



Predicting the Stage of Gastric Cancer After Gastrectomy Based on Machine Learning Algorithms

Nayereh Abdali¹, Sajad Alavimanesh¹, Mirhamid Mirsaeid Ghazi² and Sara Zareei^{3*}

¹Student Research Committee, Shahrekord University of Medical Sciences, Shahrekord, Iran

²School of Business Administration, Lakehead University, Thunder Bay, Ontario, Canada

³Department of Cell & Molecular Biology, Faculty of Biological Sciences, Kharazmi University, Tehran, Iran

Citation: Nayereh Abdali, Sajad Alavimanesh, Mirhamid Mirsaeid Ghazi, Sara Zareei (2026) Predicting the stage of gastric cancer after gastrectomy based on machine learning algorithms. *Int J. of Gastro Med & Res* 1(2), 1-9. WMJ-IJGMR-108

Abstract

Background: Gastric cancer (GC) is the fourth most common cause of cancer death worldwide, with a 5-year survival rate of less than 40%. One of the most important methods for diagnosing stomach cancer is endoscopy, which is quite costly and invasive. The aim of this study was to develop machine learning-based diagnostic prediction models for the stage of GC.

Objectives: To create a highly accurate predictive model for the stage of GC in patients via a noninvasive method based on machine learning (ML).

Methods: In this study, data from 996 patients with GC after gastrectomy were utilized. The data were split into groups, trained and tested, and a ratio of 8:2 was used to develop different machine learning models. Furthermore, the six different machine learning algorithms used in predicting the stage of GC include decision tree (DT), K nearest neighbor (KNN), logistic regression (LR), naive Bayes (NB), random forest (RF), and support vector machine (SVM) methods.

Results: The analysis of the demographic variables revealed statistically significant differences in the PLR and NLR and other parameters between the two groups of patients with stages I and III gastric cancer ($P < 0.05$). Moreover, these findings suggest that the KNN model in this study is one of the best models for predicting the stage of GC.

***Corresponding author:** Sara Zareei Department of Cell & Molecular Biology, Faculty of Biological Sciences, Kharazmi University, Tehran, Iran.

Submitted: 06.05.2026

Accepted: 12.05.2026

Published: 23.05.2026

Keywords: Gastric cancer, Machine learning, Prediction, Stage, Algorithm

Abbreviations

GC	Gastric cancer
ML	Machine Learning
DT	Decision Tree
KNN	K Nearest Neighbor
LR	Logistic Regression
NB	Naive Bayes
RF	Random forest
SVM	Support Vector Machine

Introduction

Gastric cancer (GC) is the fourth most common malignancy and the second leading cause of death worldwide. Early gastric cancer has a good prognosis; however, it presents numerous unusual and unclear clinical symptoms. More than 1 million cases of early gastric cancer are diagnosed each year. Early gastric cancer can be treated with endoscopy, whereas the only option for advanced gastric cancer is surgery. GC is usually asymptomatic in its early stages, and diagnosing GC during these stages is a challenging process. Nevertheless, early identification of patients is effective in improving their prognosis. Therefore, rapid, accurate, and reliable prediction is useful in optimizing the treatment of patients with GC (4). The 5-year survival rate for patients with GC is 90%, while this survival rate decreases to 18% for patients with stage III GC. Although the survival rate is high for surgeries performed in the early stages of the disease, patients with GC globally have an overall survival rate of 25% because most patients are diagnosed at advanced stages [1-4].

Predictive medicine is a rapidly evolving field that focuses on anticipating the onset of diseases before symptoms appear, enabling earlier and more personalized healthcare interventions. By leveraging data from genetics, lifestyle, environment, and advanced technologies such as machine learning and artificial intelligence, predictive medicine aims to identify individuals at high risk for specific conditions. This proactive approach allows healthcare providers to tailor prevention strategies, monitor high-risk patients more closely, and optimize treatment plans, ultimately improving patient outcomes and reducing healthcare costs. As the availability of large-scale health data continues to grow, predictive medicine holds significant promise for transforming the future of healthcare from reactive to preventative [5-14].

