

RESEARCH

Open Access



# Integrated clinical and genomic models using machine-learning methods to predict the efficacy of paclitaxel-based chemotherapy in patients with advanced gastric cancer

Yonghwa Choi<sup>1,2</sup>, Jangwoo Lee<sup>3,4</sup>, Keewon Shin<sup>4</sup>, Ji Won Lee<sup>5</sup>, Ju Won Kim<sup>5</sup>, Soohyeon Lee<sup>5</sup>, Yoon Ji Choi<sup>5</sup>, Kyong Hwa Park<sup>5</sup> and Jwa Hoon Kim<sup>5\*</sup>

## Abstract

**Background** Paclitaxel is commonly used as a second-line therapy for advanced gastric cancer (AGC). The decision to proceed with second-line chemotherapy and select an appropriate regimen is critical for vulnerable patients with AGC progressing after first-line chemotherapy. However, no predictive biomarkers exist to identify patients with AGC who would benefit from paclitaxel-based chemotherapy.

**Methods** This study included 288 patients with AGC receiving second-line paclitaxel-based chemotherapy between 2017 and 2022 as part of the K-MASTER project, a nationwide government-funded precision medicine initiative. The data included clinical (age [young-onset vs. others], sex, histology [intestinal vs. diffuse type], prior trastuzumab use, duration of first-line chemotherapy), and genomic factors (pathogenic or likely pathogenic variants). Data were randomly divided into training and validation sets (0.8:0.2). Four machine learning (ML) methods, namely random forest (RF), logistic regression (LR), artificial neural network (ANN), and ANN with genetic embedding (ANN with GE), were used to develop the prediction model and validated in the validation sets.

**Results** The median patient age was 64 years (range 25–91), and 65.6% of those were male. A total of 288 patients were divided into the training ( $n=230$ ) and validation ( $n=58$ ) sets. No significant differences existed in baseline characteristics between the training and validation sets. In the training set, the areas under the ROC curves (AUROC) for predicting better progression-free survival (PFS) with paclitaxel-based chemotherapy were 0.499, 0.679, 0.618, and 0.732 in the RF, LR, ANN, and ANN with GE models, respectively. The ANN with the GE model that achieved the highest AUROC recorded accuracy, sensitivity, specificity, and F1-score performance of 0.458, 0.912, 0.724, and 0.579, respectively. In the validation set, the ANN with GE model predicted that paclitaxel-sensitive patients had significantly longer PFS (median PFS 7.59 vs. 2.07 months,  $P=0.020$ ) and overall survival (OS) (median OS 14.70 vs. 7.50 months,  $P=0.008$ ). The LR model predicted that paclitaxel-sensitive patients showed a trend for longer PFS (median PFS 6.48 vs. 2.33 months,  $P=0.078$ ) and OS (median OS 12.20 vs. 8.61 months,  $P=0.099$ ).

\*Correspondence:

Jwa Hoon Kim  
jhoonkim@korea.ac.kr

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Conclusions** These ML models, integrated with clinical and genomic factors, offer the possibility to help identify patients with AGC who may benefit from paclitaxel chemotherapy.

**Keywords** Advanced gastric cancer, Paclitaxel, Machine learning, Prediction, Artificial neural network

## Background

Over the past decades, fluoropyrimidines (5-fluorouracil, capecitabine, and S-1), platinum (cisplatin and oxaliplatin), taxanes (docetaxel and paclitaxel), and irinotecan have demonstrated survival benefits for the treatment of patients with unresectable or metastatic gastric cancer. Fluoropyrimidine- or platinum-based regimens are widely accepted first-line therapies for patients with advanced gastric cancer (AGC) [1]. Since the REGARD and RAINBOW studies [2, 3], a combination of ramucirumab, a monoclonal antibody targeting vascular endothelial growth factor receptor-2, and paclitaxel has been widely used, and irinotecan has been recommended as a second- or later-line treatment [4, 5]. Recently, novel treatment strategies, including immune checkpoint inhibitors and new targeted inhibitors, have improved the survival of patients with AGC [6–10].

The proportion of patients receiving second- or later-line treatment, along with response and survival rates, have progressively decreased than those observed in first-line treatment [1]. Certain patients may experience clinical deterioration with rapid progression, resulting in missed opportunities for further treatment. Patient fragility, stemming from prior chemotherapy exposure and various disease characteristics, could contribute to this phenomenon. Several factors, such as poor performance status or cumulative toxicity due to the first-line chemotherapy, extent of disease, and history of agents used as the first-line therapy, could influence whether a patient benefits from further treatment [11–14]. Issues have been continuously raised to identify patients who are more likely to benefit from second- or later-line therapy, especially vulnerable patients with AGC.

