

Using machine learning models to predict synchronous genitourinary cancers among gastrointestinal stromal tumor patients

Mohammad Alghafees, Raouf M Seyam, Turki Al-Hussain¹, Tarek Mahmoud Amin², Waleed Altaweel, Belal Nedal Sabbah³, Ahmad Nedal Sabbah³, Razan Almesned, Laila Alessa

Departments of Urology and ¹Pathology and Laboratory Medicine, King Faisal Specialist Hospital and Research Center, ²Department of Surgical Oncology, Oncology Center, King Faisal Specialist Hospital and Research Center, ³College of Medicine, Alfaisal University, Riyadh, Saudi Arabia

Abstract

Objectives: Gastrointestinal stromal tumors (GISTs) can occur synchronously with other neoplasms, including the genitourinary (GU) system. Machine learning (ML) may be a valuable tool in predicting synchronous GU tumors in GIST patients, and thus improving prognosis. This study aims to evaluate the use of ML algorithms to predict synchronous GU tumors among GIST patients in a specialist research center in Saudi Arabia.

Materials and Methods: We analyzed data from all patients with histopathologically confirmed GIST at our facility from 2003 to 2020. Patient files were reviewed for the presence of renal cell carcinoma, adrenal tumors, or other GU cancers. Three supervised ML algorithms were used: logistic regression, XGBoost Regressor, and random forests (RFs). A set of variables, including independent attributes, was entered into the models.

Results: A total of 170 patients were included in the study, with 58.8% ($n = 100$) being male. The median age was 57 (range: 9–91) years. The majority of GISTs were gastric (60%, $n = 102$) with a spindle cell histology. The most common stage at diagnosis was T2 (27.6%, $n = 47$) and N0 (20%, $n = 34$). Six patients (3.5%) had synchronous GU tumors. The RF model achieved the highest accuracy with 97.1%.

Conclusion: Our study suggests that the RF model is an effective tool for predicting synchronous GU tumors in GIST patients. Larger multicenter studies, utilizing more powerful algorithms such as deep learning and other artificial intelligence subsets, are necessary to further refine and improve these predictions.

Keywords: Artificial intelligence, gastrointestinal oncology, gastrointestinal stromal tumors, genitourinary oncology, urologic oncology, urology

Address for correspondence: Dr. Belal Nedal Sabbah, Takhassusi Street, Alfaisal University, Riyadh 11533, Saudi Arabia.

E-mail: bsabbah@alfaisal.edu

Received: 07.03.2023, **Revised:** 24.08.2023, **Accepted:** 13.10.2023, **Published:** 25.01.2024.

INTRODUCTION

Gastrointestinal stromal tumors (GISTs) are a rare type of mesenchymal tumor that commonly develop in the gastrointestinal tract. In fact, GISTs are the most

frequently occurring mesenchymal tumor in this anatomical region.^[1] GISTs are known to have several distinct molecular subtypes, including those with mutations in KIT or PDGFRa. Detecting these molecular alterations at an

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Alghafees M, Seyam RM, Al-Hussain T, Amin TM, Altaweel W, Sabbah BN, *et al.* Using machine learning models to predict synchronous genitourinary cancers among gastrointestinal stromal tumor patients. *Urol Ann* 2024;16:94-7.

Access this article online	
Quick Response Code:	Website: www.urologyannals.com
	DOI: 10.4103/ua.ua_32_23

early stage is critical as it can significantly impact the choice of adjuvant and metastatic treatments.^[2] Existing literature suggests that GISTs have a nearly equal distribution between genders, with a higher incidence among individuals over the age of 60. Furthermore, GISTs tend to present with symptoms, indicating a symptomatic nature of the disease.^[3] Studies conducted in Saudi Arabia have shown that GISTs are predominantly located in the stomach and have a higher incidence in males over the age of 40 years.^[4]

Although GISTs primarily occur in the stomach and intestine, some patients may experience lower urinary tract symptoms that suggest synchronous genitourinary (GU) tumors. In addition, extragastrointestinal stromal tumors of the urinary bladder wall have been observed in rare cases.^[5] Currently, an accurate diagnosis of GISTs requires extensive imaging studies, pathological examination, and immunohistochemical analysis.^[6] Early diagnosis is imperative to achieve high rates of disease-free survival, yet the extensive testing required for a diagnosis takes substantial time.^[7] Therefore, implementing technology that predicts a concomitant tumor of other organs among GIST patients could significantly impact the overall prognosis of this condition.

