



Research Article

An explainable algorithm for survival prediction in gastric cancer patients

Leila Alaei¹ , * Ali Broumandnia² , Amin Golabpour³ , Sina Dami⁴ 

1. Department of Computer Engineering, Ki.C., Islamic Azad University, Kish, Iran.
2. Department of Computer Engineering, ST.C., Islamic Azad University, Tehran, Iran.
3. School of Allied Medical Sciences, Shahroud University of Medical Sciences, Shahroud, Iran.
4. Department of Computer Engineering, WT.C., Islamic Azad University, Tehran, Iran.



Citation: Alaei L, Broumandnia A, Golabpour A, Dami S. An explainable algorithm for survival prediction in gastric cancer patients. *Journal of Modern Medical Information Science*. 2025; 10(4):453-468. [In Persian]

 10.48312/jmis.10.4.909.2

Article Info:

Received: 5 Mar 2025
Accepted: 27 Jul 2025
Available Online: 30 Jun 2025

Key Words:

Gastric Cancer, Survival Duration, Artificial Intelligence, Support Vector Regression, LIME.

ABSTRACT

Introduction: Accurate and interpretable prediction of survival duration in patients with gastric cancer can significantly enhance clinical decision-making and treatment planning. This study presents a hybrid model based on Support Vector Regression (SVR) and the LIME algorithm to predict the survival time of patients with gastric cancer.

Methods: This retrospective study included data from 384 patients diagnosed with gastric cancer over a 20-year period. To predict survival time, a Support Vector Regression (SVR) model with an RBF kernel was applied. SVR was selected because of its strong capability to model complex nonlinear relationships in continuous data. To enhance the interpretability of the results, the LIME algorithm was used to analyze the influence of individual variables. The model performance was evaluated using the C-Index, mean absolute error (MAE), and mean squared error (MSE).

Results: The SVR model achieved a C-Index of 0.87, MAE of 45.3 days, and MSE of 56.7. LIME analysis showed that while addiction, family history of gastric cancer, and cause of death had negative effects on survival prediction, factors such as combination therapy, adenocarcinoma histology, education level, and age at diagnosis had a substantial beneficial impact.

Discussion: A dependable and understandable model for forecasting survival time in patients with stomach cancer was developed using the combination of SVR and LIME. The model's interpretability makes it appropriate for clinical settings, where decision-making procedures require transparency and trust.

* Corresponding Author:

Dr Ali Broumandnia
Address: Islamic Azad University, Tehran, Iran.
Tel: +98 9120218552
E-mail: ali.broumandnia@iau.ac.ir





Extended Abstract

Introduction:

Gastric cancer remains one of the most lethal malignancies worldwide, with a high mortality rate, particularly due to late diagnosis and limited treatment options. Accurate prediction of survival duration in patients with gastric cancer is crucial for individualized treatment planning, resource allocation, and improving patient outcomes. Traditional statistical models, such as the Cox Proportional Hazards model and Kaplan-Meier estimators, have been widely used in survival analyses. However, these models often struggle with the complex nonlinear relationships that are characteristic of real-world clinical data. In recent years, machine learning approaches have emerged as powerful tools for modeling these complexities. Despite their predictive power, many machine learning models are criticized for their ‘black box’ nature, which limits their acceptance in clinical decision-making owing to their lack of interpretability. To overcome this barrier, the present study introduces an explainable artificial intelligence (XAI) approach that combines Support Vector Regression (SVR) with the Local Interpretable Model-Agnostic Explanations (LIME) technique. The proposed model aims to improve predictive accuracy and provide transparent, understandable outputs that can support clinicians in making informed decisions. This study leveraged a 20-year retrospective dataset of patients with gastric cancer to validate the performance of the proposed model and explore the influence of key clinical, demographic, and pathological variables on survival duration.

Methods:

This retrospective study analyzed the clinical data of 384 patients diagnosed with gastric cancer over 20 years. The primary objective was to develop a predictive model for survival duration using Support Vector Regression (SVR) with a radial basis function (RBF) kernel[6]. The dataset included 14 independent variables covering demographic information, comorbidities, treatment modalities, pathological types, and family history. Survival duration, the target variable, was calculated from the date of diagnosis to the date of death and expressed in days. Censored data (i.e., patients with unknown survival outcomes) were excluded from

the analyses. The dataset was split into training (70%) and testing (30%) sets for the model. The hyperparameters of the SVR model were tuned using cross-validation. To enhance the interpretability of the SVR predictions, the Local Interpretable Model-Agnostic Explanations (LIME) algorithm was applied. LIME generates local, human-interpretable explanations for the model’s predictions by approximating the complex SVR model with simpler surrogate models in the vicinity of each prediction. Model performance was assessed using standard metrics: Concordance Index (C-Index), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Median Survival Time (MST). Comparisons were also made with traditional statistical models (Cox and Kaplan-Meier) and other machine learning models (K-Nearest Neighbors, Artificial Neural Networks, and Random Forests).

Results:

The SVR model demonstrated strong predictive performance, with a C-Index of 0.87, indicating high concordance between predicted and actual survival rankings. Mean Absolute Error (MAE) was 45.3 days and Mean Squared Error (MSE) was 56.7, suggesting low prediction error. The model-predicted Median Survival Time (MST) was 342 days.

Table 1. Performance Metrics of the SVR Model

C-Index	MAE (days)	MSE (days)	MST (days)
0.87	45.3	56.7	342

LIME analysis revealed the most influential variables affecting survival predictions. The positive predictors included combination therapy, adenocarcinoma pathology, and younger age at diagnosis. The negative predictors included addiction and a family history of gastric cancer. Figure One shows the relative impact of the variables derived from the LIME analysis

The rules extracted from LIME provided clinically interpretable insights. For example, patients aged < 65 years who received combination therapy had a 30% longer predicted survival than those who received only chemotherapy.

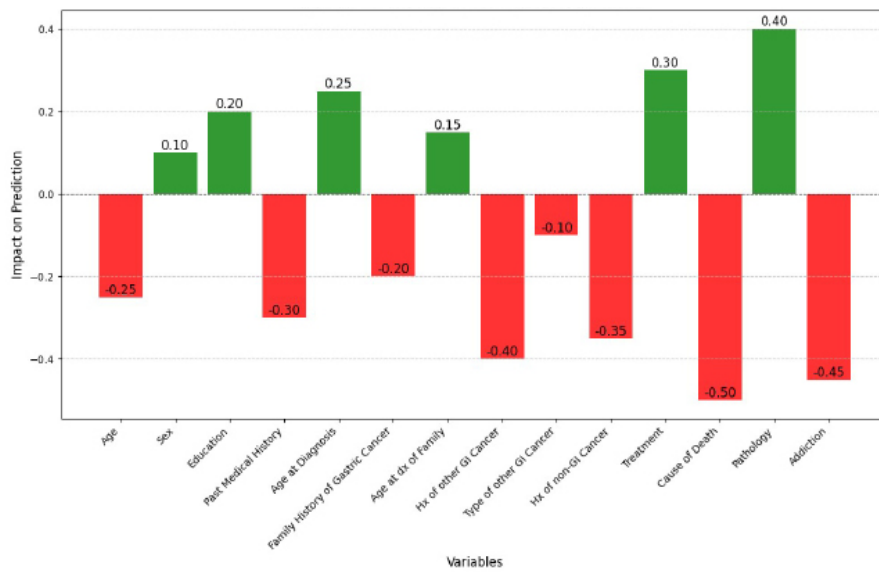


Figure 1. Variable Impact on Survival Prediction

Conclusion:

This study introduced a hybrid machine learning approach that combines the predictive power of Support Vector Regression with the interpretability of LIME to forecast survival duration in patients with gastric cancer. The SVR-LIME model outperformed both traditional survival analysis techniques and several machine learning models in terms of accuracy and reliability. Importantly, the model provided interpretable outputs that could be understood and trusted by clinicians. The ability to extract simple decision rules from complex data enhances the model's usability in real-world clinical settings, offering personalized risk assessments and supporting tailored treatment strategies. Key findings, such as the importance of combina-

tion therapy and early diagnosis, reinforce existing clinical knowledge, while the quantitative impact of sociodemographic and behavioral factors (e.g., education and addiction) offers new perspectives for holistic cancer care. Despite the promising results, this study has some limitations, including the single-center nature of the dataset and the relatively small sample size. Future research should aim for external validation across multicenter datasets and potentially integrate genomic or imaging data to further improve model performance. Overall, this study underscores the value of XAI in bridging the gap between data-driven predictions and clinical applicability, paving the way for more transparent and patient-centered healthcare innovations.



مقاله پژوهشی

ارائه یک الگوریتم توضیح‌پذیر برای پیش‌بینی مدت بقا بیماران مبتلا به سرطان معده

لیلا علانی^۱ ID، * علی برومندنیا^۲ ID، امین گلاب‌پور^۳ ID، سینا دامی^۴ ID

۱. گروه مهندسی کامپیوتر، واحد بین‌المللی کیش، دانشگاه آزاد اسلامی، کیش، ایران.
۲. گروه مهندسی کامپیوتر و هوش مصنوعی، دانشگاه آزاد واحد تهران جنوب، تهران، ایران.
۳. دانشکده پیراپزشکی، دانشگاه علوم پزشکی شاهرود، شاهرود، ایران.
۴. گروه مهندسی کامپیوتر و هوش مصنوعی، دانشگاه آزاد واحد تهران غرب، تهران، ایران.

