearest Neighbors, and Neural Networks were evaluated to predict FT. The dataset was then split into training and testing sets using an 80/20 stratified sampling strategy, which ensured class distribution was preserved across both sets. Among all models tested, the Random Forest classifier demonstrated superior performance, achieving the highest accuracy and area under the curve. In addition, feature importance analysis revealed that social well-being, emotional well-being, and age were the most influential predictors of FT. These findings suggest that financial toxicity is not only prevalent but also predictable using readily available clinical and sociodemographic data. Integrating such predictive tools into preoperative workflows may support early intervention strategies aimed at mitigating the adverse effects of financial hardship on patient quality of life.

Categories: Health Informatics, Machine Learning (ML), Predictive Analytics
Keywords: financial toxicity, predictive system, cancer patient care, modeling, machine learning

Introduction

Advances in cancer treatment have improved survival rates for many patients diagnosed with different types of cancers. However, many forms of state-of-the-art cancer treatment can incur substantial expenses. When a sizable portion of such costs is borne by individual patients, this can lead to significant financial hardship. In the context of oncology, financial toxicity (FT) is a term that describes the financial hardship faced by cancer patients as they undergo potentially lifesaving treatment.

Financial hardship can have broad and deep impacts on individual patients. Studies, e.g., [1,2], have found that such impacts can be multifaceted, affecting patients' health and finance. Understanding the effects of FT is important because financial burden on patients can negatively impact their physical and mental well-being [3]. In some cases, the harmful effects of financial hardship can be greater than any actual or perceived benefits of undergoing treatment. Therefore, modeling and early prediction of FT can be highly desirable for patients and their caregivers to develop a treatment plan that is optimal for each individual patient based on well-informed decisions.

Gastrointestinal (GI) cancers are a major cause of morbidity and mortality worldwide, resulting in over 4.8 million new cases and 3.4 million deaths in 2020 alone [4,5]. GI cancers affect organs in the digestive tract. Common types of GI cancers include stomach, colorectal, liver, gallbladder, pancreas, and small intestine. For many of these cancers, treatment involves complex surgical procedures. While many of these procedures are effective, they can also lead to temporary or even long-term disability. Such disability, even if limited in time, can negatively affect a patient's - and potentially their immediate family members' - ability to participate in the workforce. This can place a heavier financial burden on the patients than the treatment alone [6,7].

The risks of FT associated with different types of cancer and treatment are known to vary significantly [8]. Despite the prevalence of some GI cancers, such as colorectal cancer, the risks of FT have not been widely studied in the context of GI surgical procedures. It has been reported that FT rates can be as high as 50%

in gynecologic oncology [9]. In a relatively rare study [10] on the topic of FT modeling for GI surgery, the authors applied statistical data analyses on real-world data using SPSS version 27.0 [11]. Using validated tools like Functional Assessment of Chronic Illness Therapy Comprehensive Score for Financial Toxicity (FACIT-COST) and Functional Assessment of Cancer Therapy: General (FACT-G), the study [10] surveyed 188 patients who underwent GI surgery at Western Michigan Cancer Center, identifying that 31.4% experienced FT.

This paper builds on the seminal work of Young et al. [10] by leveraging contemporary machine learning (ML) algorithms to significantly enhance the predictive capability of FT models for the same anonymized dataset. The primary objective of this paper is to present a comprehensive comparative analysis of a suite of supervised ML models including multivariable logistic regression, random forest, gradient-boosted trees (GBTs), support vector machines (SVM), naïve Bayes, neural networks, and K-nearest neighbors (KNN). To ensure a holistic view of each ML model's performance, predictive performance of each model was evaluated using multiple metrics, namely accuracy, sensitivity, positive and negative predictive values, and area under the receiver operating characteristic curve (AUC-ROC). In addition to classification accuracy, our secondary objective is to identify the most influential predictors of FT. Taken together, our overarching goal is to determine the most effective model for predicting FT while providing actionable insights to guide early interventions in surgical oncology care.

Materials And Methods

In this study, we adopt a standard ML workflow that begins with data preparation.

Data preparation

The dataset for this study was provided by the Western Michigan University Homer Stryker M.D. School of Medicine (WMED) and includes a total of 188 patient responses collected from individuals who underwent complex GI surgery from 2016 to 2023. The dataset contains 25 variables that span demographic, clinical, socioeconomic, and patient-reported domains. A breakdown of the dataset can be seen in Table *1*.