The adoption of artificial intelligence (AI)-based solutions, such as machine learning (ML), can overcome the limitations of invasive diagnostic procedures in the screening and diagnosis of GC due to their computational capabilities. ML-based techniques for developing predictive models and analyzing data rely on existing data and provide valuable insights from raw datasets. ML models can make predictions on the basis of various measurements. ML methods have been applied with high accuracy to predict several types of cancer, including breast cancer and colorectal cancer [15-20].

Therefore, the use of artificial intelligence can optimally address intractable medical problems that were previously thought to be unsolvable. Identifying patients in the early stages of GC via highly accurate machine learning techniques is an efficient, rapid, and low-cost method that can assist physicians in diagnosing the disease stage more accurately and quickly. The present study employed machine learning to predict various stages of GC in patients after surgery [21-27].

Methods and Materials**Study Design and Population**

The present study included data from 1,331 patients with GC after gastrectomy, some of whom were excluded because of missing information. Therefore, this study was conducted on 996 patients. All patients were diagnosed with stage I-III GC based on histological tests after surgery. Tumors were staged according to the TNM staging system of the 7th edition of the American Joint Cancer Commission (AJCC). In this study, the inclusion and exclusion criteria included the absence of radiation or chemotherapy in the patients, the presence of complete clinical pathology and follow-up data of potential prognostic factors, the lack of other malignant cancers coexisting with the disease, the absence of GC recurrence, the absence of residual GC, and the lack of any acute inflammation or infection two weeks before surgery [28-30].

Data, including sex, age, tumor characteristics after surgery, necessary laboratory tests before surgery, and survival time, were then collected. Blood samples were taken from patients one week before surgery. The samples were divided into two groups: highly differentiated (papillary and moderately differentiated GC) and poorly differentiated (mucinous, signet-ring cell, and undifferentiated GC). To monitor the

patients, various examinations, including physical examinations, dynamic CT examinations, complete blood counts, gastroscopy, and measurements of tumor markers in the blood serum, were performed every three months for the first two years and every six months thereafter. The date of the last follow-up was considered June 2015 or the date of death due to any cause [25-27]. Overall survival (OS) was defined as the time interval from the date of surgery to the date of the patient's last follow-up or death from any cause [31-32].

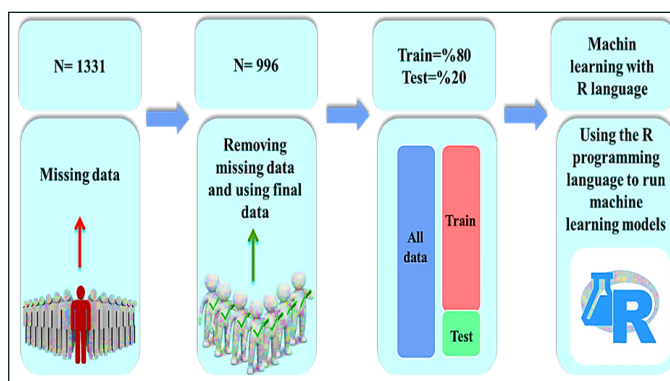


Figure 1: Overview of the steps involved in developing a GC staging prediction model

Calculations for Biomarkers

The NLR is obtained by dividing the absolute number of neutrophils by the absolute number of lymphocytes, and the PLR is obtained by dividing the absolute number of platelets by the absolute number of lymphocytes (. COP-NLR was calculated on the basis of previous studies, with a score of 0 or 1 assigned to patients with normal values and a score of 2 assigned to patients with high platelet counts ($> 300 \times 10^9 /L$) and a neutrophil–lymphocyte ratio greater than 3. Finally, the prediction of the disease stage will be made based on the data available in two phases, including the initial phase of GC, stage I, and the final phase of GC, stage III [33-34].

Machine Learning Algorithms

This study used six ML algorithms to predict the stage of GC, including DT, KNN, LR, NB, RF, and SVM. The DT algorithm is a model of classical supervised learning that uses a tree diagram structured like a flowchart. The KNN algorithm is a nonparametric, supervised learning model that utilizes the nearest neighboring individuals in the population for identification. LR employs a logistic function to model a dependent binary variable and is also a tool for

statistical measurements and regression analyses. NB is a statistical technique that applies Bayes' theorem for classification. It is one of the simplest ML algorithms, examine a feature in a class independently of other features. RF is an algorithm that uses the results of multiple DTs to achieve a specific output. SVM is a supervised ML algorithm that classifies data via an optimal line or plane [35-44].