Use your device to scan and read the article online



Citation: Alaei L, Broumandnia A, Golabpour A, Dami S. An explainable algorithm for survival prediction in gastric cancer patients. *Journal of Modern Medical Information Science*. 2025; 10(4):453-468. [In Persian]

doi 10.48312/jmis.10.4.909.2

چکیده

هدف: پیش‌بینی مدت بقا در بیماران مبتلا به سرطان معده با استفاده از مدل‌های دقیق و قابل تفسیر می‌تواند نقش مهمی در تصمیم‌گیری‌های بالینی و مدیریت درمان داشته باشد. در این مطالعه، مدلی ترکیبی مبتنی بر الگوریتم رگرسیون بردار پشتیبان (SVR) و روش توضیح‌پذیر LIME برای پیش‌بینی مدت بقا در بیماران مبتلا به سرطان معده ارائه شده است.

روش‌ها: این مطالعه به صورت گذشته‌نگر بر روی داده‌های مربوط به ۳۸۴ بیمار مبتلا به سرطان معده، در بازه زمانی ۲۰ ساله انجام شد. برای پیش‌بینی مدت بقا، از مدل رگرسیون بردار پشتیبان (SVR) با کرنل RBF استفاده گردید. مدل SVR به دلیل توانایی بالا در مدل‌سازی روابط غیرخطی پیچیده در داده‌های پیوسته انتخاب شد. به منظور افزایش تفسیرپذیری نتایج، الگوریتم LIME برای تحلیل تأثیر متغیرها به کار گرفته شد. عملکرد مدل با استفاده از شاخص‌هایی مانند C-Index، میانگین خطای مطلق (MAE) و میانگین مربعات خطا (MSE) مورد ارزیابی قرار گرفت.

نتایج: مدل SVR به ترتیب C-Index معادل ۰/۸۷، MAE برابر با ۴۵/۳ روز و MSE برابر با ۵۶/۷ را نشان داد. تحلیل LIME نشان داد که متغیرهایی مانند درمان ترکیبی، پاتولوژی آدنوکارسینوما، سطح تحصیلات و سن در زمان تشخیص تأثیر مثبت و عواملی مانند اعتیاد، سابقه خانوادگی سرطان معده و علت مرگ تأثیر منفی در پیش‌بینی مدت بقا داشتند.

نتیجه‌گیری: ترکیب الگوریتم SVR با تحلیل LIME، مدلی تفسیرپذیر برای پیش‌بینی مدت بقا در بیماران مبتلا به سرطان معده ارائه داد. تفسیرپذیری این مدل، آن را به گزینه‌ای مناسب برای استفاده در محیط‌های بالینی تبدیل می‌کند که در آن شفافیت و اعتمادپذیری در تصمیم‌گیری اهمیت بالایی دارد.

اطلاعات مقاله

تاریخ دریافت: ۱۵ اسفند ۱۴۰۳

تاریخ پذیرش: ۸ تیر ۱۴۰۴

تاریخ انتشار: ۶ مرداد ۱۴۰۴

کلیدواژه‌ها:

سرطان معده، مدت بقا، هوش مصنوعی، رگرسیون بردار پشتیبان، LIME.

*نویسنده مسئول:

دکتر علی برومندنیا

نشانی: دانشگاه آزاد واحد تهران جنوب، تهران، ایران.

تلفن: +98 9120218552

پست الکترونیک: ali.broumandnia@iau.ac.ir



Copyright © 2024 The Author[s].

This is an open access article distributed under the terms of the Creative Commons Attribution License [CC-BY-NC; <https://creativecommons.org/licenses/by-nc/4.0/legalcode.en>], which permits use, distribution, and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

مقدمه:

مدت بقا یکی از شاخص‌های کلیدی در ارزیابی وضعیت بیماران مبتلا به بیماری‌های مزمن و تهدیدکننده حیات محسوب می‌شود [۱]. این شاخص، نقشی اساسی در تعیین نوع و شدت مراقبت‌های درمانی، تصمیم‌گیری‌های پزشکی، برنامه‌ریزی مراقبت‌های حمایتی و حتی مدیریت منابع نظام سلامت دارد [۲]. در بسیاری از بیماری‌ها از جمله انواع مختلف سرطان‌ها، اطلاع از مدت بقا می‌تواند بینش دقیقی نسبت به روند بیماری، پیش‌آگهی بیمار و کارایی مداخلات درمانی فراهم آورد، در این میان، سرطان معده به‌عنوان یکی از شایع‌ترین و مرگ‌بارترین سرطان‌های دستگاه گوارش، جایگاه ویژه‌ای دارد [۳]؛ چراکه اغلب در مراحل پیشرفته تشخیص داده می‌شود و نرخ بقای پایینی دارد، موضوعی که اهمیت پیش‌بینی دقیق مدت بقا را در این بیماران دوچندان می‌کند [۴].

مطالعات جهانی حاکی از آن است که سرطان معده سالانه بیش از یک‌میلیون نفر را درگیر می‌کند و به‌عنوان یکی از پنج عامل اصلی مرگ ناشی از سرطان در سطح بین‌المللی شناخته می‌شود [۵]. در کشورهایمانند ایران نیز، این بیماری از شیوع و مرگ‌ومیر بالایی برخوردار است [۶]؛ به‌طوری‌که در سال ۲۰۲۲ ایران در جایگاه هشتم جهانی از نظر مرگ‌ومیر ناشی از سرطان معده قرار داشته است [۷]. با توجه به ماهیت تهاجمی و اغلب دیر تشخیص این سرطان، ابزارهای دقیق و قابل‌اعتماد برای پیش‌بینی طول عمر بیماران می‌تواند نقشی حیاتی در مدیریت درمانی و مراقبتی ایفا کند [۸].

برای تحلیل مدت بقا، عموماً از مدل‌هایی مانند مدل خطر نسبی کاکس^۱ و روش کاپلان-مایر^۲ استفاده می‌شود [۹،۱۰]. این روش‌ها به دلیل ساختار آماری ساده و قابلیت استفاده در نمونه‌های نسبتاً کوچک، در مطالعات بالینی بسیار رایج هستند [۱۱]. با این حال، این مدل‌ها محدودیت‌هایی نیز دارند؛ به‌ویژه در مواجهه با روابط پیچیده و غیرخطی بین متغیرهای بالینی و زیستی، عملکرد

1 Cox Proportional Hazards Model
2 Kaplan-Meier Estimator

آن‌ها کاهش می‌یابد [۱۲]. علاوه بر این، در شرایطی که داده‌ها دارای ابعاد بالا یا ساختارهای غیرمتعارف باشند، دقت و انعطاف‌پذیری این روش‌ها برای پیش‌بینی دقیق‌تر کاهش می‌یابد [۱۳]. در چنین شرایطی، روش‌های مبتنی بر یادگیری ماشین می‌توانند عملکرد بهتری از خود نشان دهند و بینش‌های عمیق‌تری از داده‌ها استخراج کنند [۱۴،۱۵].

الگوریتم‌های یادگیری ماشین به دلیل قابلیت پردازش داده‌های پیچیده و مدل‌سازی روابط غیرخطی، توانسته‌اند در سال‌های اخیر جایگاه برجسته‌ای در تحلیل‌های پیش‌بینی محور پزشکی پیدا کنند. با این حال، استفاده از آن‌ها با چالش‌هایی نیز همراه است. یکی از مهم‌ترین این چالش‌ها، عدم شفافیت در فرایند تصمیم‌گیری این الگوریتم‌ها است که باعث می‌شود آن‌ها به‌عنوان «مدل‌های جعبه سیاه» شناخته شوند [۱۶]. در حوزه پزشکی، که تصمیم‌گیری‌ها باید براساس شواهد و دلایل قابل‌فهم و قابل‌اعتماد اتخاذ شود، این ویژگی می‌تواند مانعی جدی برای کاربرد عملی این مدل‌ها باشد.

برای غلبه بر این چالش، حوزه‌ای نوین به نام هوش مصنوعی توضیح‌پذیر^۳ ظهور یافته است [۱۷]. این رویکرد تلاش می‌کند تا با حفظ دقت مدل، تفسیرپذیری تصمیمات آن را نیز ممکن سازد. از جمله روش‌های پرکاربرد در این زمینه، الگوریتم LIME^۴ است که با ایجاد مدل‌های محلی ساده، به تحلیل تأثیر ویژگی‌ها در تصمیم نهایی مدل کمک می‌کند و امکان ارائه توضیحاتی شفاف برای پزشکان را فراهم می‌آورد. در کنار LIME، روش‌های دیگری مانند SHAP^۵ نیز توسعه یافته‌اند، اما LIME به دلیل ساختار ساده‌تر و قابل‌پیاده‌سازی روی مدل‌های متنوع، در مطالعات پزشکی کاربرد بیشتری یافته است [۱۸].