Variable Name	Data Type	Unique Values	Missing Values
Race	Object	3	1
Age	Float64	41	1
Sex	Object	3	1
Type of Cancer	Object	87	1
Type of Operation	Object	101	1
Date of Surgery	Object	97	1
Systemic Therapy	Object	3	1
Radiation Therapy	Object	3	1
Employment Status	Object	5	1
Type of Insurance	Object	4	1
Marital Status	Object	4	1
Rural-Urban Continuum	Object	2	1
ADI Quintile	Float64	5	1
Education Level	Float64	4	1
Distance to Cancer Center	Float64	57	1
Number of Admissions	Float64	3	1
Number of ED Visits	Float64	3	1
COST Score (FT)	Float64	61	0
Time Since Surgery	Float64	90	1
Social Well-Being Score	Float64	36	1
Emotional Well-Being Score	Float64	36	1
Complications	Object	2	1
Clavien-Dindo Grade	Object	4	183
Discharged	Object	2	1

TABLE 1: Summary of dataset variables

ADI, Area of Deprivation Index; COST, COmprehensive Score for financial Toxicity; ED, Emergency Department; FT, Financial Toxicity

The primary outcome variable in this study is FT, assessed using the validated FACIT-COST instrument. FT severity was categorized based on COST scores as follows: no FT (COST > 26), mild FT (COST 14-26), moderate FT (COST > 0-14), and severe FT (COST = 0). A COST score of ⩽26 was considered indicative of any degree of FT, while a score >26 was classified as no FT. This allowed us to frame the analysis as a binary classification problem. Threshold values were determined based on established scoring guidelines from the FT questionnaire [12].

Initial preprocessing began with the removal of the Clavien-Dindo grade variable due to its high rate of missing values. Categorical entries were standardized to eliminate inconsistencies - for example, "F" and "M" were converted to "Female" and "Male," respectively. Another inconsistency was found in the distance to cancer center variable, where some entries included the unit "mi," which was removed to ensure accurate categorization. Missing values were imputed using the mode for categorical variables and the median for continuous variables.

The dataset was then split into training and testing sets using an 80/20 stratified sampling strategy, which ensured class distribution was preserved across both sets. The cleaned and processed dataset was used to train and evaluate several supervised learning models.

Metrics

To evaluate the performance of the predictive models developed in this study, we implemented a comprehensive set of evaluation metrics that quantify different aspects of classification performance. These metrics include overall accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and AUC-ROC. Selection of these metrics was guided by their application in the original study on FT in complex GI surgery [10], enabling consistent performance benchmarking across modeling approaches.

ML models

Using the above metrics for evaluation, we performed a comparative analysis of seven ML models. We were interested to discover how each ML model would perform, and whether any of the models would emerge as a clear winner using multiple metrics mentioned above. The following subsections explain and describe the seven ML models investigated in this study.

Multivariable Logistic Regression

Multivariable logistic regression is a probabilistic, parametric model widely used in medical and clinical research for its balance of interpretability, simplicity, and predictive capability. It estimates the probability of a binary outcome - such as the presence or absence of FT - using a linear combination of predictor variables. To ensure compatibility with the logistic model, all categorical variables were binarized or one-hot encoded, and continuous variables were standardized where appropriate, as shown in Table *2*. This model was selected as the baseline comparator due to its strong interpretability, clinical relevance, and established use in healthcare decision-making. Model implementation was conducted using the ‘LogisticRegression’ class from the scikit-learn library (v1.2) in Python.

Parameters	Binary Classification = 1	Binary Classification = 0
Sex	Male	Female
Race	Black	White
Type of Cancer	Pancreas	Other
Employment Status	Employed	Other
Type of Insurance	Private	Other
Marital Status	Married	Other
Education Level	Completed college	Other
Discharged	Home	Other
Rural-Urban Continuum	Urban	Other
Complications	Other	None
# of Inpatient Admissions	>1	Other
# of ER Visits	>1	Other
Systemic Therapy	Yes	No
Radiation Therapy	Yes	No

TABLE 2: Encoding for multivariable logistic regression and KNN binary classification

KNN, K-Nearest Neighbor

Random Forest

Random Forest is a supervised learning algorithm that builds multiple decision trees and combines their outputs to make more accurate and stable predictions. It is especially useful when working with datasets that have many features and potential nonlinear relationships, as it can handle both categorical and continuous variables well. This approach reduces overfitting and captures complex, nonlinear interactions. It was chosen for its robust performance on tabular data, ability to handle mixed feature types, and resilience to outliers and class imbalance, all of which are characteristics of this dataset.

Additionally, Random Forest offers variable importance scores, which aid interpretability. This model was implemented using scikit-learn (v1.2) in Python.

Gradient Boosted Trees

GBTs iteratively build an ensemble of weak learners, typically shallow trees, with each new tree correcting the errors of its predecessors. By minimizing a differentiable loss function using gradient descent, GBTs can capture subtle patterns and interactions in the data. This model was included due to its strong predictive accuracy in clinical applications and its capacity to model complex, nonlinear relationships. GBTs often outperform other models in structured datasets like ours. This model was implemented using XGBoost (v1.7) in Python.

Support Vector Machine

The SVM classifier was implemented to predict FT as a binary outcome. SVM is a powerful supervised learning model that identifies the optimal hyperplane that maximizes the margin between classes in a high-dimensional feature space. For data that is not linearly separable, SVM leverages kernel functions - such as the radial basis function - to map inputs into a higher-dimensional space where a separating hyperplane can be found. This model was chosen for its ability to handle high-dimensional feature sets and its robustness to overfitting, especially in scenarios with complex but non-linear relationships between predictors. To prepare the data, categorical features were one-hot encoded and numeric features standardized using a preprocessing pipeline built with ColumnTransformer. Model development and evaluation were carried out using scikit-learn (v1.2).