Through recent next-generation sequencing (NGS), molecular classification of heterogeneous AGC has become more important and its prognostic significance with chemotherapy efficacy is well known [15]. Specifically, taxanes are considered to exhibit anti-cancer effects through aberrant stabilization of microtubules, causing defects in chromosome segregation, mitotic arrest and activation of the spindle assembly checkpoint, where prolonged activation results in cell death. There were previous studies suggesting that altered expression of genes involved in the spindle assembly checkpoint may affect cellular sensitivity to paclitaxel [16–18]. However, there are still no definite predictive biomarkers for each palliative chemotherapy in AGC.

Machine learning (ML), a form of artificial intelligence (AI), is widely used and has great potential in precision

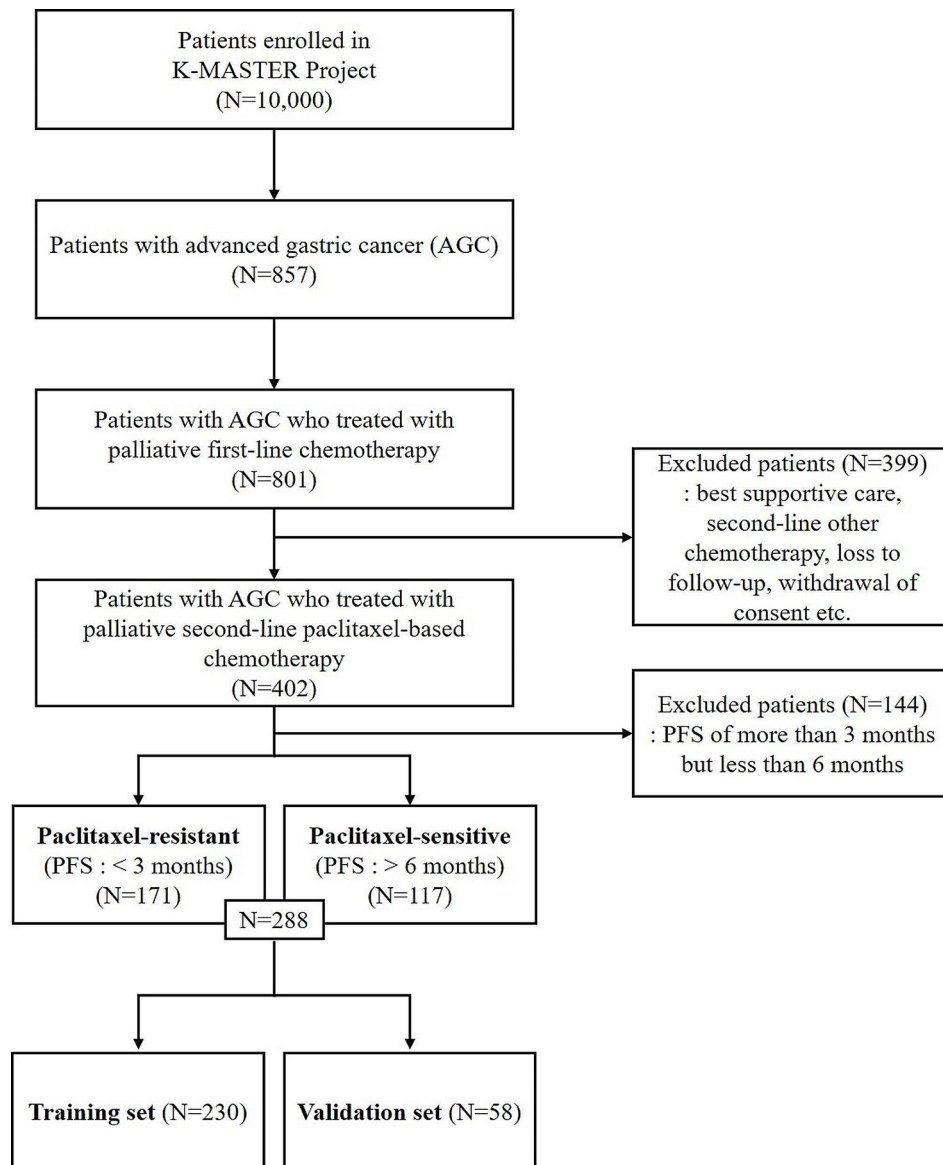
oncology. Random forest (RF) utilizes multiple decision trees trained on random subsets of data to collectively make predictions for classification or regression tasks. Each tree independently learns the optimal feature splits, and the final prediction is determined by aggregating the outputs from these trees. Logistic regression (LR) is a statistical method used for binary classification that estimates the probability of a binary outcome. It models the relationship between one or more independent variables and a dependent variable using a logistic function, transforming the inputs into probabilities between zero and one. An artificial neural network (ANN) is a computational model consisting of interconnected nodes, called neurons, organized in layers to process information. Through training, ANNs adjust the connections between neurons to learn patterns and make predictions based on the data. Earlier studies have attempted to predict overall survival (OS) and disease-free survival in patients with gastric cancer and the benefits of adjuvant chemotherapy using ML-based methods [18, 19]. Recently, various methods for generating continuously distributed representations of words, for example, Word2Vec [20], have been introduced for joint use with ANN-based machine learning techniques. Similar attempts to represent genetic mutations or protein sequences in a continuous vector space have been made in the biomedical domain [21, 22], showing remarkable improvements in the ability to capture the characteristics of proteins or relationships between mutations.