در این پژوهش، با هدف ارائه مدلی دقیق، قابل تفسیر و کاربردی برای پیش‌بینی مدت بقا در بیماران مبتلا به سرطان معده، از الگوریتم رگرسیون بردار پشتیبان^۶ SVR

3 Explainable Artificial Intelligence – XAI

4 Local Interpretable Model-agnostic Explanations

5 Shapley Additive exPlanations

6 Support Vector Regression

مدت بقا در بیماران مبتلا به سرطان معده، از الگوریتم رگرسیون بردار پشتیبان به‌عنوان مدل اصلی استفاده شد. این الگوریتم به دلیل توانایی بالا در حل مسائل پیچیده پیوسته و مدیریت داده‌های غیرخطی انتخاب گردید. SVR به‌خوبی با داده‌های چندبعدي سازگار است و قادر به ارائه پیش‌بینی‌های دقیق هست [۱۹]. با این حال، برای افزایش شفافیت مدل و تفسیرپذیری پیش‌بینی‌ها، از الگوریتم LIME استفاده خواهد شد.

در این پژوهش، ابتدا مدل SVR بر روی مجموعه داده‌های آموزشی (۷۰ درصد) آموزش داده شد و سپس عملکرد آن بر روی مجموعه آزمایشی (۳۰ درصد) با استفاده از معیارهایی مانند شاخص توافق^۷، میانگین خطای مطلق^۸، میانگین مربعات خطا^۹ و میانه مدت بقا^{۱۰} ارزیابی گردید. C-Index معیاری برای سنجش دقت مدل‌های پیش‌بینی بقا است که نشان می‌دهد مدل چقدر به‌درستی بیماران را براساس مدت بقا رتبه‌بندی می‌کند. MAE برای اندازه‌گیری اختلاف میان مقادیر پیش‌بینی و واقعی، MSE برای تحلیل پراکندگی خطاها و MST به‌عنوان یک شاخص مرکزی از پیش‌بینی مدل، اطلاعات ارزشمندی برای پزشکان فراهم می‌آورد. این شاخص می‌تواند در ارزیابی روند بیماری، مقایسه اثر روش‌های درمانی مختلف و شناسایی بیماران پرخطر که نیاز به مداخلات خاص دارند، مفید باشد. به‌عنوان مثال، بیمارانی که پیش‌بینی می‌شود مدت بقای کمتری از میانه داشته باشند، ممکن است به برنامه‌های درمانی پیشرفته‌تری نیاز داشته باشند [۲۰]. این مقدار همچنین به تحلیل تأثیر متغیرهای مستقل، مانند سن یا نوع درمان، بر طول عمر بیماران کمک می‌کند و ابزاری برای برنامه‌ریزی بهتر درمانی و پیش‌آگهی فراهم می‌آورد.

هدف از این روش، ارائه مدلی بود که هم دقت بالایی در پیش‌بینی مدت بقا داشته باشد و هم توضیحات قابل‌فهمی برای پزشکان و متخصصان فراهم آورد. این مدل می‌توانست به بهبود تصمیم‌گیری‌های درمانی و ارتقای کیفیت زندگی بیماران کمک کند.

به‌عنوان الگوریتم اصلی و از LIME به‌عنوان ابزار تحلیل تفسیرپذیری استفاده شده است. شایان‌ذکر است که هدف از به‌کارگیری LIME در این مطالعه نه بهبود عملکرد پیش‌بینی، بلکه افزایش شفافیت، درک‌پذیری و تفسیر خروجی‌های مدل بوده است تا پزشکان بتوانند از نتایج آن در تصمیم‌گیری بالینی استفاده کنند. این رویکرد ترکیبی، امکان استفاده از توانمندی‌های یادگیری ماشین را در کنار قابلیت اطمینان و تفسیر بالینی فراهم می‌آورد. هدف نهایی این مطالعه، توسعه مدلی است که ضمن برخورداری از دقت بالا در پیش‌بینی مدت بقا، با ارائه قوانین ساده و روشن مبتنی بر تحلیل LIME، راهگشای پزشکان در انتخاب استراتژی‌های درمانی مؤثرتر باشد.

مواد و روش‌ها:

در این مطالعه، متغیر هدف یا وابسته، مدت بقا بود که به‌صورت پیوسته و برحسب روز اندازه‌گیری شد. مقدار این متغیر از حداقل ۱۸ روز تا حداکثر ۷۱۴ روز در بین بیماران متغیر بود. این متغیر براساس تاریخ تشخیص تا زمان فوت محاسبه شده است. داده‌های مربوط به بیماران سانسور شده (موردی که اطلاعات دقیقی از زمان فوت در دسترس نبود) از تحلیل حذف شدند؛ بنابراین، در تحلیل نهایی فقط نمونه‌هایی که بقا به‌طور کامل مشخص بود ($n=384$) لحاظ شدند. در مجموع، ۱۴ متغیر مستقل در این تحقیق بررسی شد. این متغیرها از نظر نوع مقیاس شامل سه متغیر پیوسته (سن بیمار، سن در زمان تشخیص و سن تشخیص بیماری در بستگان)، شش متغیر اسمی (جنسیت، وضعیت اعتیاد، سابقه بیماری‌های مزمن، سابقه خانوادگی سایر سرطان‌های گوارشی یا غیر گوارشی) و پنج متغیر رتبه‌ای (سطح تحصیلات، نوع درمان دریافتی، نوع پاتولوژی، علت مرگ و سابقه خانوادگی سرطان معده) بودند.

ترکیب متوازن متغیرهای کیفی و کمی در این مجموعه داده، شرایط مناسبی برای تحلیل‌های پیش‌بینی محور فراهم کرده است. به‌ویژه متغیرهای رتبه‌ای که بیانگر شدت یا سطح عوامل خطر هستند، در مدل‌سازی بقا نقش مهمی ایفا می‌کنند.

در این مطالعه، به‌منظور توسعه مدلی برای پیش‌بینی

7 C-Index

8 Mean Absolute Error: MAE

9 Mean Squared Error: MSE

10 Median Survival Time: MST

خیر؟ همچنین، از معیارهای آماری مانند آزمون لاگ-رنک برای مقایسه منحنی‌های بقا استفاده شد. این تحلیل‌ها به بررسی جامع قابلیت‌های مدل پیشنهادی و کاربرد آن در تصمیم‌گیری‌های درمانی کمک نمود. فلوجارت متد در شکل یک نمایش داده شده است.

مدل خطرات نسبی کاکس یکی از روش‌های نیمه پارامتریک رایج در تحلیل بقا است که برای بررسی تأثیر هم‌زمان چندین متغیر مستقل بر زمان وقوع یک رویداد (مانند مرگ یا عود بیماری) به کار می‌رود. این مدل فرض می‌کند که نسبت خطر بین افراد با ویژگی‌های مختلف، در طول زمان ثابت باقی می‌ماند. در مقابل روش کاپلان-مایر یک روش غیر پارامتریک در تحلیل بقا است که برای برآورد تابع بقا در داده‌های زمان-تا-رویداد به کار می‌رود. این روش به‌ویژه برای داده‌های سانسور شده (مواردی که زمان دقیق رویداد مشخص نیست) بسیار مناسب است و احتمال بقا در طول زمان را نشان می‌دهد. برای مقایسه بین منحنی‌های بقا در گروه‌های مختلف (مانند انواع درمان یا سطوح تحصیلات)، از آزمون لاگ-رنک استفاده می‌شود که تفاوت معنادار بین توزیع‌های بقای گروه‌ها را ارزیابی می‌کند.

برای ارزیابی عملکرد مدل پیشنهادی مبتنی بر الگوریتم SVR، خروجی‌های آن با روش‌های آماری رایج در تحلیل مدت بقا مانند مدل کاکس و روش کاپلان-مایر مقایسه شد. مدل کاکس که به‌عنوان یکی از استانداردهای طلایی در تحلیل بقا شناخته می‌شود، اثر متغیرهای مستقل را بر خطر نسبی مدل‌سازی می‌کند و تأثیر هر متغیر را بر مدت بقا تعیین می‌نماید. همچنین، روش کاپلان-مایر توزیع مدت بقا را با استفاده از تخمین‌های غیر پارامتریک نمایش داده و امکان مقایسه گروه‌های مختلف (مانند نوع درمان یا سابقه خانوادگی سرطان) را فراهم می‌کند. همچنین در ادامه الگوریتم SVR به سه روش نزدیک‌ترین همسایگی، شبکه‌های عصبی مصنوعی و رگرسیون جنگل تصادفی مقایسه می‌شود.