Naive Bayes

Naïve Bayes is a probabilistic classifier based on Bayes' Theorem, assuming feature independence within each class. Despite this assumption, it often performs well in high-dimensional, low-sample-size settings. Naïve Bayes was chosen for its computational efficiency, ability to work well with categorical variables, and strong performance under conditional independence assumptions. It serves as a fast and lightweight benchmark against more complex models. This model was implemented using scikit-learn (v1.2).

Neural Networks

Artificial Neural Networks consist of interconnected nodes organized in layers, enabling them to model complex, nonlinear relationships. Each node applies a weighted transformation followed by a non-linear activation function. Neural Networks were considered in this study due to their ability to capture deeper, more complex relationships between variables. The implemented architecture consisted of a feedforward network with two hidden layers (128 and 64 neurons, respectively) using ReLU activation, followed by an output layer with softmax activation for classification. Although Artificial Neural Networks typically require larger datasets, this shallow architecture was intentionally designed with L2 regularization (α = 0.001) and early stopping to minimize overfitting. The model was implemented using scikit-learn's MLPClassifier (v1.2.2), with training optimized via the Adam algorithm and adaptive learning rate.

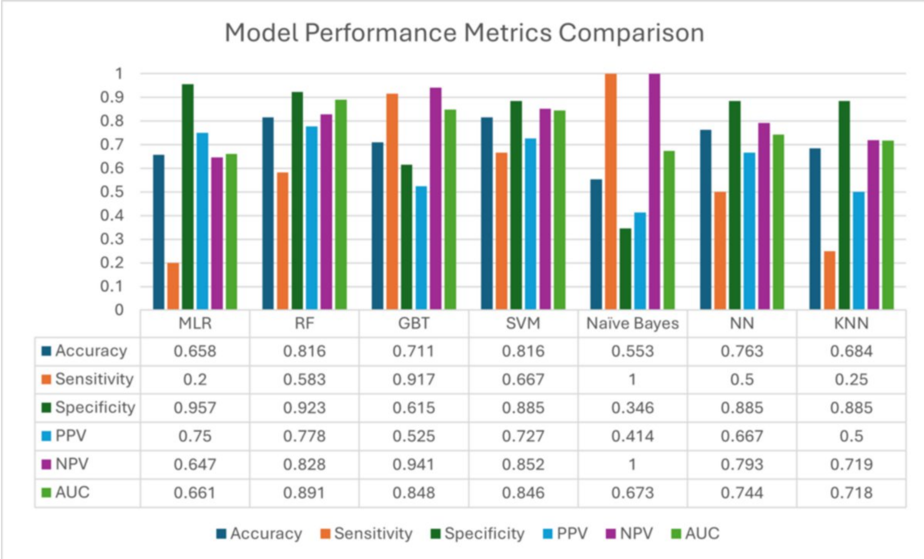
K-Nearest Neighbor

KNN classification is a supervised ML algorithm that predicts the class of a data point based on the majority class of its k nearest neighbors in the feature space. Unlike parametric models, KNN makes no assumptions about data distribution, making it flexible for nonlinear relationships. KNN was selected as one of the comparative models in this study because it is interpretable, simple to implement, and effective for small-to-medium datasets. Similar to the approach used in multivariable logistic regression, non-numeric predictors were converted into binary format as shown in Table *2*. The KNN model was able to capture local patterns in the data, which was valuable for examining the nonlinear interactions between socioeconomic, clinical, and demographic factors associated with FT. Its performance also serves as a baseline to compare against more complex regression models like random forests and GBTs.

Results

In this study, a total of seven different ML models were used to predict financial toxicity: Multivariable Logistic Regression, Random Forest, GBTs, SVM, Naïve Bayes, Neural Networks, and KNN Classification. This selection was intentional to ensure a comprehensive evaluation and comparative analysis of predictive models. It captures diverse modeling approaches: linear discriminative modeling (Multivariable Logistic Regression) for interpretability, tree-based methods (Random Forest, GBTs) for handling nonlinear feature interactions, distance-based modeling (KNN, SVM with kernel functions) for similarity-driven classification, probabilistic modeling (Naïve Bayes) for baseline performance and deep learning (Neural Networks) to assess complex pattern recognition.

Figure 1 summarizes the performance metrics for each of the predictive models evaluated in this study. Corresponding ROC curves are presented in Figure 2.