Recent research suggests that using artificial intelligence (AI) and deep learning algorithms may provide more accurate confirmation of the malignant potential of GISTs.^[8] The implementation of machine learning (ML) techniques, including supervised learning algorithms, has shown promising results in improving the accuracy of predictions for various medical conditions.

In this study, we aim to utilize ML to predict GU cancer in GIST patients, with a particular focus on the Saudi Arabian population. By utilizing a large dataset of patients diagnosed with GIST from our specialist research center between 2003 and 2020, we aim to determine the accuracy and effectiveness of three supervised ML algorithms: logistic regression, XGBoost Regressor, and random forests (RF). The identification of predictive variables and the accuracy of these models will provide valuable insight into the potential for AI and ML to improve the diagnosis and management of GIST patients, particularly in the context of having a synchronous GU neoplasm.

MATERIALS AND METHODS

This retrospective study included all patients with a histopathological diagnosis of GIST at King Faisal Specialist Hospital and Research Centre between 2003 and 2020. Any concomitant GU cancer was identified. Data were analyzed using SPSS v26 (IBM, New York, United

States). Continuous data are presented as averages with standard deviations, while categorical data are depicted using absolute numbers and percentages.

Four types of AI algorithms were employed in this study to predict the presence of GU cancer in the presence of GIST. These include RF, XGBoost classifier, CatBoost classifier, and support vector machine. After running a baseline prediction model, some variables were dropped because they were not significant to the prediction of the model. The ML models were fitted using scikit-learn 0.18 modules of Python throughout this study. The data set was randomly divided into 80% of the training set and 20% of the test set at 8:2 (136: 34). The target variable was encoded in a binary format with 1 (presence of GU cancer) and 0 (absence of GU cancer). The RF model is a decision tree-based ML model. Each node of the decision tree divides the data into two groups using a cutoff value inside one of the features. By building an ensemble of randomized decision trees, each of which overfits the data and averages the results to obtain a better classification, the RF technique can reduce the effect of the overfitting problem.

This retrospective chart review study involving human participants followed the standards of the 1964 Helsinki Declaration and its later amendments. This study is a secondary analysis of datasets from an already approved study by the Human Investigation Committee (IRB) and Research Ethics Committee of King Faisal Specialist Hospital and Research Center.

RESULTS

A total of 170 GIST patients were detected. As shown in Table 1, most of the patients (58.8%; $n = 100$) were males. The median age was 57 (9–91) years. The majority of the GISTs were gastric (60%; $n = 102$) with spindle cell histology. The most common stage at diagnosis is T2 (27.6%; $n = 47$) and N0 (20%; $n = 34$). Six patients (3.5%) had synchronous GU tumors. Of them, three patients had renal cell carcinomas (RCC). Two were histologically diagnosed to have clear cell RCC and one with only a radiological diagnosis of RCC. Three other patients had adrenal tumors (one adrenal carcinoma, one isolated adrenal GIST, and one pheochromocytoma).

After all modes of hyperparameter tuning were done to the model, the RF model achieved the highest accuracy with 97.1%. It predicted that based on the input variables and patient characteristics, 97.1% still did not have associated GU cancer and that only 2.9% of those who had GIST had associated GU cancer. On more analysis to ascertain the specificity of the model, Figure 1 shows the confusion

matrix for the RF models which explains the specificity of the model in terms of how true the predicted values are accurate to the original values. It showed that out of a random 34 number of patients, the model predicts 32 patients to be GU cancer-free even in the presence of GIST and only 1 patient to have associated GU cancer in the presence of GIST.

Figure 2 shows the feature importance of each variable column used for the RF model which is the one with the best prediction accuracy. It is evident that variables in the

Table 1: Demographic and tumor-related characteristics of patients (n=170)

Continuous variables	n	Median (range)
Age at diagnosis (years)	170	57 (9-91)
GIST size (cm)	161	6 (0.3-36)
Categorical variables	n (%)	
Gender		
Male	100 (58.8)	
Female	70 (41.2)	
GIST primary site		
Gastric	102 (60)	
Small intestine	47 (27.6)	
Omentum/peritoneum/mesenteric	12 (7.1)	
Other	9 (5.3)	
GIST TNM stage		
T1	25 (14.7)	
T2	47 (27.6)	
T3	44 (25.9)	
T4	45 (26.5)	
N0	34 (20.0)	
N1	5 (2.9)	
M0	13 (7.6)	
M1	25 (14.7)	
Histopathological subtype		
Spindle cell	85 (50.0)	
Epithelioid type	16 (9.4)	
Mixed epithelioid and spindle	10 (5.9)	
Other	2 (1.2)	