در این مطالعه، مدل SVR با معیارهایی نظیر میانه مدت بقا، نرخ بقای یک‌ساله و دوساله و همچنین شاخص‌های C-Index و AUC ارزیابی شد و نتایج آن با خروجی‌های روش‌های کاکس و کاپلان-مایر مقایسه گردید. این مقایسه‌ها مشخص کردند که آیا مدل پیشنهادی از نظر دقت پیش‌بینی و توانایی در تفکیک بیماران با ریسک‌های متفاوت، عملکرد بهتری نسبت به روش‌های سنتی دارد یا



شکل ۱: فلوجارت مدل پیشنهادی

در نمونه مورد بررسی ۳۸۴ بیمار مبتلا به سرطان معده، ویژگی‌های جمعیت شناختی، بالینی و سبک

یافته‌ها:

درصد در بستگان درجه اول بوده است. همچنین ۲۱/۶۱ درصد از بیماران سابقه اعتیاد داشتند. در مورد نوع درمان، بیشترین گروه (۷۴/۷۴ درصد) تنها شیمی درمانی دریافت کرده بودند و تنها شش درصد بیماران درمان ترکیبی شامل جراحی و شیمی درمانی را دریافت نمودند. از نظر نوع پاتولوژی، آدنوکارسینوما شایع ترین نوع بوده و در ۷۰.۰۵ درصد از موارد گزارش شده است. علت مرگ نیز در اکثر موارد (۷۲/۱۴ درصد) مستقیماً به سرطان معده نسبت داده شده است. اطلاعات تفصیلی تمام متغیرهای فوق به همراه نوع داده، توضیحات طبقه بندی و فراوانی آن‌ها در جدول ۱ ارائه شده است.

زندگی بیماران از تنوع قابل توجهی برخوردار بود. بیش از سه چهارم بیماران (۷۸/۳۹ درصد) بالای ۵۰ سال سن داشتند و توزیع جنسیتی نیز با درصد بالاتر مردان (۵۶/۷۷ درصد) نسبت به زنان (۴۳/۲۳ درصد) همراه بود. در زمینه سطح تحصیلات، حدود ۷۶ درصد بیماران بی سواد یا دارای تحصیلات کمتر از دیپلم بودند. بررسی سابقه بیماری‌های مزمن نشان داد که فشارخون (۲۸/۳۹ درصد) دیابت (۲۷/۸۶ درصد) و بیماری عروق کرونر (۲۰/۳۱ درصد) از شیوع نسبتاً بالایی برخوردارند، در حالی که ترکیب این بیماری‌ها نیز در مواردی مشاهده شد. سابقه خانوادگی سرطان معده در ۵۲/۶ درصد از بیماران گزارش شد که از این میان، ۲۷/۸۶

جدول ۱: آمار توصیفی داده‌های مطالعه				
ردیف	نام متغیر	نوع متغیر	تعداد (درصد)	توضیحات
۱	سن	فاصله‌ای	کمتر از ۵۰ سال: ۲۱/۶۱ (درصد) بیشتر از ۵۰ سال: ۷۸/۳۹ (درصد)	حداقل ۴۸ سال و حداکثر ۸۸ سال
۲	جنسیت	فاصله	مرد: ۵۶/۷۷ (درصد) زن: ۴۳/۲۳ (درصد)	-
۳	تحصیلات	رتبه‌ای	بی سواد: ۷۶/۳ (درصد) باسواد: ۲۳/۷ (درصد)	مدرک پایین‌تر از دیپلم بی سواد در نظر گرفته شده است و بالای با دیپلم باسواد در نظر گرفته شده است.
۴	سوابق پزشکی	اسمی	فشارخون: ۲۸/۳۹ (درصد) عروق کرونر: ۲۰/۳۱ (درصد) دیابت: ۲۷/۸۶ (درصد) دیابت و فشارخون: ۸/۳۳ (درصد) فشارخون و عروق کرونر: ۱۱/۲ (درصد) فشارخون و عروق کرونر و دیابت: ۳/۹۱ (درصد)	سوابق پزشکی نشان دهنده وضعیت‌های پزشکی پیشین بیماران است که شامل شش دسته اصلی می‌شود: فشارخون بالا، بیماری عروق کرونر قلب، دیابت و ترکیب این شرایط
۵	سن در زمان تشخیص	فاصله‌ای	کمتر از ۶۵ سال: ۵۶/۲۵ (درصد) بیشتر از ۶۵ سال: ۴۳/۷۵ (درصد)	دارای مقدار حداقل ۴۶ سال و حداکثر ۸۷ است.
۶	سابقه خانوادگی سرطان معده	رتبه‌ای	خویشاوند درجه اول: ۲۷/۸۶ (درصد) خویشاوند درجه دوم: ۲۷/۷۴ (درصد) عدم وجود سابقه سرطان: ۴۷/۴ (درصد)	سابقه خانوادگی سرطان معده نشان دهنده وجود سرطان معده در اعضای خانواده بیمار است. این متغیر به سه سطح تقسیم می‌شود: ۱. خویشاوند درجه اول: شامل والدین، خواهر و برادر، یا فرزندان بیمار؛ ۲. خویشاوند درجه دوم: شامل پدربزرگ، مادربزرگ، عمو، عمه، دایی، یا خاله. ۳- عدم وجود سابقه سرطان معده در خانواده
۷	سن تشخیص سرطان در خانواده	فاصله‌ای	کمتر از ۵۸ سال: ۲۰/۸۳ (درصد) بیشتر از ۵۸ سال: ۳۱/۷۷ (درصد) عدم وجود سابقه سرطان معده: ۴۷/۴ (درصد)	سن تشخیص در خانواده دارای مقدار حداقل ۴۵ و حداکثر ۸۲ است.
۸	سابقه سایر سرطان‌های دستگاه گوارش	رتبه‌ای	خویشاوند درجه اول: ۱۴/۰۶ (درصد) خویشاوند درجه دوم: ۱۹/۵۳ (درصد) عدم وجود سایر سرطان‌های دستگاه گوارش: ۶۶/۴۱ (درصد)	سابقه سایر سرطان‌های دستگاه گوارش نشان دهنده وجود سرطان‌های دیگر دستگاه گوارش در اعضای خانواده بیمار است. این متغیر به سه سطح تقسیم می‌شود. ۱- خویشاوند درجه اول: شامل والدین، خواهر و برادر، یا فرزندان بیمار. ۲- خویشاوند درجه دوم: شامل پدربزرگ، مادربزرگ، عمو، عمه، دایی، یا خاله ۳- عدم وجود سایر سرطان‌های دستگاه گوارش

جدول ۱: آمار توصیفی داده‌های مطالعه			
ردیف	نام متغیر	نوع متغیر	تعداد (درصد)
۹	نوع سایر سرطان‌های دستگاه گوارش	رتبه‌ای	سرطان روده کوچک: ۱۰ (۲/۶ درصد) سرطان کبد: ۲۴ (۶/۲۵ درصد) سرطان مری: ۳۸ (۹/۹ درصد) سرطان روده معده: ۵۷ (۱۴/۸۴ درصد) هیچ کدام: ۲۵۵ (۶۶/۴۱ درصد)
۱۰	سابقه سرطان‌های غیر گوارشی	رتبه‌ای	خوبشاوند درجه اول: ۱۸ (۴/۶۹ درصد) خوبشاوند درجه دوم: ۱۲ (۳/۱۳ درصد) هیچ کدام: ۳۵۴ (۹۲/۱۹ درصد)
۱۱	نوع درمان	رتبه‌ای	شیمی‌درمانی: ۲۸۷ (۷۴/۷۴ درصد) جراحی: ۷۴ (۱۹/۲۷ درصد) شیمی‌درمانی و جراحی: ۲۳ (۵/۹۹ درصد)
۱۲	علت مرگ	رتبه‌ای	سرطان: ۲۷۷ (۷۲/۱۴ درصد) سکته قلبی: ۶۸ (۱۷/۷۱ درصد) آمبولی ریه: ۳۹ (۱۰/۱۶ درصد)
۱۳	نوع پاتولوژی	رتبه‌ای	Adenocarcinoma: ۲۶۹ (۷۰/۰۵ درصد) signet ring cell carcinoma: ۴۸ (۱۲/۵ درصد) mucinous adenocarcinoma: ۳۸ (۹/۹ درصد) neuroendocrine carcinoma: ۸ (۲/۰۹ درصد) Gastrointestinal Stromal Tumo: ۷ (۱/۸۲ درصد) inflammatory tumor: ۷ (۱/۸۲ درصد) undifferentiated carcinoma: ۷ (۱/۸۲ درصد)
۱۴	وضعیت اعتیاد	اسمی	سالم: ۳۰۱ (۷۸/۳۹ درصد) معتاد: ۸۳ (۲۱/۶۱ درصد)
۱۵	مدت بقا	فاصله‌ای	یک سال: ۲۲۸ (۵۹/۳۸ درصد) دو سال: ۱۵۶ (۴۰/۶۲ درصد)
	وضعیت اعتبار		حداقل ۱۸ و حداکثر ۷۱۴

مشخص کردن خطای مجاز در پیش‌بینی بود. همچنین، برای اطمینان از قابلیت تعمیم مدل، از روش اعتبارسنجی 5-Fold استفاده شد عملکرد مدل پیشنهادی با استفاده از معیارهای استاندارد MAE، C-Index و MSE ارزیابی شد. مدل SVR نتایج بسیار خوبی ارائه داد. در جدول ۲ نتایج الگوریتم SVR بیان شده است.