This study aimed to develop a prediction model to identify patients with AGC who would benefit from paclitaxel-based chemotherapy after failure of fluoropyrimidine and platinum-based chemotherapy.

## Materials and methods

### Patients and K-MASTER datasets

Patients eligible for the study were 20 years of age or older, diagnosed with metastatic or recurrent stomach adenocarcinoma through histological or cytological methods, following the unsuccessful treatment with first-line fluoropyrimidine- and platinum-based chemotherapy. These individuals participated in the second-line, paclitaxel-based chemotherapy as part of the K-MASTER project from 2017 to 2022 (Fig. 1) [23]. The K-MASTER initiative, a comprehensive precision medicine trial across 51 Korean institutions, focused on identifying treatable mutations through Next-Generation Sequencing (NGS) in 10,000 Korean patients with advanced solid tumors, and led to the strategic enrollment of patients in



**Fig. 1** Patient flow diagram

clinical trials tailored to their genetic profiles [23]. Additionally, the K-MASTER involved a nationwide effort to map out genomic profiles and systematically gather data on common clinical attributes across various solid tumors [23].

#### Clinical and genetic features

Clinical data included age, sex, tumor histology (intestinal vs. diffuse), prior trastuzumab use, and duration of first-line chemotherapy. The clinical utility of NGS in the K-MASTER project was previously reported [23–25]. The dataset included molecular changes, such as single nucleotide variants, insertions, deletions, copy number variations, and structural variants, all of which have the potential to influence clinical decision-making

(Additional file 1). These genetic alterations were classified within databases as either “likely-pathogenic” or “pathogenic” according to COSMIC and ClinVar, and as “likely-oncogenic” or “oncogenic” according to OncoKB databases.

#### Preprocess data

Participants in this study were divided into two groups based on their response to second-line paclitaxel-based chemotherapy: those who experienced progression-free survival (PFS) of more than six months were considered paclitaxel-sensitive, while individuals with a PFS of less than three months were deemed paclitaxel-resistant. Following this classification, the cohort was then randomly split into training and validation datasets in an 80:20

ratio, utilizing the ‘StratifiedShuffleSplit’ function from the scikit-learn library to maintain an equal distribution of outcomes across both datasets (Fig. 1).

Clinical and genetic information was subsequently converted into binary form to facilitate analysis, resulting in distinct sets of binary clinical and genetic features. For genetic embeddings, each patient’s genetic variants were first assigned to a vector space, creating vector representations for individual variants. These vectors were aggregated to form a comprehensive genetic feature for each patient. These vector representations were initially set at random values and tuned throughout the training phase.