Machine Learning Data Processing Methods

First, the data were analyzed based on descriptive statistics. A P-value < 0.05 was considered significant. The necessary analyses were performed to investigate different ML algorithms via the R programming language (V.4.3.1). For this purpose, 80% of the data were used for training, and the other 20% were used for testing to investigate the prediction of the GC stage [45].

Results

Demographic characteristics of the study population The study population included 1,331 patients with GC, of whom 996 patients were used to design the model: 220 patients in stage I (65 women and 155 men) and 776 patients in stage III (258 women and 518 men). The remaining studies were excluded from the analysis because of a lack of information. The mean age of the study population with stage I GC was 56.10 ± 12.60 years, and for stage III GC, it was 57.77 ± 11.50 years (Table 1).

The results of the comparison of the two groups in stages I and III of GC revealed that age and sex were not significant demographic criteria in the disease stage. However, the PLR, NLR, and other characteristics were significantly different between the two groups in stages I and III of GC ($P < 0.05$) (Table 1). Patients with stage III GC exhibit more advanced disease characteristics, such as larger tumor sizes, higher NLRs and PLRs, and poorer histological differentiation. Additionally, these patients had worse overall survival and higher mortality rates than patients with stage I GC did (Table 1).

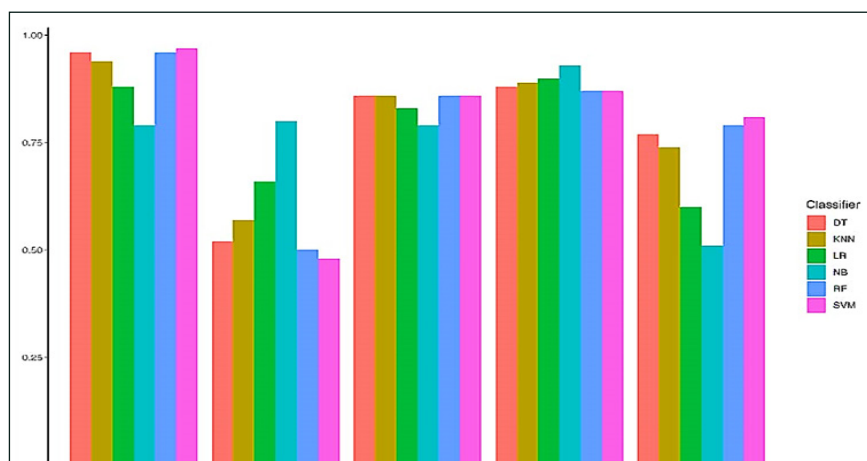
Table 1: Basic data of variables and their significance levels between two different stages of GC

	level	I	III	p
n		220	776	
Sex (%)	Female	65 (29.5)	258 (33.2)	0.34
Age (mean (SD))		56.10 (12.60)	57.77 (11.50)	0.063
Tumor location (%)	Lower third	136 (61.8)	267 (34.4)	<0.001***
	Middle third	44 (20.0)	160 (20.6)	
	Upper third	40 (18.2)	349 (45.0)	
Tumor size (mean (SD))		2.88 (1.81)	5.34 (2.70)	<0.001***
Histological grade (%)	Poorly differentiated	168 (76.4)	672 (86.6)	<0.001***
NLR (mean (SD))		2.07 (1.04)	2.75 (2.68)	<0.001***
PLR (mean (SD))		125.60 (63.55)	157.23 (87.61)	<0.001***
OS months (mean (SD))		54.39 (30.30)	34.78 (31.15)	<0.001***
Survival status (%)	0	203 (92.3)	336 (43.3)	<0.001***
COP NLR (%)	0	174 (79.1)	449 (57.9)	<0.001***
	1	35 (15.9)	270 (34.8)	
	2	11 (5.0)	57 (7.3)	

Analyzing Machine Learning

The results in Figure 2 show the performance of the six ML classifier algorithms based on the selected features. The results in Figure 2 indicate that the performance of the KNN classifier is better than that of the other five algorithms. The detection accuracy of the SVM algorithm was excellent, with a high sensitivity of 0.97. This algorithm works effectively for all stages of cancer and results in few false negatives. However, since its specificity is 0.48, it misclassifies healthy cases.