در این مطالعه، الگوریتم SVR به‌عنوان روش پیشنهادی برای پیش‌بینی مدت بقا در بیماران مبتلا به سرطان معده به کار گرفته شد. این الگوریتم به دلیل توانایی بالا در مدل‌سازی روابط غیرخطی و مدیریت داده‌های پیچیده انتخاب شد. پارامترهای اصلی مدل شامل هسته، $C=0/1$ برای تنظیم تعادل بین دقت و پیچیدگی، $\Gamma=0/1$ برای تأثیر داده‌های آموزشی و $\text{Epsilon}=0/1$ برای

جدول ۲: نتایج الگوریتم پیشنهادی			
MST	MSE	MAE	C-Index
۳۴۲ (۹۵ درصد فاصله اطمینانی: ۳۱۵-۳۷۸)	۵۶/۷	۴۵/۳	۰/۸۷

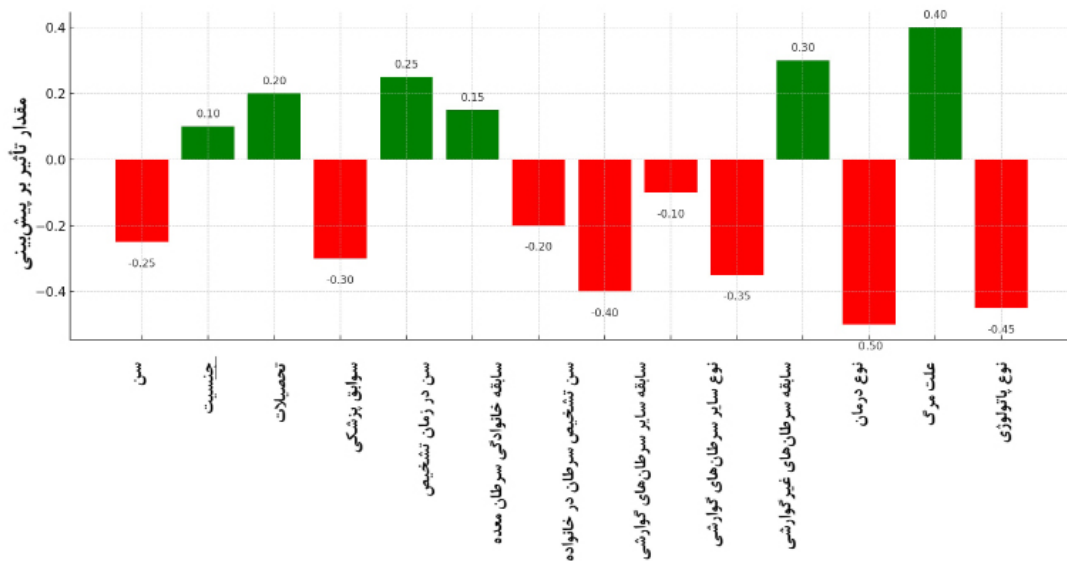
(رتبه‌بندی کاملاً صحیح) متغیر است. مقدار ۰/۸۷ به این معناست که در ۸۷ درصد موارد، مدل به‌درستی بیمارانی

شاخص C-Index در این مطالعه برابر با ۰/۸۷ است. مقدار C-Index بین ۰/۵ (بدون توانایی رتبه‌بندی) تا ۱/۰

MST پیش‌بینی شده که در این مطالعه برابر با ۳۴۲ روز (۳۷۸-۳۱۵: اطمینانی ۹۵ درصد) است، یک شاخص مرکزی از توزیع مدت بقا در پیش‌بینی‌های مدل است. این مقدار نشان می‌دهد که ۵۰ درصد بیماران در پیش‌بینی مدل، مدت بقای کمتر از ۳۴۲ روز و ۵۰ درصد دیگر مدت بقای بیشتر از ۳۴۲ روز خواهند داشت.

تأثیر متغیرهای مختلف بر پیش‌بینی مدت بقا در بیماران مبتلا به سرطان معده است که با استفاده از الگوریتم LIME تحلیل شده‌اند. (شکل ۲).

با خطر بالاتر را در رتبه بالاتری قرار داده است. مقدار MAE برابر با ۴۵/۳ روز است که نشان می‌دهد پیش‌بینی مدل برای مدت بقا به‌طور متوسط ۴۵/۳ روز با مقدار واقعی فاصله دارد. مقدار MSE برابر با ۵۶/۷ روز است که شاخص پراکندگی خطاها را اندازه‌گیری می‌کند. این مقدار نشان‌دهنده عملکرد نسبتاً پایدار مدل است. در مقایسه با MAE، MSE نسبت به خطاهای بزرگ حساس‌تر است و به همین دلیل یک معیار تکمیلی برای ارزیابی دقت مدل محسوب می‌شود.



شکل ۲: تأثیر متغیرهای مطالعه در مدت بقا سرطان معده

جمله عوامل منفی مهم می‌توان به سوابق پزشکی مانند بیماریهای مزمن (ضریب منفی ۰/۳۰)، سابقه خانوادگی سرطان معده (ضریب منفی ۰/۲۰)، و به ویژه سابقه سایر سرطانهای دستگاه گوارش (ضریب منفی ۰/۴۰) اشاره کرد. در مقابل، تشخیص بیماری در سنین پایینتر (ضریب مثبت ۰/۲۵) و تشخیص سرطان در اعضای خانواده در سنین پایینتر (ضریب مثبت ۰/۱۵) ممکن است نشان‌دهنده عوامل ژنتیکی مؤثر بر بقا باشد.

درمان‌های ترکیبی مانند جراحی همراه با شیمی درمانی (ضریب مثبت ۰/۳۰) و نوع خاصی از پاتولوژی مانند آدنوکارسینوما (ضریب مثبت ۰/۴۰) از عوامل بهبوددهنده بقا محسوب میشوند. در نقطه مقابل، اعتیاد (ضریب منفی

تأثیر متغیرها بر مدت بقا از طریق الگوریتم LIME تحلیل شده است. در این نمودار، هر میله نمایانگر تأثیر یک متغیر خاص است و رنگ میله‌ها نشان‌دهنده تأثیر مثبت یا منفی آن متغیر بر پیش‌بینی مدت بقا هست. در ادامه توضیح هر متغیر بیان می‌شود. عوامل مختلفی با تأثیرات مثبت یا منفی بر پیش‌بینی مدت بقای بیماران بررسی شده‌اند. سن بیماران یک عامل منفی با ضریب ۰/۲۵ محسوب میشود، چرا که افراد مسن معمولاً دوره بقای کوتاهتری دارند. جنسیت تأثیر خنثی تا کمی مثبت (ضریب ۰/۱۰) دارد، با احتمال بقای بهتر در برخی زنان. سطح تحصیلات بالاتر به عنوان یک عامل مثبت (ضریب ۰/۲۰) در نظر گرفته شده که ممکن است ناشی از آگاهی بیشتر و دسترسی بهتر به مراقبتهای بهداشتی باشد. از

پیچیده‌های از عوامل بالینی، اجتماعی و ژنتیکی قرار دارد. با استفاده از خروجی LIME، چهار قوانین ساده و قابل تفسیر ارائه می‌شود که در جدول ۳ نشان داده است.

۰/۴۵) و مرگ ناشی از سرطان (ضریب منفی ۰/۵۰) به عنوان شدیدترین عوامل کاهنده بقا شناسایی شده‌اند. این یافته‌ها نشان می‌دهند که پیش‌آگهی بیماران تحت تأثیر

شماره قانون	جزئیات قانون
قانون یک	اگر سن بیمار کمتر از ۶۵ سال باشد و درمان ترکیبی (جراحی و شیمی‌درمانی) دریافت کند، مدت بقا ۳۰ درصد بیشتر از بیمارانی است که تنها شیمی‌درمانی دریافت کرده‌اند.
قانون دو	اگر پاتولوژی آدنوکارسینوما باشد، پیش‌بینی مدت بقا بیشتر از کارسینوم غیر تمایز است.
قانون سه	اگر سابقه خانوادگی سرطان معده در خویشاوندان درجه اول وجود داشته باشد، مدت بقا کمتر از بیماران بدون سابقه خانوادگی است.
قانون چهار	اگر متغیر تحصیلات بیمار بالاتر از دیپلم باشد، این متغیر تأثیر مثبت بر مدت بقا دارد.
قانون پنج	اگر بیمار معتاد باشد، تأثیر منفی بر مدت بقا دارد.