	MLR	RF	GBT	SVM	Naïve Bayes	NN	KNN
Accuracy	0.658	0.816	0.711	0.816	0.553	0.763	0.684
Sensitivity	0.2	0.583	0.917	0.667	1	0.5	0.25
Specificity	0.957	0.923	0.615	0.885	0.346	0.885	0.885
PPV	0.75	0.778	0.525	0.727	0.414	0.667	0.5
NPV	0.647	0.828	0.941	0.852	1	0.793	0.719
AUC	0.661	0.891	0.848	0.846	0.673	0.744	0.718

FIGURE 1: Comparative performance of classification models across multiple evaluation metrics

AUC, Area Under the Curve; GBT, Gradient Boosted Trees; KNN, K-Nearest Neighbors; MLR, Multivariable Logistic Regression; NN, Neural Network; NPV, Negative Predictive Value; PPV, Positive Predictive Value; RF, Random Forest; SVM, Support Vector Machine

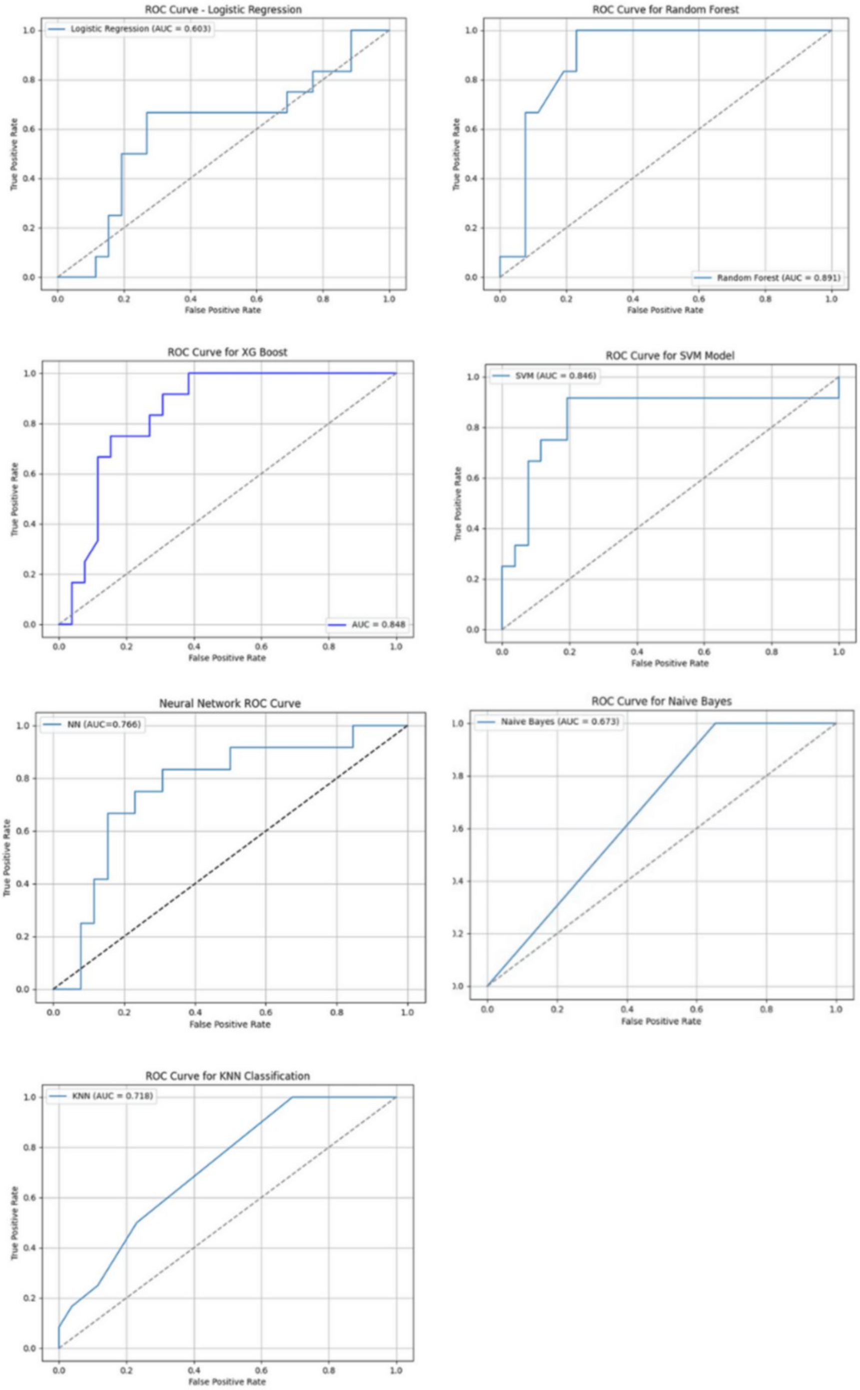


FIGURE 2: ROC curves for all seven models, demonstrating the superior AUC performance of the Random Forest classifier compared to other models

AUC, Area Under the Curve; KNN, K-Nearest Neighbors; ROC, Receiver Operating Characteristic; SVM, Support Vector Machine; XGBoost, eXtreme Gradient Boosting

Multivariable Logistic Regression relies on a single linear decision boundary, which limits its ability to capture complex, non-linear interactions among predictors. As a result, the model tends to default to the majority class, yielding very high specificity (0.957) but low sensitivity (0.200). This makes it poorly suited for identifying true positive cases unless they fall cleanly on the positive side of the linear boundary. These limitations are reflected in its lower AUC (0.661) and modest accuracy (0.658).