NB is another algorithm that is suitable in terms of specificity (0.80) and positive predictive value (PPV = 0.93); it is used to confirm true positives and is reliable. However, because of its sensitivity (0.79), it is not effective in detecting actual cases of cancer. Although KNN, DT, and RF achieved balanced performances, with an above-average accuracy of 0.86 and sensitivity, they presented relatively low to moderate specificity. In the trade-off of specificity to sensitivity, SVM stands the best bet; however, to reduce the incidence of false positives, the RF or DT models would be better candidates since they delivered balanced performance (Figure 2).



Classifier	Sensitivity	Specificity	Accuracy	PPV	NPV
KNN	0.94	0.57	0.86	0.89	0.74
SVM	0.97	0.48	0.86	0.87	0.81
NB	0.79	0.8	0.79	0.93	0.51
DT	0.96	0.52	0.86	0.88	0.77
RF	0.96	0.5	0.86	0.87	0.79
LR	0.88	0.66	0.83	0.9	0.6

Figure 2: Rectangular diagram of six ML algorithms, including DT, KNN, LR, NB, RF, and SVM, along with their performance comparison.

Feature Analysis

This analysis studied the feature contribution to the RF model for predicting the GC stage. The importance of a feature is generally expressed by the %IncMSE-rank: the higher, the more important a feature is when the model performance decreases significantly due to changes in the feature values [46]. The top ten percent features show, in order of importance, the most important features contributing to GC. Tumor size and survival status appeared to be the most important features. The importance of features decreased from top to bottom (Figure 3).

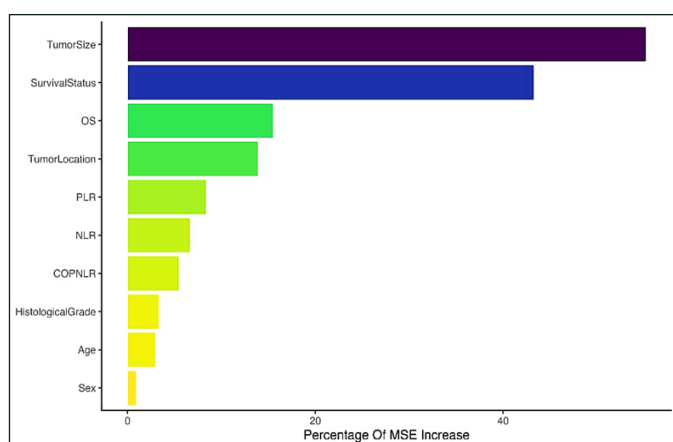


Figure 3: Important features for predicting disease stage in patients with GC after surgery, IncMSE (percentage increase in the mean squared error)

Discussion

In this study, based on the available data for diagnosing stage I and stage III GC, six different ML algorithms were used to predict the stage of GC. Compared with other ML algorithms, the KNN algorithm provided appropriate results and reported more balanced outcomes. Additionally, the stage of GC had a statistically significant relationship with various variables in this study, including tumor location, tumor size, histological grade of tissue, NLR, PLR,

OS months, survival status, and COP NLR. Among these variables, tumor size and survival status were the most important features associated with disease stage. However, just as sex and age did not have a statistically significant relationship with cancer stage, they were deemed the least important features.

GC is one of the most common and deadly tumors globally, primarily because patients with GC are often diagnosed at advanced stages of the disease. Therefore, efforts to develop noninvasive screening methods are necessary to increase patient survival and quality of life. Studies indicate that several factors affect the disease stage and prognosis of patients with GC, including tumor size and tumor location. The results of this study also emphasize this point, as tumor location and tumor size are significantly correlated with disease stage. Furthermore, as shown in Figure 3, the results indicated that the most important feature was tumor size [47-50]