**Development and validation of ML models in training and validation sets**

Four ML models – Random Forest (RF), Logistic Regression (LR), Artificial Neural Network (ANN), and ANN incorporating genetic embedding (ANN with GE) – were employed on the training sets and then tested on the validation sets. Patient data was synthesized by combining clinical binary features and genetic information into comprehensive feature vectors. For the LR, RF, and standard ANN models, genetic information was represented through binary genetic features. Conversely, the ANN with GE model utilized vectors of embedded genetic features. These comprehensive patient feature vectors were then used to train each respective model (Fig. 2). The genetic embedding dimension was set at 20, and both ANN configurations included a single hidden layer with 20 nodes. The effectiveness of the models, particularly in predicting patient outcomes following second-line paclitaxel-based chemotherapy, was measured by the area under the Receiver Operating Characteristics (ROC)

curve (AUROC), focusing on the progression-free survival (PFS) duration.

Model development and training processes were executed using Python version 3.9.12. The LR, RF, and standard ANN models used tools from the scikit-learn package version 1.1.1, specifically ‘LogisticRegression,’ ‘RandomForestClassifier,’ and ‘MLPClassifier,’ respectively. The ANN with GE model was developed using functionalities from the PyTorch package, version 1.13.0.

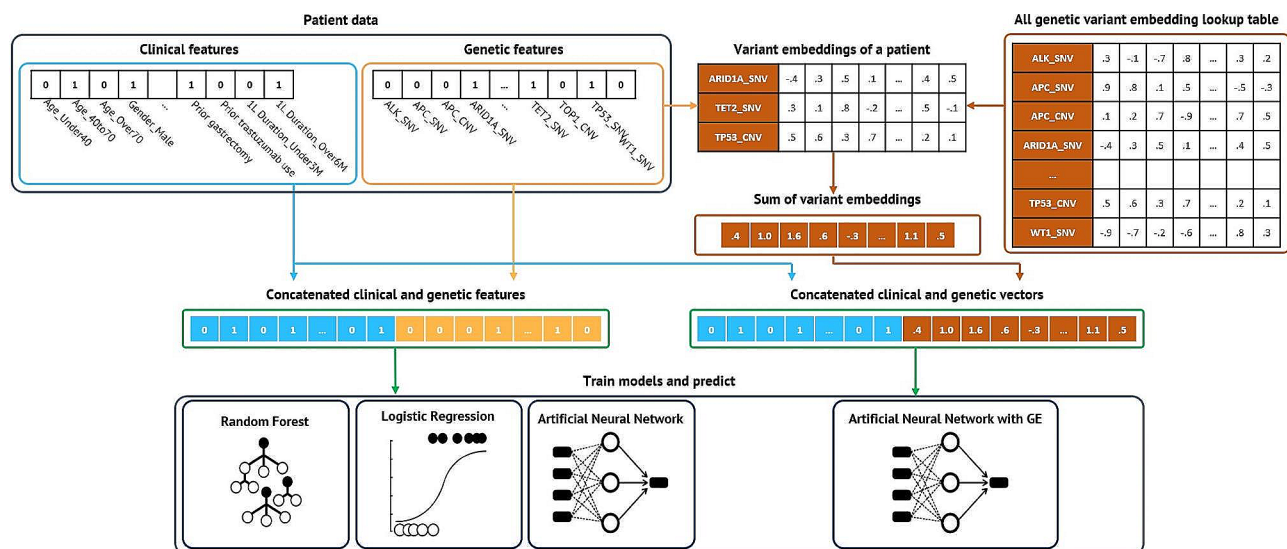
**Statistical analysis**

All statistical analyses, along with the development of prediction models, were conducted using Python (version 3.9.12) with the scikit-learn (version 1.1.1) and lifelines (version 0.27.7) packages. PFS was measured from the start of second-line paclitaxel-based chemotherapy until the occurrence of progression or death from any cause. OS was also calculated from the commencement of the same chemotherapy until death due to any cause. Survival rates were determined using the Kaplan-Meier method, and differences between survival curves were assessed with the log-rank test. Statistical significance was established at a *P*-value of less than 0.05, using a two-sided test.

**Results**

**Patient characteristics**

A total of 288 patients with AGC were treated with second-line paclitaxel-based chemotherapy between 2017 and 2022 (Fig. 1). The median age was 63 years (range 25–91), and 65.6% of patients were male. First-line chemotherapy included fluoropyrimidine- and platinum-based regimes, with trastuzumab use observed in 16.7% of the patients. Second-line chemotherapy comprised



**Fig. 2** Structures of random forest, logistic regression, artificial neural network, and artificial neural network with genetic embedding