درحالی‌که درمان ترکیبی موجب کاهش خطر مرگ شد (HR=۰/۷۸). با این حال، در مقایسه مدل پیشنهادی با این دو روش سنتی، نتایج نشان داد که مدل پیشنهادی (SVR-LIME) عملکرد دقیق‌تری در پیش‌بینی مدت بقا دارد. این مدل با C-Index برابر با ۰/۸۷ دقت بالاتری در رتبه‌بندی بیماران نشان داد، درحالی‌که این شاخص برای مدل کاکس ۰/۸۲ و برای کاپلان-مایر ۰/۷۸ بود. همچنین، از نظر خطای پیش‌بینی، مدل پیشنهادی با MAE برابر با ۴۵/۳ روز و MSE برابر با ۵۶/۷ روز کمترین میزان خطا را در مقایسه با کاکس (MAE=۵۰/۲، MSE=۶۵/۰) و کاپلان-مایر (MAE=۵۵/۱، MSE=۷۲/۳) داشت. میان مدت بقای پیش‌بینی‌شده (MST) نیز در مدل پیشنهادی برابر با ۳۴۲ روز بود که نسبت به کاکس (۳۲۸ روز) و کاپلان-مایر (۳۱۹ روز) به مقادیر واقعی نزدیک‌تر بود. این برتری در شاخص‌های دقت و خطا نشان‌دهنده توانایی بالاتر مدل پیشنهادی در شناسایی روابط غیرخطی و پیچیده میان متغیرها و ارائه پیش‌بینی‌های قابل اتکاتر است.

مقایسه مدل پیشنهادی SVR با روش‌های آماری کاکس و کاپلان-مایر نشان داد که SVR با دقت بالاتر، خطای کمتر و تفسیرپذیری بهتر عملکرد بهتری در پیش‌بینی مدت بقا در بیماران مبتلا به سرطان معده دارد. در جدول چهار مقایسه مدل پیشنهادی با دو روش مدت بقا کاکس و کاپلان-مایر مشاهده شده است.

در تحلیل بقا با استفاده از روش‌های آماری کلاسیک، هر دو روش کاپلان-مایر و مدل خطرات نسبی کاکس نتایج معناداری را ارائه دادند. منحنی‌های بقا به روش کاپلان-مایر نشان داد که بیماران دریافت‌کننده درمان ترکیبی (شیمی‌درمانی و جراحی) دارای میان مدت بقای بالاتری (۳۴۷ روز) نسبت به بیماران دریافت‌کننده تنها شیمی‌درمانی (۳۲۱ روز) بودند؛ این تفاوت با آزمون لاگ‌رنک معنادار بود ($P < 0.05$). همچنین، بقا در بیماران فاقد سابقه خانوادگی سرطان معده به‌طور معناداری بیشتر از بیماران دارای این سابقه بود. در مدل کاکس نیز متغیرهایی مانند سن بالا (HR=۱/۱۵)، اعتیاد (HR=۱/۳۲) و سابقه خانوادگی سرطان معده (HR=۱/۲۶) با افزایش خطر مرگ همراه بودند،

معیار/مدل	کاکس	کاپلان-مایر	SVR
C-Index	۰/۸۲	۰/۷۸	۰/۸۷
MAE	۵/۲	۵۵/۱	۴۵/۳
MSE	۶۵/۰	۷۲/۳	۵۶/۷

جدول ۴: مقایسه مدل پیشنهادی با دو روش مدت بقا کاکس و کاپلان-مایر

معیار/مدل	کاکس	کاپلان-مایر	SVR
MST (روز)	۳۲۸	۳۱۹	۳۴۲
حساسیت به تغییر متغیرها	متوسط	کم	بالا
تفسیرپذیری	بله (نسبی)	خیر	بله

در مقایسه مدل پیشنهادی با روش‌های آماری کاکس و کاپلان-مایر، نتایج نشان داد که مدل پیشنهادی عملکرد بهتری در پیش‌بینی مدت بقا دارد. مدل پیشنهادی با C-Index برابر با ۰/۸۷، دقت بالاتری در رتبه‌بندی بیماران براساس خطر بقا نشان داد، درحالی‌که این مقدار برای کاکس ۰/۸۲ و برای کاپلان-مایر ۰/۷۸ بود. این دقت بالاتر بیانگر توانایی بهتر پیشنهادی در شناسایی ارتباطات پیچیده میان متغیرها و پیش‌بینی نتایج است.

از نظر خطای پیش‌بینی، مدل پیشنهادی با میانگین خطای مطلق (MAE) برابر با ۴۵/۳ روز و میانگین مربعات خطا (MSE) برابر با ۵۶/۷ روز، کمترین خطا را نشان داد. این مقادیر در مقایسه با کاکس و کاپلان-مایر بیانگر دقت بالاتر پیشنهادی در پیش‌بینی مقادیر واقعی است. همچنین، MST پیش‌بینی شده توسط پیشنهادی برابر با ۳۴۲ روز بود که نسبت به مدل کاکس (۳۲۸ روز) و کاپلان-مایر (۳۱۹ روز) نزدیک‌تر به مقادیر واقعی مشاهده شده بود.

جدول ۵: مقایسه مدل پیشنهادی با سه روش نزدیک‌ترین همسایگی، شبکه عصبی مصنوعی و رگرسیون جنگل تصادفی

معیار/مدل	نزدیک‌ترین همسایگی	شبکه عصبی مصنوعی	رگرسیون جنگل تصادفی	SVR
C-Index	۰/۸۶	۰/۸۳	۰/۸۱	۰/۸۷
MAE	۴۷/۲	۴۸/۱	۴۹/۳	۴۵/۳
MSE	۵۹/۳	۵۹/۴	۶/۱	۵۶/۷
MST (روز)	۳۴۰	۳۳۹	۳۳۲	۳۴۲

بحث و نتیجه‌گیری:

در این پژوهش مشخص گردید که مدت بقا در بیماران مبتلا به سرطان معده را می‌توان با استفاده از الگوریتم‌های توضیح‌پذیر هوش مصنوعی (XAI)، نظیر SVR همراه با LIME، به‌طور قابل تفسیر پیش‌بینی کرد. مدل پیشنهادی توانست با C-Index برابر با ۰/۸۷ و مقادیر پایین‌تر MAE و MSE نسبت به روش‌های سنتی مانند کاکس و کاپلان-مایر عملکرد بهتری داشته باشد. علاوه بر این، استفاده از LIME برای تحلیل تأثیر متغیرها نشان داد که این الگوریتم امکان ارائه تفسیرهای شفاف و قوانین ساده را برای پیش‌بینی مدت بقا فراهم می‌کند. در مقایسه با روش‌های سنتی که فاقد قابلیت‌های تفسیرپذیری دقیق هستند، الگوریتم XAI پیشنهادی به‌عنوان روشی برتر برای

پیش‌بینی و تحلیل بقا در سرطان معده معرفی می‌شود. همچنین مشخص گردید الگوریتم SVR نسبت به سه الگوریتم نزدیک‌ترین همسایگی، رگرسیون جنگل تصادفی و شبکه عصبی مصنوعی عملکرد بهتر دارد. هر چند حجم نسبتاً محدود نمونه (۳۸۴ بیمار) است که می‌تواند قابلیت تعمیم نتایج به جمعیت‌های وسیع‌تر را تحت تأثیر قرار دهد. هرچند تحلیل‌های آماری با حجم نمونه فعلی انجام‌پذیر بود، اما اعتبار خارجی مدل در جمعیت‌های متنوع‌تر نیازمند بررسی در مطالعات آینده با حجم نمونه بزرگ‌تر است.

در مقایسه با مطالعه حاضر، افراش و همکاران با استفاده از الگوریتم XGBoost و انتخاب ویژگی‌های بهینه‌شده، دقت ۸۹/۱ درصد، AUC برابر با ۸۹/۳۷ درصد و f1-Score برابر



تأثیر منفی بر پیش‌بینی مدت بقا داشتند [۲۶،۲۷]. بیماران با سابقه خانوادگی بیشتر در معرض ریسک بالاتری قرار دارند که این ریسک به پیش‌بینی کوتاه‌تر بودن مدت بقا منجر می‌شود. همچنین، اعتیاد به‌عنوان یک عامل خطر دیگر در کاهش مدت بقا تأثیرگذار بود، چراکه اعتیاد ممکن است مشکلات بهداشتی اضافی ایجاد کند که روند درمان را پیچیده‌تر می‌کند [۲۷]. هرچند که در مطالعه لیو و همکاران قائل بر این است که اعتیاد بر پیش‌بینی‌های بقا تأثیر منفی نمی‌گذارد [۲۸].

پاتولوژی سرطان نیز تأثیر مهمی داشت، به‌ویژه برای بیماران مبتلا به آدنوکارسینوما که به‌طورمعمول پیش‌بینی مدت بقا بالاتری نسبت به انواع دیگر سرطان معده دارند [۲۹]. این نتیجه تأکید بر اهمیت نوع پاتولوژی در تعیین شانس بقا دارد.

در نهایت، متغیرهایی مانند تحصیلات و سن بیمار در زمان تشخیص که با دسترسی بهتر به مراقبت‌های بهداشتی و تشخیص زودتر مرتبط هستند، تأثیر مثبت داشتند [۳۰]. این امر نشان می‌دهد که عواملی فراتر از ویژگی‌های پزشکی می‌توانند در تعیین پیش‌آگهی بیماران نقش داشته باشند.

به‌طورکلی، این تحلیل‌ها اهمیت شناسایی و مدیریت متغیرهای مختلف در روند درمان و پیش‌بینی طول عمر بیماران مبتلا به سرطان معده را نشان می‌دهد و می‌تواند در بهبود تصمیم‌گیری‌های درمانی و پیش‌بینی نتایج بالینی مفید باشد.