Several biomarkers of the inflammatory response in cancer, including the neutrophil-to-lymphocyte ratio (NLR) and the platelet-to-lymphocyte ratio (PLR), which are crucial prognostic indicators, are important prognostic factors. PLR is a particularly significant factor in stage III/IV CRC. Compared with those of tumor markers, the levels of systemic inflammatory markers are notably increased in stage I GC patients. Consequently, inflammatory markers present in peripheral blood are vital for the early diagnosis and screening of GC. These data confirmed the results of the present study, demonstrating that biomarkers such as the PLR and NLR are significantly related to the stage of GC. However, one study indicated that NLR and PLR markers were more strongly associated with the diagnosis of GC in male patients, suggesting that sex plays a role in this context. In contrast, the

present study showed no relationship between sex or biomarkers and the stage of the disease [51-53]

L. Lian et al. investigated the NLR and PLR and their relationships with GC prognosis and reported that the NLR and PLR in patients with GC were significantly greater before surgery than they were in healthy subjects. Their results revealed that lower levels of two factors, PLR and NLR, before surgery indicated less damage, including a lower depth of invasion and fewer lymph node metastases. Zhou et al. demonstrated that GBM was the best model for predicting the prognosis of inflammatory factors in GC; however, the model still requires further training to achieve greater accuracy [28]. In this study, like our study, models such as LR were also utilized, with the accuracy of the current study being slightly higher in this model. Xu et al. also predicted colorectal cancer recurrence via DT, gradient boosting, LR, and GBM algorithms, among which the well-developed algorithms were gradient boosting and GBM (52). The current study did not use the GBM algorithm but instead employed other algorithms. In a study aimed at predicting bladder cancer stage identification via ML, Kouznetsova et al. reported that MLP_PCA_1 has a high average accuracy [54-55].

In summary, the present study investigated the relationships between various features and the stage of GC and demonstrated that tumor size, survival status, overall survival status, and tumor location are the most important effective features in the disease stage. Additionally, among the algorithms used, the KNN algorithm is one of the best, exhibiting the highest accuracy, along with the SVM, DT, and RF algorithms, while it is more specific than the other algorithms. Therefore, the use of ML is a noninvasive and efficient method for predicting the stage of GC.

Data Availability

Data from this study are available in the Biostudies database: <https://www.ebi.ac.uk/biostudies/studies?query=S-EPMC5373584>

References

- Zhou C, Hu J, Wang Y, Ji M-H, Tong J, et al. (2021) A machine learning-based predictor for the identification of the recurrence of patients with gastric cancer after operation 11: 1571.
- Afrash MR, Shafiee M, Kazemi-Arpanahi HJBg (2023) Establishing machine learning models to predict the early risk of gastric cancer based on lifestyle factors 23: 6.
- Liu X, Tao P, Su H, Li YJSR (2025) Machine learning-random forest model was used to construct gene signature associated with cuproptosis to predict the prognosis of gastric cancer 15: 41-70.
- Baradaran Rezaei H, Amjadian A, Sebt MV, Askari R, Gharaei AJAoOR (2023) An ensemble method of the machine learning to prognosticate the gastric cancer 328: 151-192.
- Hesami Z, Sabzehali F, Khorsand B, Alipour S, Sadeghi A, et al. (2025) Microbiota as a State-of-the-art Approach in Precision Medicine for Pancreatic Cancer Management: A Comprehensive Systematic Review. *iScience* 112-314.
- Haghzad T, Khorsand B, Razavi SA, Hedayati M (2024) A computational approach to assessing the prognostic implications of BRAF and RAS mutations in patients with papillary thyroid carcinoma. *Endocrine* 86: 707-722.
- Houri H, Aghdaei HA, Firuzabadi S, Khorsand B, Soltanpoor F, et al. (2022) High prevalence rate of microbial contamination in patient-ready gastrointestinal endoscopes in Tehran, Iran: an alarming sign for the occurrence of severe outbreaks. *Microbiology Spectrum* 10: e01897-22.
- Irankhah L, Khorsand B, Naghibzadeh M, Savadi A (2024) Analyzing the performance of short-read classification tools on metagenomic samples toward proper diagnosis of diseases. *Journal of bioinformatics and computational biology* 22: 2450012.
- Kharaghani AA, Harzandi N, Khorsand B, Rajabnia M, Kharaghani AA (2023) High prevalence of Mucosa-Associated extended-spectrum β -Lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae* among Iranian patients with inflammatory bowel disease (IBD). *Annals of Clinical Microbiology and Antimicrobials* 22: 86.
- Khorsand B, Khammari A, Shirvanizadeh N, Zahiri J, Arab SS (2019) OligoCOOL: a mobile application for nucleotide sequence analysis. *Biochemistry and Molecular Biology Education* 47: 201-206.
- Khorsand B, Savadi A, Naghibzadeh M (2020) Parallelizing assignment problem with DNA strands. *Iranian Journal of Biotechnology* 18: e2547.
- Khorsand B, Savadi A, Naghibzadeh M (2020)