Random Forest combines multiple decision trees built on bootstrapped samples, enabling it to learn flexible, non-linear relationships. The ensemble approach helps smooth out overfitting and enhances robustness to both outliers and class imbalance. As a result, Random Forest consistently demonstrated strong performance across all metrics, achieving the highest AUC (0.891), high specificity (0.923), solid sensitivity (0.583), and joint-highest accuracy (0.816).

GBT corrects errors made by previous trees through iterative updates. This leads to a strong emphasis on identifying positive cases, reflected in very high sensitivity (0.917). However, this comes at the cost of increased false positives, resulting in lower specificity (0.615) and a reduced accuracy (0.711). The model's aggressive focus on the minority class improves recall but undermines overall balance.

SVM, especially with a radial basis function kernel, constructs flexible decision boundaries while controlling for overfitting via margin maximization. By tuning the penalty and kernel width, the model achieved a balanced performance: moderately high sensitivity (0.667), specificity (0.885), and strong AUC (0.846). However, unlike Random Forest, the SVM applies a global kernel function, limiting its ability to adapt locally across different regions of the feature space.

Naïve Bayes assumes conditional independence between predictors - an assumption rarely valid in clinical datasets. This leads to overly confident probability estimates, pushing many cases to extremes. With a threshold of 0.5, the model correctly classified all true positives (sensitivity = 1.000) but generated a relatively large number of false positives, reflected in low specificity (0.346). Consequently, Naïve Bayes recorded one of the lowest accuracies (0.553) and a modest AUC (0.673).

Neural Networks were implemented with a relatively simple architecture to prevent overfitting on the small dataset. While this simplicity limited the model's ability to learn deeper feature interactions, it still achieved moderate sensitivity (0.500) and specificity (0.885) with a respectable AUC of 0.744. Although it did not perform as well, likely due to the relatively small sample size and mix of variable types, it was still useful for comparison and helped confirm that simpler, tree-based models were more interpretable and effective for predicting financial toxicity in this case.

KNN classification model was evaluated, yielding an accuracy of 0.684 and an AUC-ROC of 0.718, indicating moderate discriminatory capability. High specificity (0.885) was observed, suggesting that the model was effective in correctly identifying patients without financial toxicity. However, sensitivity was low (0.25) because it tended to favor the majority class, correctly identifying patients without financial toxicity but missing many who experienced it. The positive predictive value was 50%, meaning that only half of the predicted positive cases were correct, while the negative predictive value was 71.9%, indicating better reliability in ruling out financial toxicity. These results suggest that the model may have been influenced by class imbalance or overlapping feature distributions between groups. While the KNN model showed some strengths, particularly in specificity, it underperformed compared to the random forest model. This was largely due to its reliance on distance-based calculations, which make it vulnerable to noise and outliers. Its inability to learn patterns from the data limited its capacity to capture nuanced relationships, resulting in comparatively weaker performance.

Random Forest posted the highest AUC and joint-highest accuracy across the full spectrum of thresholds. Unlike GBTs and Naïve Bayes, it does not sacrifice specificity for sensitivity, offering more stable clinical operating points. In addition, Random Forest is robust to class imbalance and outliers, and its built-in variable importance scores provide interpretability and actionable insights that single-kernel SVMs do not readily supply.

Taken together, these advantages justify the adoption of Random Forest as the primary predictive model for this study. The following section outlines how this model may be further leveraged in practical applications.

Discussion

Best predictive model analysis

Among the ML models investigated, Random Forest emerged as the best model that demonstrated strong performance in predicting FT. As shown in the confusion matrix (Figure *3*), the model correctly identified 24 true negatives (patients without FT) and 7 true positives (patients with FT), while misclassifying only 5 false negatives and 2 false positives. These results translate to high precision and recall, indicating the model's robust discriminatory power.