مطالعه حاضر نشان داد که استفاده از الگوریتم رگرسیون بردار پشتیبان (SVR) در ترکیب با روش توضیح‌پذیر LIME می‌تواند پیش‌بینی نسبتاً مناسبی از مدت بقا در بیماران مبتلا به سرطان معده ارائه دهد. در مقایسه با روش‌های آماری سنتی، این مدل عملکرد بهتری از نظر شاخص C-Index و میزان خطا نشان داد. همچنین، تحلیل LIME امکان تفسیر تأثیر متغیرهای بالینی را فراهم کرد که می‌تواند در تصمیم‌گیری‌های درمانی مفید باشد. باوجوداین، به دلیل محدودیت‌هایی همچون حجم نمونه و

با ۹۰/۸ درصد را در پیش‌بینی پنج‌ساله بقای بیماران مبتلا به سرطان معده گزارش کردند [۲۱]. در مطالعه‌ای دیگر، کووایاما و همکاران با استفاده از شبکه عصبی و داده‌های بالینی معمولی، دقت ۷۶/۹ درصد برای بقای دوساله به دست آوردند [۲۲]. هرچند دقت این الگوریتم‌ها بالاست، اما تفسیرپذیری نتایج آن‌ها برای کاربرد بالینی محدود است. در مطالعه حاضر، با ترکیب مدل SVR و تحلیل LIME، امکان ارائه خروجی‌های قابل تفسیر و استخراج قوانین بالینی ساده فراهم شده که می‌تواند به تصمیم‌گیری پزشکان کمک کند، حتی اگر دقت مطلق آن کمی کمتر از برخی الگوریتم‌های پیچیده‌تر باشد.

براساس نتایج حاصل از مدل پیشنهادی و تحلیل تفسیرپذیر با استفاده از الگوریتم LIME، اهمیت متغیرهای مستقل در پیش‌بینی مدت بقا یکسان نبوده و میزان تأثیر آن‌ها با یکدیگر تفاوت دارد. همان‌گونه که در شکل ۲ مشاهده می‌شود، برخی متغیرها نظیر نوع درمان، سابقه خانوادگی سرطان معده و سطح تحصیلات نقش برجسته‌تری در تعیین پیش‌آگهی بیماران ایفا کرده‌اند، درحالی‌که متغیرهایی مانند وضعیت تأهل یا جنسیت تأثیر کمتری در مدل داشته‌اند. این تفاوت در وزن متغیرها نشان‌دهنده توانایی مدل در شناسایی روابط پیچیده و وزن دهی واقع‌بینانه به عوامل بالینی است.

یکی از مهم‌ترین متغیرهای تأثیرگذار سن در زمان تشخیص است؛ که در مطالعات چویی و همکاران و دیازدل و همکاران نیز این مسئله بیان شده است [۲۳،۲۴]. بیماران جوان‌تر معمولاً مدت بقا طولانی‌تری دارند، زیرا تشخیص زودهنگام می‌تواند شانس درمان را افزایش دهد [۲۳]. این تأثیر مثبت نشان می‌دهد که مدل به‌درستی ارتباط میان سن و پیش‌آگهی بیماری را شبیه‌سازی کرده است. همچنین، نوع درمان، به‌ویژه درمان‌های ترکیبی شامل جراحی و شیمی‌درمانی، تأثیر مثبت قابل‌توجهی بر بقا داشت. این یافته‌ها با مطالعه الضامر و همکاران نشان می‌دهند درمان‌های ترکیبی معمولاً نتیجه بهتری می‌دهند، همخوانی دارد [۲۵].

از سوی دیگر، سابقه خانوادگی سرطان معده و اعتیاد

علی برومند نیا: طراحی مفهومی مطالعه، تفسیر نتایج، بازبینی انتقادی مقاله و تأیید نهایی نسخه قابل انتشار نقش؛ امین گلاب‌پور: گردآوری داده‌ها، تحلیل آماری و تفسیر بالینی یافته‌ها، بازنگری نهایی مقاله و بررسی صحت علمی؛ سینا دامی: طراحی الگوریتم، پیاده‌سازی مدل‌های یادگیری ماشین، تحلیل نتایج و نگارش بخش فنی مقاله.

تعارض منافع

هیچ گونه تضاد منافی از سوی پژوهشگران گزارش نشده است.

تشکر و قدردانی

این مقاله برگرفته از رساله دکتری نویسنده اول در مقطع دکتری تخصصی رشته مهندسی کامپیوتر، گرایش نرم‌افزار، در دانشگاه آزاد اسلامی واحد بین‌المللی کیش با عنوان «ارائه یک الگوریتم قابل توضیح برای پیش‌بینی مدت بقا بیماران مبتلا به سرطان معده با استفاده از الگوریتم‌های ماشین بردار پشتیبان و منطق فازی» هست. بدین‌وسیله از دانشگاه آزاد اسلامی واحد بین‌المللی کیش به‌عنوان نهاد علمی حمایت‌کننده این پایان‌نامه و همچنین از دانشگاه علوم پزشکی شاهرود بابت فراهم‌سازی داده‌های بیماران، صمیمانه قدردانی می‌شود.

تمرکز داده‌ها بر یک مرکز درمانی، اعتبارسنجی خارجی مدل در مطالعات آینده ضروری به نظر می‌رسد. به‌کارگیری چنین مدل‌هایی در محیط‌های بالینی، مستلزم ارزیابی‌های بیشتر در جمعیت‌های مختلف و شرایط عملیاتی متنوع است.

نتیجه‌گیری:

این مطالعه نشان داد که مدل ترکیبی SVR-LIME با دقت پیش‌بینی بالا ($C\text{-Index}=0/87$) و تفسیرپذیری مناسب، ابزار مؤثری برای تحلیل بقا در سرطان معده است. یافته‌های کلیدی شامل:

عملکرد برتر: SVR نسبت به ANN، KNN و RF در معیارهای ارزیابی $MAE=45/3$ ، $MSE=56/7$ عملکرد بهتری نشان داد.

عوامل کلیدی: شناسایی متغیرهای تأثیرگذار مانند سن تشخیص $\beta=-0/25$ ، درمان ترکیبی $\beta=+0/30$ و پاتولوژی $\beta=+0/40$ که تأییدکننده مطالعات قبلی است.

کاربرد بالینی: قوانین تفسیرپذیر استخراج شده (مثل تأثیر مثبت درمان ترکیبی بر بقا) امکان تصمیم‌گیری شخصی‌سازی‌شده را فراهم می‌کند.

ملاحظات اخلاقی:

پیروی از اصول اخلاق در پژوهش

این مطالعه دارای تاییدیه اخلاقی به شماره IR.SHMU. REC.1403.182 از دانشگاه علوم پزشکی شاهرود است.

حامی مالی

این مقاله از طرف هیچ‌گونه نهاد یا موسسه‌ای حمایت مالی نشده و تمام منابع مالی آن از طرف نویسنده اول یا نویسندگان تأمین‌شده است.

مشارکت نویسندگان

لیلا علائی: طراحی مطالعه، جمع‌آوری و تحلیل داده‌ها، نگارش پیش‌نویس اولیه و بازنگری محتوای علمی مقاله؛