GIST: Gastrointestinal stromal tumor, TNM: Tumor, node, and metastasis

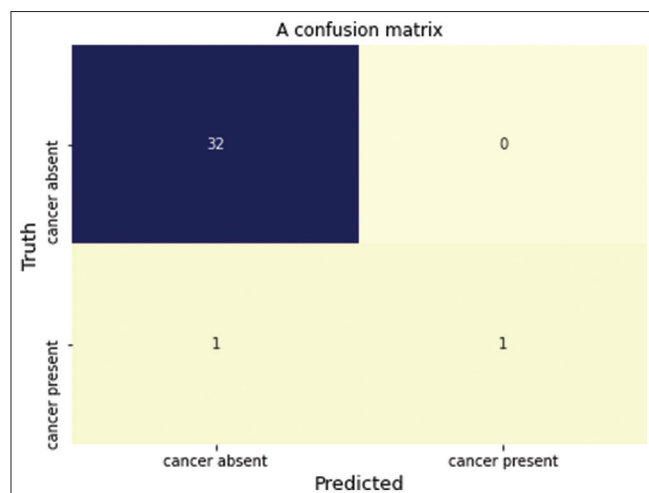


Figure 1: Confusion Matrix

index 5, 3, and 6 contributed more in the prediction. These variables were associated with cancer taking the highest, gender, and site of GIST, respectively. Therefore, even with the presence of GIST-associated cancer, there is a rare correlation between GIST and GU cancer.

DISCUSSION

The study’s findings demonstrate the potential of AI technology to accurately predict having synchronous GU cancer among GIST patients, as evidenced by the RF model’s 97.1% accuracy. The patient population analyzed was mostly male. Only a small portion of patients had an accompanying GU cancer, at <5%. The diagnoses for these patients included RCC, adrenal carcinoma, adrenal GIST, and pheochromocytoma.

Our study’s findings are consistent with existing literature regarding patient demographics and disease characteristics, showing that GISTs are predominantly located in the stomach (61%). The reported age of onset varies across studies, with median diagnosis age ranging from 50 to 60 years.^[9,10] However, a study conducted in Saudi Arabia reported a lower mean age at diagnosis of 40 years, which is substantially lower than the median age reported in other studies.^[4] Thus, our study’s results indicate that the age of onset of GIST in our cohort is higher than what has been reported in other studies conducted in Saudi Arabia. This difference may be due to various factors, including differences in sample sizes, selection criteria, and genetic and environmental factors. However, further studies are required to confirm this observation.

This study represents an initial attempt to utilize ML algorithms to predict the presence of GU tumors in GIST patients. However, ML models have recently been the subject of numerous research studies across various cancer types, including ovarian, thyroid, and breast cancer.^[11-13] These studies demonstrate the potential of ML

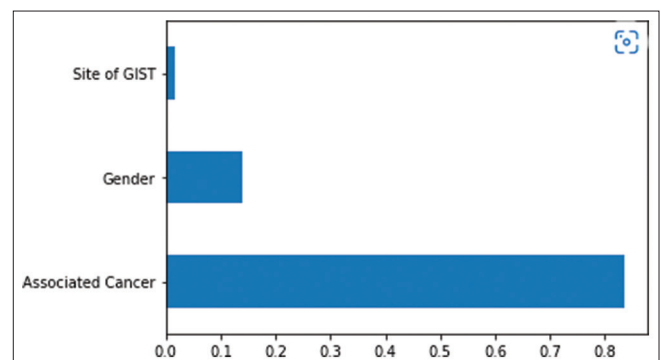


Figure 2: Most Effective Parameters

in predicting disease outcomes and identifying biomarkers for early diagnosis. Toth *et al.* demonstrated the successful use of the RF model in clinical practice for the detection of biomarkers for prostate cancer progression. Their study utilized an RF-based classification model to predict the aggressive behavior of prostate cancer, achieving an accuracy of 95%. The application of the RF model in their study allowed for the identification of a set of biomarkers that could predict the likelihood of disease progression and guide clinical decision-making.^[14] The high accuracy of the RF model in predicting prostate cancer behavior suggests its potential for use in other cancer types, including the prediction of GU cancers in GIST patients as demonstrated in our study. These findings support AI as an externally valid classification model to support the clinical management of prostate cancer.^[14] Another study by Xiao *et al.* reported on similar outcomes predicting the occurrence of prostate cancer using the RF algorithm. Here, transrectal ultrasound findings, age, and serum levels of prostate-specific antigen were taken into account, yielding a predictive accuracy of 83.10%. The results of this study permitted the statement that the adoption of an RF model and AI technology demonstrates superior diagnostic performance than individual diagnostic indicators alone.^[15] This is supported by the findings of the present study.