**Table 1** Baseline characteristics

	Training set (N=230, %)	Validation set (N=58, %)	P
Age (y)			
Median, years (range)	64 (25–91)	62 (28–89)	0.154
< 40 years, n (%)	8 (3.5)	5 (8.6)	0.153
≥ 70 years, n (%)	74 (32.2)	14 (24.1)	
Sex			0.209
Male	155 (67.4)	34 (58.6)	
Female	75 (32.6)	24 (41.4)	
Histology			0.146
Intestinal type	143 (62.1)	42 (72.4)	
Diffuse type*	87 (37.8)	16 (27.6)	
Clinical settings			0.835
Initially metastatic disease	154 (67.0)	38 (65.6)	
Recurrent disease	76 (33.0)	20 (34.5)	
Prior gastrectomy	120 (52.2)	26 (44.8)	0.353
Prior trastuzumab use	40 (17.4)	8 (13.8)	0.511
Duration of prior 1st -line fluoro- pyrimidine plus platinum- based chemotherapy			0.508
< 3 months	57 (24.8)	16 (27.6)	
≥ 6 months	87 (37.8)	19 (32.7)	

\* Mixed type was categorized into diffuse type: training set (n=13) and validation set (n=1).

**Table 2** Baseline characteristics between paclitaxel-sensitive and-resistant patients in the training set

	Paclitaxel- sensitive (N=93, %)	Paclitaxel- resistant (N=137, %)	P
Age (y)			
Median, years (range)	64 (37–87)	64 (25–91)	0.910
< 40 years, n (%)	3 (3.2)	5 (3.6)	0.833
≥ 70 years, n (%)	28 (30.1)	46 (33.6)	
Sex			0.116
Male	57 (61.3)	98 (71.5)	
Female	36 (28.5)	39 (28.5)	
Histology			0.072
Intestinal type	51 (54.8)	92 (67.2)	
Diffuse type	42 (45.2)	45 (32.8)	
Clinical settings			0.132
Initially metastatic disease	57 (61.3)	97 (70.8)	
Recurrent disease	36 (38.7)	40 (29.2)	
Prior gastrectomy	56 (60.2)	64 (46.7)	0.059
Prior trastuzumab use	20 (21.5)	20 (14.6)	0.215
Duration of prior 1st -line fluoropyrimidine plus platinum- based chemotherapy			0.034
< 3 months	17 (18.3)	40 (29.2)	
≥ 6 months	44 (47.3)	43 (31.4)	

paclitaxel+ramucirumab administration to 237 patients (82.3%), while the other patients were treated with paclitaxel alone or paclitaxel+others in clinical trials. The patients were divided into training (n=230) and validation sets (n=58). There were no significant differences in baseline characteristics between the training and validation sets (Table 1). All pathogenic variants observed in all patients were used as genetic features, comprising 73 SNVs and 29 CNVs for 87 genes.

At a median follow-up duration of 19.07 months (95% confidence interval [CI], 15.947–22.193), median PFS and OS were 2.70 months (95% CI, 2.364–3.036) and 13.28 months (95% CI, 10.271–16.289), respectively. Comparisons between the training and validation sets revealed no notable differences in PFS and OS following second-line paclitaxel-based chemotherapy. Specifically, the median PFS was 2.53 months in the training set versus 2.79 months in the validation set (P=0.911), and the median OS was 13.61 months in the training set versus 10.45 months in the validation set (P=0.280).