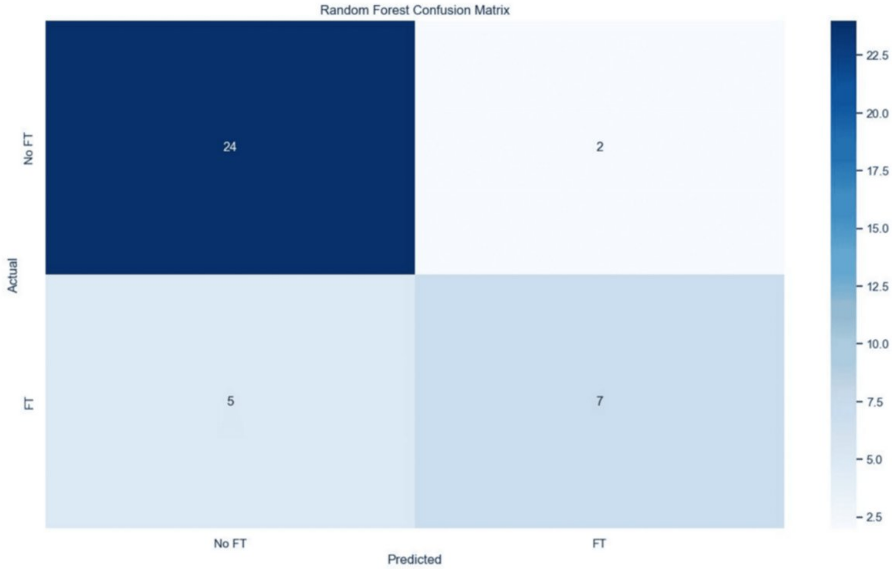


FIGURE 3: Random forest confusion matrix

FT, Financial Toxicity

The learning curves in Figure 4 further validate the model's reliability. Early trees contributed significantly to performance, with the model reaching optimal accuracy with approximately 60-80 trees. Both training accuracy and validation accuracy (out-of-bag score) converged to high, stable values as the number of trees increased. This convergence suggests the model achieved optimal generalization without overfitting. The minimal gap between training and out-of-bag scores confirms the model's consistency, and the plateau in performance beyond 60 trees indicates computational efficiency. Together, these results highlight Random Forest as the best-fitting model due to its balance of accuracy, interpretability, and robustness to overfitting, making it well-suited for predicting financial toxicity.

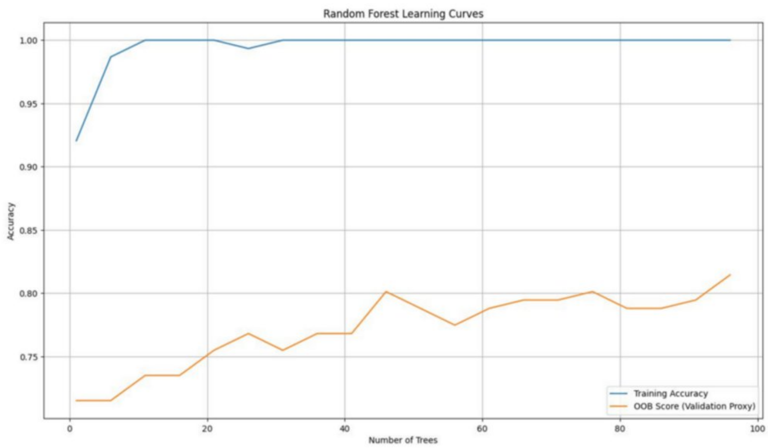


FIGURE 4: Random forest learning curve (Training vs. OOB score)

OOB, Out-of-Bag

Identifying the most influential predictors

The Random Forest model identified key variables that significantly influenced FT among patients undergoing GI surgery. As illustrated in the predictor importance plot in Figure 5, FT is influenced by a combination of clinical, socioeconomic, and psychosocial factors. The most significant predictors were Social Well-Being (SWB) Score and Emotional Well-Being (EWB) Score, highlighting the critical role of mental and social health in financial distress. This observation further validates a key finding presented by Young et al. [10]. Additionally, Distance to Cancer Center, Age, and Area Deprivation Index were highly influential, suggesting that accessibility and neighborhood disadvantage exacerbate financial hardship.

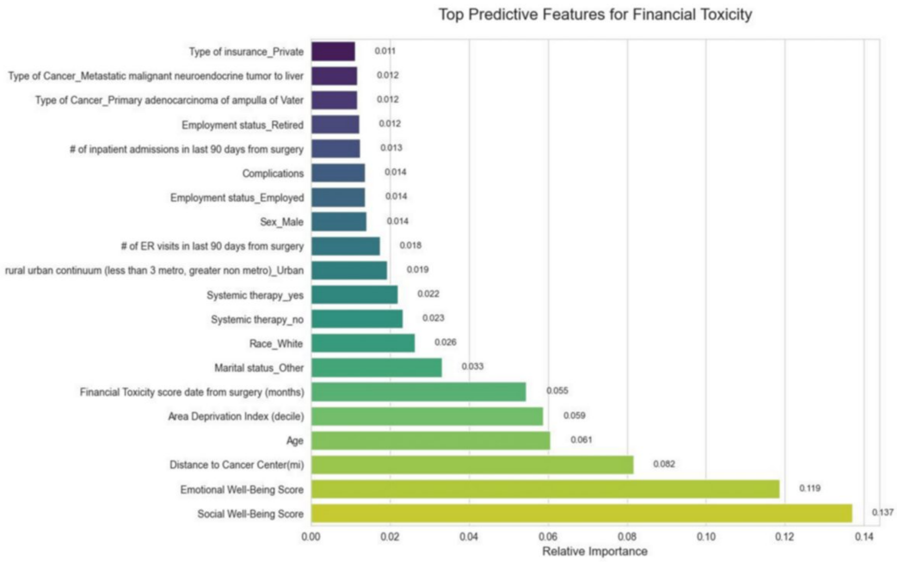


FIGURE 5: Predictor importance from trained random forest

Demographic factors such as Marital Status, Race, Type of Insurance, and Employment Status, along with treatment-related factors like receipt of Systemic Therapy and the presence of Complications, also emerged as relevant predictors of FT. Notably, certain parameters, such as Type of Cancer and Education Level, showed minimal impact on FT. These findings highlight that FT is driven more by systemic and psychosocial stressors than by disease pathology alone.