- Comprehensive host-pathogen protein-protein interaction network analysis. *BMC bioinformatics* 21: 1-22.
13. Khorsand B, Asadzadeh Aghdaei H, Nazemalhosseini-Mojarad E, Nadalian B, Nadalian B, et al. (2022) Overrepresentation of Enterobacteriaceae and Escherichia coli is the major gut microbiome signature in Crohn's disease and ulcerative colitis; a comprehensive metagenomic analysis of IBDMDB datasets. *Frontiers in cellular and infection microbiology* 12: 1015890.
 14. SA, Khorsand B, Salehipour P, Hedayati M (2024) Metabolite signature of human malignant thyroid tissue: A systematic review and meta-analysis. *Cancer Medicine* 13: e7184.
 15. Reza A, Mohsen S, Hadi K-AJBG (2023) Establishing machine learning models to predict the early risk of gastric cancer based on lifestyle factors 23: 1-13.
 16. Khorsand B, Teichrow D, Ghanbarian E, Zheng L, Sajjadi SA, et al. (2025) Scalable Markers for Early Cognitive Decline: Plasma p-tau217, Subjective Cognitive Concerns, and Digital Testing: Results from the A4/LEARN studies. medRxiv. <https://www.medrxiv.org/content/10.1101/2025.10.14.25338009v1>
 17. Khorsand B, Ghanbarian E, Rabin LA, Sajjadi SA, Ezzati A (2026) Incremental value of plasma biomarkers in predicting clinical decline among cognitively unimpaired older adults: Results from the A4 trial. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 18: e70321.
 18. Zhou C, Wang Y, Ji M-H, Tong J, Yang J-J, et al. (2020) Predicting peritoneal metastasis of gastric cancer patients based on machine learning 27.
 19. Zhu S-L, Dong J, Zhang C, Huang Y-B, Pan WJPo (2020) Application of machine learning in the diagnosis of gastric cancer based on noninvasive characteristics 15: e0244869.
 20. Khorsand B, Naderi N, Karimian SS, Mohaghegh M, Aghaahmadi A, et al. (2025) Comprehensive transcriptomic analysis of hepatocellular Carcinoma: Uncovering shared and unique molecular signatures across diverse etiologies. *Biochemistry and Biophysics Reports* 43: 102123.
 21. Khorsand B, Teichrow D, Jicha CJ, Minen MT, Seng E, et al. (2026) Digital Assessment of Objective and Patient-Reported Cognition Across Migraine Phases: Results from the MIND Cohort. medRxiv <https://www.medrxiv.org/content/10.64898/2026.04.14.26350892v1.full.pdf+html>.
 22. Khorsand B, Teichrow D, Lipton RB, Ezzati A (2026) The MIND Study: Design, Feasibility, and Baseline Characteristics of a Smartphone-Based Migraine Cohort. medRxiv <https://www.medrxiv.org/content/10.64898/2026.04.14.26350866v1>.
 23. Rostami-Nejad M, Asri N, Olfatifar M, Khorsand B, Hourri H, et al.(2023) Systematic review and dose-response meta-analysis on the relationship between different gluten doses and risk of coeliac disease relapse. *Nutrients* 15: 1390.
 24. Sadeghnezhad E, Sharifi M, Zare-maivan H, Khorsand B, Zahiri J (2019) Cross talk between energy cost and expression of Methyl Jasmonate-regulated genes: from DNA to protein. *Journal of Plant Biochemistry and Biotechnology* 28: 230-243.
 25. Abolghasemi S, Torbati MB, Pakzad P, Ghafouri-Fard S (2023) Gene expression analysis of SOCS, STAT and PIAS genes in lung cancer patients. *Pathology-Research and Practice* 249: 154760.
 26. Koushki EH, Abolghasemi S, Mollica A, Aghaeepoor M, Moosavi SS, et al. (2020) Structure-based virtual screening, molecular docking and dynamics studies of natural product and classical inhibitors against human dihydrofolate reductase. *Network Modeling Analysis in Health Informatics and Bioinformatics* 9:1-21.
 27. Samarin ZE, Abolghasemi S, Dehnavi E, Akbarzadeh A, Hadian A, et al. (2018) Response Surface Optimization of the Expression Conditions for Synthetic Human Interferon α -2b Gene in Escherichia coli. *Indian Journal of Pharmaceutical Sciences* 80: 470-479.
 28. Zhou C-M, Wang Y, Yang J-J, Zhu YJBMI, et al.(2023) Predicting postoperative gastric cancer prognosis based on inflammatory factors and machine learning technology 23: 53.
 29. Hematpour A, Habibi P, Alavimanesh S, Dadkhah K, Babaie K (2025) Machine learning approach to predict protein-protein interactions between human and hepatitis E virus: revealing links to hepatocellular carcinoma. bioRxiv <https://www.biorxiv.org/content/10.1101/2025.02.23.639757v1>.
 30. Jalali S, Dadkhah K, Ghazi MM (2025) Peritoneal Metastasis Prediction in Gastric Cancer: A Machine Learning Approach. medRxiv <https://www.semanticscholar.org/paper/Peritoneal-Metastasis->