References

1. Bandeira-de Oliveira M, Aparicio-González T, Del Cura-González I, Suárez-Fernández C, Rodríguez-Barrientos R, et al. Adjusted morbidity groups and survival: A retrospective cohort study of primary care patients with chronic conditions. *BMC Prim Care*. 2023; 24(1):103. DOI: [10.1186/s12875-023-02059-9](https://doi.org/10.1186/s12875-023-02059-9) PMID: [37081395](https://pubmed.ncbi.nlm.nih.gov/37081395/)
2. Jaul E, Rosenzweig JP. A retrospective study of the impact of pressure Ulcers on Survival in elderly persons with chronic diseases. *Ostomy Wound Manage*. 2017; 63(5):26-32. PMID: [28570246](https://pubmed.ncbi.nlm.nih.gov/28570246/)
3. Pradhan SP, Agarwala S, Sahoo J, Pradhan SK, Jena S, Satpathy N, et al. Survival of gastric cancer patients at a tertiary care Hospital in Eastern India: A retrospective data analysis. *Cureus*. 2023; 15(4):e37064. DOI: [10.7759/cureus.37064](https://doi.org/10.7759/cureus.37064) PMID: [37153312](https://pubmed.ncbi.nlm.nih.gov/37153312/)
4. Mahawar R, S Devi Y, D Devi N, Panmei R, Babu B M, Malsawmtluanga B. A five-year retrospective analysis of clinical, pathological and treatment aspects on stomach cancer from a regional cancer centre in north east India. *IJRMS*. 2023;11(11):4165-9. DOI: [10.18203/2320-6012.ijrms20233393](https://doi.org/10.18203/2320-6012.ijrms20233393)
5. Ajibawo T. Trends and disparities in the incidence and mortality of stomach cancer in the United States. *American Society of Clinical Oncology*. 2024; 42(16):. DOI: [10.1200/JCO.2024.42.16_suppl.e16102](https://doi.org/10.1200/JCO.2024.42.16_suppl.e16102)
6. Cheng M, Li S. Lifestyle and dietary patterns as risk factors affect gastric cancer in Iran. *Highl Sci Eng Technol*. 2022; 2:82-9. DOI: [10.54097/hset.v2i.558](https://doi.org/10.54097/hset.v2i.558)
7. Kalan Farmanfarma K, Mahdaviifar N, Hassanipour S, Salehiniya H. Epidemiologic study of Gastric Cancer in Iran: A Systematic Review. *Clin Exp Gastroenterol*. 2020; 13:511-42. DOI: [10.2147/CEG.S256627](https://doi.org/10.2147/CEG.S256627) PMID: [33177859](https://pubmed.ncbi.nlm.nih.gov/33177859/)
8. Lin Z, Wang R, Zhou Y, Wang Q, Yang CY, Hao BC, et al. Prediction of distant metastasis and survival prediction of gastric cancer patients with metastasis to the liver, lung, bone, and brain: research based on the SEER database. *Ann Transl Med*. 2022; 10(1):16. DOI: [10.21037/atm-21-6295](https://doi.org/10.21037/atm-21-6295) PMID: [35242861](https://pubmed.ncbi.nlm.nih.gov/35242861/)
9. Deo SV, Deo V, Sundaram V. Survival analysis-part 2: Cox proportional hazards model. *Indian J Thorac Cardiovasc Surg*. 2021; 37(2):229-33. DOI: [10.1007/s12055-020-01108-7](https://doi.org/10.1007/s12055-020-01108-7) PMID: [33642726](https://pubmed.ncbi.nlm.nih.gov/33642726/)
10. Etikan I, Abubakar S, Alkassim R. The Kaplan-Meier estimate in survival analysis. *Biom Biostat Int J*. 2017; 5(2):55-9. DOI: [10.15406/bbij.2017.05.00128](https://doi.org/10.15406/bbij.2017.05.00128)
11. Martínez Pérez JA, Pérez Martínez PS. Survival analysis. *Semergen*. 2023; 49(5):101986. Spanish. DOI: [10.1016/j.semerg.2023.101986](https://doi.org/10.1016/j.semerg.2023.101986) PMID: [37182235](https://pubmed.ncbi.nlm.nih.gov/37182235/)
12. Zhang Y, Wittrup E, Najarian K. CoxFNN: Interpretable machine learning method for survival analysis. *Annu Int Conf IEEE Eng Med Biol Soc*. 2024; 2024:1-5. DOI: [10.1109/EMBC53108.2024.10781910](https://doi.org/10.1109/EMBC53108.2024.10781910) PMID: [40031518](https://pubmed.ncbi.nlm.nih.gov/40031518/)
13. Jiang F, Zhao G, Rodriguez-Monguio R, Ma Y. Causal effect estimation in survival analysis with high dimensional confounders. *Biometrics*. 2024; 80(4):ujae110. DOI: [10.1093/biomtc/ujae110](https://doi.org/10.1093/biomtc/ujae110) PMID: [39400258](https://pubmed.ncbi.nlm.nih.gov/39400258/)
14. Potluri D, Vatluri S, Reddy G, Rathod S, Ramesh D, Devi K. Application of Machine Learning Techniques Models for Forecasting of Redgram Prices of Andhra Pradesh, India. *J Sci Res Rep*. 2024; 30:252-71. DOI: [10.9734/jsrr/2024/v30i72142](https://doi.org/10.9734/jsrr/2024/v30i72142)
15. Rožman M, Kišić A, Oreški D. Comparative analysis of nonlinear models developed using machine learning algorithms. *WSEAS Trans Inf Sci Appl*. 2024; 21(29):303-7. DOI: [10.37394/23209.2024.21.29](https://doi.org/10.37394/23209.2024.21.29)
16. King I, Meng H, Lam TY. Chapter 5- Black box medicine. In *Artificial Intelligence in Medicine* Academic Press; 2024. p.59-69. DOI: [10.1016/B978-0-323-95068-8.00005-4](https://doi.org/10.1016/B978-0-323-95068-8.00005-4)
17. Chennam KK, Mudrakola S, Maheswari VU, Aluvalu R, Rao KG. Black box models for explainable artificial intelligence. In *Explainable AI: foundations, methodologies and applications 2022 Oct 20* (pp. 1-24). Cham: Springer International Publishing. DOI: [10.1007/978-3-031-12807-3_1](https://doi.org/10.1007/978-3-031-12807-3_1)
18. Mutalib NH, Sabri AQ, Wahab AW, Abdullah ER, AlDahoul N. Explainable deep learning approach for advanced persistent threats (APTs) detection in cybersecurity: A review. *Artif Intell Rev*. 2024; 57(11):297. DOI: [10.1007/s10462-](https://doi.org/10.1007/s10462-)



024-10890-4

19. Awad M, Khanna R. Efficient learning machines: theories, concepts, and applications for engineers and system designers. Springer Nature; 2015. p.67-80. DOI: [10.1007/978-1-4302-5990-9](https://doi.org/10.1007/978-1-4302-5990-9)
20. Riccò B, Martinelli G, Bardasi C, Dominici M, Spallanzani A, Salati M. Optimizing the continuum of care in Gastric Cancer. *Onco Targets Ther.* 2023; 16:995-1012. DOI: [10.2147/OTT.S365505](https://doi.org/10.2147/OTT.S365505) PMID: 38021446
21. Afrash MR, Mirbagheri E, Mashoufi M, Kazemi-Arpanahi H. Optimizing prognostic factors of five-year survival in gastric cancer patients using feature selection techniques with machine learning algorithms: A comparative study. *BMC Med Inform Decis Mak.* 2023; 23(1):54. DOI: [10.1186/s12911-023-02154-y](https://doi.org/10.1186/s12911-023-02154-y) PMID: 37024885
22. Kuwayama N, Hoshino I, Mori Y, Yokota H, Iwatate Y, Uno T. Applying artificial intelligence using routine clinical data for preoperative diagnosis and prognosis evaluation of gastric cancer. *Oncol Lett.* 2023; 26(5):499. DOI: [10.3892/ol.2023.14087](https://doi.org/10.3892/ol.2023.14087) PMID: 37854867
23. Díaz Del Arco C, Ortega Medina L, Estrada Muñoz L, Molina Roldán E, García Gómez de Las Heras S, et al. Impact of age at diagnosis on clinicopathological features, prognosis, and management of Gastric Cancer: A retrospective single-center Experience from Spain. *Cancers (Basel).* 2023; 15(17):4241. DOI: [10.3390/cancers15174241](https://doi.org/10.3390/cancers15174241) PMID: 37686517
24. Choi Y, Kim N, Kim KW, Jo HH, Park J, Yoon H, et al. Gastric cancer in older patients: A retrospective study and literature review. *Ann Geriatr Med Res.* 2022; 26(1):33-41. DOI: [10.4235/agmr.21.0144](https://doi.org/10.4235/agmr.21.0144) PMID: 35298880
25. Al Zamer DS, Theeb L, Al-Zamer YS, Hammad AM, Rawashdeh B. 182P Cytoreductive surgery and chemotherapy in metastatic gastric adenocarcinoma: A population-based study. *Ann Oncol.* 2023; 34(4):S1545-6. DOI: [10.1016/j.annonc.2023.10.761](https://doi.org/10.1016/j.annonc.2023.10.761)
26. Pühr HC, Berchtold L, Zingerle L, Felfernig M, Weissenbacher L, Jomrich G, et al. Association of family history with patient characteristics and prognosis in a large European gastroesophageal cancer cohort. *Wien Klin Wochenschr.* 2025; 137(7-8):214-23. DOI: [10.1007/s00508-024-02432-3](https://doi.org/10.1007/s00508-024-02432-3) PMID: 39235615
27. Ershov PV, Veselovsky EM, Konstantinova Yu S. Role of heredity, endogenous and exogenous factors in Gastric Cancer. *Extreme Med.* 2020; 4(22):67-79. DOI: [10.47183/mes.2020.023](https://doi.org/10.47183/mes.2020.023)
28. Liu X, Cai H, Yu L, Huang H, Long Z, Wang Y. Prognostic significance of cancer family history for patients with gastric cancer: A single center experience from China. *Oncotarget.* 2016; 7(24):37305-18. DOI: [10.18632/oncotarget.9032](https://doi.org/10.18632/oncotarget.9032) PMID: 27127887
29. Leijonmarck W, Mattsson F, Lagergren J. Survival among patients cured from gastric adenocarcinoma compared to the background population. *Gastric Cancer.* 2024; 27(6):1180-88. DOI: [10.1007/s10120-024-01545-y](https://doi.org/10.1007/s10120-024-01545-y) PMID: 39230776
30. Smith O, Alkhaffaf B. OGC P42 presentation, management and outcomes in young patients diagnosed with Gastric Cancer. *BJS.* 2023; 110(Supplement_8):znad240-348. DOI: [10.1093/bjs/znad348.240](https://doi.org/10.1093/bjs/znad348.240)