Limitations

There are several limitations worth noting in this study. First, we did not include all potential predictive factors for having GU cancer in GIST patients, such as family history of malignancy and exposure to risk factors. Second, this study was conducted at a single center, which may limit the generalizability of our results to other populations. Third, there are currently no other studies in the literature that explore the use of ML to predict synchronous GU tumors and GISTs, which makes it difficult to compare and validate our findings. Fourth, our model accuracy needs to be tested and validated by an external series in the future. Future research should aim to address these limitations by exploring whether incorporating additional predictive factors into the RF model can improve its accuracy.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

- Rubin BP, Heinrich MC, Corless CL. Gastrointestinal stromal tumour. *Lancet* 2007;369:1731-41.
- Kelly CM, Gutierrez Sainz L, Chi P. The management of metastatic GIST: Current standard and investigational therapeutics. *J Hematol Oncol* 2021;14:2.
- Søreide K, Sandvik OM, Søreide JA, Giljaca V, Jureckova A, Bulusu VR. Global epidemiology of gastrointestinal stromal tumours (GIST): A systematic review of population-based cohort studies. *Cancer Epidemiol* 2016;40:39-46.
- Bokhary RY, Al-Maghrabi JA. Gastrointestinal stromal tumors in Western Saudi Arabia. *Saudi Med J* 2010;31:437-41.
- Mekni A, Chelly I, Azzouz H, Ben Ghorbel I, Bellil S, Haouet S, *et al.* Extragastric stromal tumor of the urinary wall bladder: Case report and review of the literature. *Pathologica* 2008;100:173-5.
- Yadav SC, Menon S, Bakshi G, Katdare A, Ramadwar M, Desai SB. Gastrointestinal stromal tumor presenting with lower urinary tract symptoms – A series of five cases with unusual clinical presentation. *Indian J Urol* 2021;37:357-60.
- Uzunoglu H, Tosun Y, Akinci O, Baris B. Gastrointestinal stromal tumors of the stomach: A 10-year experience of a single-center. *Niger J Clin Pract* 2021;24:1785-92.
- Seven G, Silahatoglu G, Kochan K, Ince AT, Arici DS, Senturk H. Use of artificial intelligence in the prediction of malignant potential of gastric gastrointestinal stromal tumors. *Dig Dis Sci* 2022;67:273-81.
- Joensuu H, Hohenberger P, Corless CL. Gastrointestinal stromal tumour. *Lancet* 2013;382:973-83.
- Vij M, Agrawal V, Kumar A, Pandey R. Gastrointestinal stromal tumors: A clinicopathological and immunohistochemical study of 121 cases. *Indian J Gastroenterol* 2010;29:231-6.
- Park S, Yi G. Development of gene expression-based random forest model for predicting neoadjuvant chemotherapy response in triple-negative breast cancer. *Cancers (Basel)* 2022;14:881.
- Liu YH, Jin J, Liu YJ. Machine learning-based random forest for predicting decreased quality of life in thyroid cancer patients after thyroidectomy. *Support Care Cancer* 2022;30:2507-13.
- Cheng L, Li L, Wang L, Li X, Xing H, Zhou J. A random forest classifier predicts recurrence risk in patients with ovarian cancer. *Mol Med Rep* 2018;18:3289-97.
- Toth R, Schiffmann H, Hube-Magg C, Büscheck F, Höflmayer D, Weidemann S, *et al.* Random forest-based modelling to detect biomarkers for prostate cancer progression. *Clin Epigenetics* 2019;11:148.
- Xiao LH, Chen PR, Gou ZP, Li YZ, Li M, Xiang LC, *et al.* Prostate cancer prediction using the random forest algorithm that takes into account transrectal ultrasound findings, age, and serum levels of prostate-specific antigen. *Asian J Androl* 2017;19:586-90.