- Prediction-in-Gastric-Cancer%3A-Jalali-Dadk hah/93e790a7915e4cc5070e9823d46af455d81 eb317.
31. Hesami Z, Atashrooz M, Sardarzehi R, Looha MA, Khorsand B, et al. (2025) The oro- and nasopharyngeal microbiota as a revolutionary perspective on mental disorders and related psychopathology: a systematic review and meta-analysis. *Journal of Translational Medicine* 23: 726.
 32. Ahn HS, Lee HJ, Hahn S, Kim WH, Lee KU, et al. (2010) Evaluation of the seventh American Joint Committee on Cancer/International Union Against Cancer Classification of gastric adenocarcinoma in comparison with the sixth classification 116: 5592-5598.
 33. Yodying H, Matsuda A, Miyashita M, Matsumoto S, Sakurazawa N, et al. (2016) Prognostic significance of neutrophil-to-lymphocyte ratio and platelet-to-lymphocyte ratio in oncologic outcomes of esophageal cancer: a systematic review and meta-analysis 23: 646-654.
 34. Ishizuka M, Oyama Y, Abe A, Kubota K, Joso. (2014) Combination of platelet count and neutrophil to lymphocyte ratio is a useful predictor of postoperative survival in patients undergoing surgery for gastric cancer 110: 935-941.
 35. Samandari-Bahraseman MR, Khorsand B, Zareei S, Amanlou M, Rostamabadi H (2023) Various concentrations of hesperetin induce different types of programmed cell death in human breast cancerous and normal cell lines in a ROS-dependent manner. *Chemico-Biological Interactions* 382: 110642.
 36. Shiralipour A, Khorsand B, Jafari L, Salehi M, Kazemi M, et al. (2022) Identifying key lysosome-related genes associated with drug-resistant breast cancer using computational and systems biology approach. *Iranian Journal of Pharmaceutical Research: IJPR* 21 :e130342.
 37. Soltanyzadeh M, Khorsand B, Baneh AA, Hour H (2022) Clarifying differences in gene expression profile of umbilical cord vein and bone marrow-derived mesenchymal stem cells; a comparative in silico study. *Informatics in Medicine Unlocked* 33: 101072.
 38. Zareei S, Khorsand B, Dantism A, Zareei N, Asgharzadeh F, et al. (2024) PeptiHub: a curated repository of precisely annotated cancer-related peptides with advanced utilities for peptide exploration and discovery. Database <https://portal.research.lu.se/en/publications/peptihub-a-curated-repository-of-precisely-annotated-cancer-relat/>.
 39. Hourfar H, Taklifi P, Razavi M, Khorsand B (2025) Machine Learning-Driven Identification of Molecular Subgroups in Medulloblastoma via Gene Expression Profiling. *Clinical Oncology* 103-789.
 40. Khorsand B, Vaghf A, Salimi V, Zand M, Ghoreishi SA (2025) Enhancing ischemic stroke management: leveraging machine learning models for predicting patient recovery after Alteplase treatment. *Brain Injury* 1-7.
 41. Khorsand B, Rajabnia M, Jahanian A, Fathy M, Taghvaei S (2025) Enhancing the accuracy and effectiveness of diagnosis of spontaneous bacterial peritonitis in cirrhotic patients: a machine learning approach utilizing clinical and laboratory data. *Advances in Medical Sciences* 70: 1-7.
 42. Khorsand B, Savadi A, Zahiri J, Naghibzadeh M (2020) Alpha influenza virus infiltration prediction using virus-human protein-protein interaction network. *Mathematical Biosciences and Engineering* 17: 3109-3129.
 43. Khorsand B, Savadi A, Naghibzadeh M (2020) SARS-CoV-2-human protein-protein interaction network. *Informatics in medicine unlocked* 20: 100413.
 44. Samandari Bahraseman MR, Khorsand B, Esmaeilzadeh-Salestani K, Sarhadi S, Hatami N, et al. (2022) The use of integrated text mining and protein-protein interaction approach to evaluate the effects of combined chemotherapeutic and chemopreventive agents in cancer therapy. *Plos one* 17: e0276458.
 45. Mousazadeh M, Khorsand B, Modarres Mousavi SM, Mahmoudi Aznaveh H, Fouani MH, et al. (2025) Liposomal Cancer Drug Database (LCDD): a comprehensive resource for liposome research in cancer therapy and diagnosis. Database <https://visualize.jove.com/41171289-liposomal-cancer-drug-database-lcdd-a-comprehensive-resource-for-liposome-research-in-cancer-therapy-and-diagnosis>.
 46. Halimi H, Hesami Z, Asri N, Khorsand B, Rostami-Nejad M (2025) Exploring the biliary microbiome in hepatopancreatobiliary disorders: a comprehensive systematic review of microbial signatures and diagnostic potential. *BMC gastroenterology* <https://pubmed.ncbi.nlm.nih>.