The data were narrowed down to only patients exhibiting financial hardship (FT score ≤26) to specifically analyze correlations between the top predictors and FT severity. Based on the predictor importance plot from the Random Forest model in Figure *5*, the top nine predictors of FT were selected for further correlation analysis. These included SWB Score, EWB Score, Age, Distance to Cancer Center, Area Deprivation Index, Marital Status, Race, Employment Status, and Type of Insurance. The correlation heatmap as shown in Figure *6* reveals the strength and direction of the top predictors' relationships with FT.

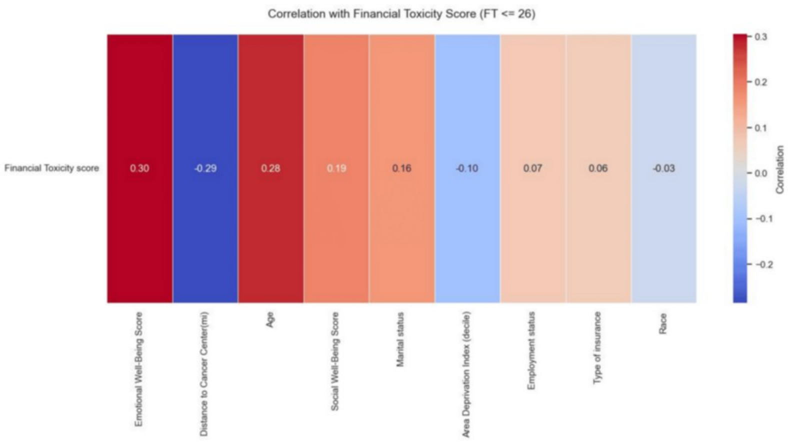


FIGURE 6: Correlation of key variables with financial toxicity score

According to Figure 6, EWB (0.30) showed the strongest positive correlation. In contrary, distance to treatment center (-0.29) demonstrated a strong negative correlation, suggesting patients who live farther from the cancer center experienced greater financial toxicity. This likely reflects added costs of travel, lodging, missed work, or logistical burdens that strain finances. Age (0.28) and SWB (0.19) showed moderate positive correlations, while marital status (0.16) had a weaker positive association. The weakest correlations were observed for employment status (0.07), insurance type (0.06), and race (-0.03).

Statistical analysis

Table 3 provides a summary of the statistical analysis conducted on the top predictive features related to FT. Among the patients included in this study, the mean SWB score was 22.25, with a median of 24, suggesting a slight left-skew in the distribution, as more patients reported relatively higher SWB. The EWB score had a mean of 19.71 and a median of 21, indicating that while the average EWB was somewhat lower, many patients still reported moderate to high emotional health. The average distance to the cancer center was approximately 20.02 miles, with a median of 18.4 miles, highlighting that most patients traveled a considerable distance for treatment.

Feature	Mean	Median
Social Well-Being Score	22.253	24
Emotional Well-Being Score	19.712	21
Distance to Cancer Center (mi)	20.015	18.4
Age	67.683	69
Area Deprivation Index (decile)	5.603	6

TABLE 3: Statistical analysis of top predictors

The mean age of participants was 67.68 years, and the median age was 69 years, indicating that the cohort largely consisted of older adults. The Area Deprivation Index had a mean of 5.60 and a median of 6, suggesting that patients were generally distributed across moderately deprived areas on the national scale, where higher deciles indicate greater deprivation. Out of the 188 patients included in the study, 104 (55.3%) were married, 106 (56.4%) were employed, and 167 (88.8%) identified as white.

The mean FT score was 31, with a standard deviation of 9.2. Using the FACIT-COST cutoff of <26, 59 patients (31.4%) were considered to have FT. Out of those, 51 (27.1%) had mild FT and 8 (4.2%) had moderate FT. None of the patients in this group experienced severe financial toxicity, suggesting that while FT was relatively common, it was mostly on the milder end of the spectrum. These findings agree with - and provide additional insight to - those reported by Young et al. [10].