- gov/41408188/.
47. Chen Y, Wang B, Zhao Y, Shao X, Wang M, et al. (2024) Metabolomic machine learning predictor for diagnosis and prognosis of gastric cancer 15: 1657.
 48. Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick FJTL (2020) Gastric cancer 396: 635-648.
 49. Tanaka Y, Chiwaki F, Kojima S, Kawazu M, Komatsu M, et al. (2021) Multi-omic profiling of peritoneal metastases in gastric cancer identifies molecular subtypes and therapeutic vulnerabilities 2: 962-977.
 50. Zu H, Wang F, Ma Y, Xue YJPo (2013) Stage-stratified analysis of prognostic significance of tumor size in patients with gastric cancer8: e54502.
 51. Guo J, Chen S, Chen Y, Li S, Xu DJCm (2018) research. Combination of CRP and NLR: a better predictor of postoperative survival in patients with gastric cancer 315-321.
 52. Xu Y, Ju L, Tong J, Zhou C-M, Yang J-JJSr (2020) Machine learning algorithms for predicting the recurrence of stage IV colorectal cancer after tumor resection 10: 2519.
 53. Fang T, Wang Y, Yin X, Zhai Z, Zhang Y, et al. (2020) Diagnostic sensitivity of NLR and PLR in early diagnosis of gastric cancer <https://pmc.ncbi.nlm.nih.gov/articles/PMC7081040/>
 54. Lian L, Xia Y-Y, Zhou C, Shen X-M, Li X-L, et al. (2015) Application of platelet/lymphocyte and neutrophil/lymphocyte ratios in early diagnosis and prognostic prediction in patients with resectable gastric cancer 15: 899-907.
 55. Kouznetsova VL, Kim E, Romm EL, Zhu A, Tsigelny IFJM (2019) Recognition of early and late stages of bladder cancer using metabolites and machine learning 15: 1-15.