Conclusions

FT in oncology can have major physical and mental impacts on cancer patients and their caregivers. The ability to accurately predict FT can help cancer patients, their family, and medical professionals develop personalized treatment plans to achieve the best possible healthcare outcomes by making informed decisions. Despite the prevalence of several GI cancers, such as colorectal cancer, predicting FT in the context of complex GI surgery is a relatively less studied area. Advances in ML techniques offer an opportunity for researchers to significantly improve FT predictions in the context of complex GI surgery. The research described in this paper is intended to significantly expand on the seminal paper on FT prediction in the context of complex GI surgery. We developed and evaluated seven ML models on multiple metrics and performed predictor importance analysis. Specifically, among the predictive models evaluated, the Random Forest classifier achieved the strongest overall performance with the highest accuracy, AUC, sensitivity, and specificity across cross-validation folds. Statistical analysis demonstrated that the model reliably identified patients at risk of FT with low misclassification rates and high classification precision. In addition to its predictive strength, the Random Forest model offered insights into variable importance, identifying SWB, EWB, Distance to Cancer Center, Age, and Area Deprivation as the most influential predictors. These findings not only support the robustness of the model but also validate the role of psychosocial and socioeconomic factors in shaping financial outcomes. Compared to the other six machine learning models, Random Forest provided the best balance between predictive performance and clinical interpretability, making it well-suited for future integration into preoperative screening workflows.

A limitation of the study is the relatively small dataset, and that all data were collected from a single source. Future studies should focus on expanding and diversifying the dataset to improve model generalizability. Increasing the sample size and incorporating more detailed demographic variables - such as specific racial and ethnic identities, household income, number of financial dependents, and history of financial crises, would enhance the predictive power and fairness of the models. To address class imbalance and improve model sensitivity, data augmentation and resampling techniques like Synthetic Minority Over-sampling Technique could be employed to process the data, particularly for underrepresented outcome groups. Additionally, implementing k-fold cross-validation in future modeling pipelines would provide a more rigorous assessment of model stability and reduce the risk of overfitting. Given that individual ML models can perform somewhat differently and potentially can “see” different aspects of the same dataset, an ensemble approach could be taken. While a random forest can be considered an ensemble of decision trees, a more heterogeneous ensemble made up of very different classifiers can lead to positive outcomes. This is because relative weaknesses of some classifiers can potentially be compensated for by other (very) different classifiers. These enhancements would support the development of more accurate, equitable, and clinically useful models for identifying patients at risk of FT.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: A C Fong, Ruth Bahre, Alex Masterson, Gitonga Munene, Zachary D. Asher, Sahar Abedzadeh

Drafting of the manuscript: A C Fong, Ruth Bahre, Alex Masterson

Critical review of the manuscript for important intellectual content: A C Fong, Ruth Bahre, Alex Masterson, Gitonga Munene, Zachary D. Asher, Sahar Abedzadeh

Supervision: A C Fong, Sahar Abedzadeh

Acquisition, analysis, or interpretation of data: Ruth Bahre, Alex Masterson, Gitonga Munene

Disclosures

Human subjects: Consent was obtained or waived by all participants in this study. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Coroneos CJ, Lin YL, Sidey-Gibbons C, et al.: Correlation between financial toxicity, quality of life, and patient satisfaction in an insured population of breast cancer surgical patients: A single-institution retrospective study. Journal of the American College of Surgeons. 2021, 232:253-263. 10.1016/j.jamcollsurg.2020.10.023
2. Tran TB, Malhotra G, Razavi M, et al.: Emotional distress and financial toxicity in younger adult patients undergoing oncologic surgery. Annals of Surgery. 2022, 276:694-700. 10.1097/sla.0000000000005593
3. Delgado-Guay M, Ferrer J, Rieber AG, et al.: Financial distress and its associations with physical and emotional symptoms and quality of life among advanced cancer patients. The Oncologist. 2015, 20:1092-1098. 10.1634/theoncologist.2015-0026
4. Global Cancer Observatory, International Agency for Research on Cancer. https://gco.iarc.fr/.
5. World Health Organization. Cancer Fact Sheet. (2025). https://www.who.int/news-room/fact-sheets/detail/cancer.
6. Scott JW, Knowlton LM, Murphy P, et al.: Financial toxicity after trauma and acute care surgery: From understanding to action. Journal of Trauma and Acute Care Surgery. 2023, 95:800-805. 10.1097/ta.0000000000003979
7. LaRocca CJ, Li A, Lafaro K, et al.: The impact of financial toxicity in gastrointestinal cancer patients. Surgery. 2020, 168:167-172. 10.1016/j.surg.2020.02.006
8. Arastu A, Patel A, Mohile SG, et al.: Assessment of financial toxicity among older adults with advanced cancer. JAMA Network Open. 2020, 3:e2025810. 10.1001/jamanetworkopen.2020.25810
9. Bouberhan S, Shea M, Kennedy A, et al.: Financial toxicity in gynecologic oncology. Gynecologic Oncology. 2019, 154:8-12. 10.1016/j.ygyno.2019.04.003
10. Young L, Vergara R, Henriquez J, Fong A, Al-Assil T, Shebrain S, Munene G: Financial toxicity in complex gastrointestinal surgery and correlation with patient reported outcomes. Annals of Surgery. 10.1097/SLA.0000000000006559
11. IBM SPSS 27. https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-27.
12. de Souza JA, Yap BJ, Wroblewski K, et al.: Measuring financial toxicity as a clinically relevant patient-reported outcome: The validation of the COmprehensive Score for financial Toxicity (COST). Cancer. 2017, 123:476-484. 10.1002/cncr.30369