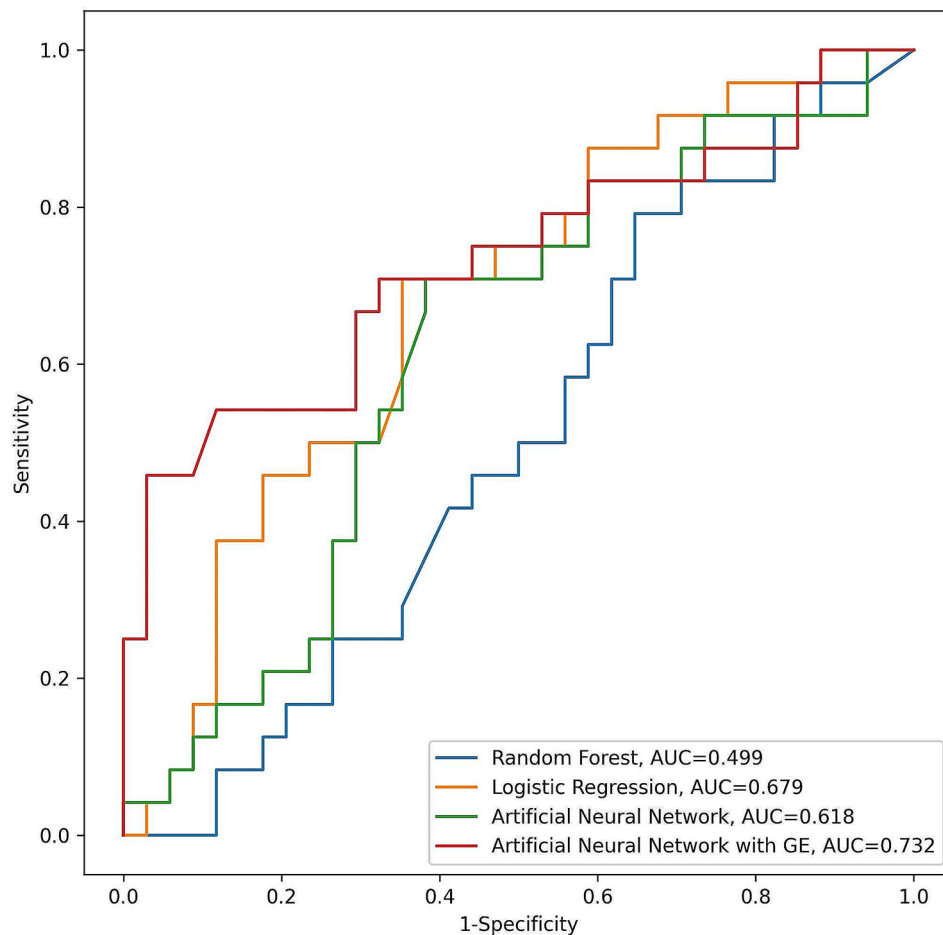
#### Development of four ML-based prediction models

The baseline characteristics of the paclitaxel-sensitive (n=93) and paclitaxel-resistant (n=137) patients in the training set showed no substantial variations (Table 2). The only exception was the length of prior first-line chemotherapy. Specifically, patients with paclitaxel-sensitive advanced gastric cancer (AGC) experienced a significantly longer duration of initial chemotherapy compared to those who were paclitaxel-resistant (47.3% versus 31.4%, P=0.034).

The AUROC scores for predicting paclitaxel-sensitive patients varied across different models: 0.499 (95% CI 0.378–0.626) for the RF, 0.679 (95% CI 0.562–0.798) for the LR, 0.597 (95% CI 0.475–0.722) for the ANN, and 0.732 (95% CI 0.610–0.842) for the ANN with E models (Fig. 3). The sensitivity, specificity, accuracy, and F1 scores for these models are detailed in Table 3. Among these, the ANN with GE model demonstrated the highest effectiveness with an AUROC of 0.732, whereas the RF model was the least effective, recording an AUROC of 0.499.

#### Validation of four ML-based prediction models

In the validation sets, the RF model was unable to effectively predict a longer PFS for paclitaxel-sensitive patients compared to paclitaxel-resistant ones, with median PFS figures of 1.51 vs. 2.79 months, respectively (P=0.075) (Fig. 4a). Conversely, the LR model suggested a trend towards longer PFS for paclitaxel-sensitive patients (median PFS 6.48 vs. 2.33 months, P=0.078) (Fig. 4b), while the ANN model indicated a non-significant numerical advantage in PFS for paclitaxel-sensitive patients over paclitaxel-resistant patients (median PFS



**Fig. 3** Receiver operating characteristic curves for progression-free survival

**Table 3** Performance metrics of machine learning models to predict the progression-free survival of second-line paclitaxel in patients with advanced gastric cancer

Models	AUROC	Sensitivity	Specificity	Accuracy	F1 score
RF	0.499	0.517	0.417	0.588	0.417
LR	0.679	0.638	0.375	0.823	0.461
ANN	0.618	0.500	0.706	0.621	0.522
ANN with GE	0.732	0.458	0.912	0.724	0.579

RF, random forest; LR, logistic regression; ANN, artificial neural network; ANN with GE, artificial neural network with genetic embedding

6.38 vs. 2.33 months,  $P=0.719$ ) (Fig. 4c). The ANN with GE model was the only one to significantly predict longer PFS for paclitaxel-sensitive patients (median PFS 7.59 vs. 2.07 months,  $P=0.020$ ) (Fig. 4d).

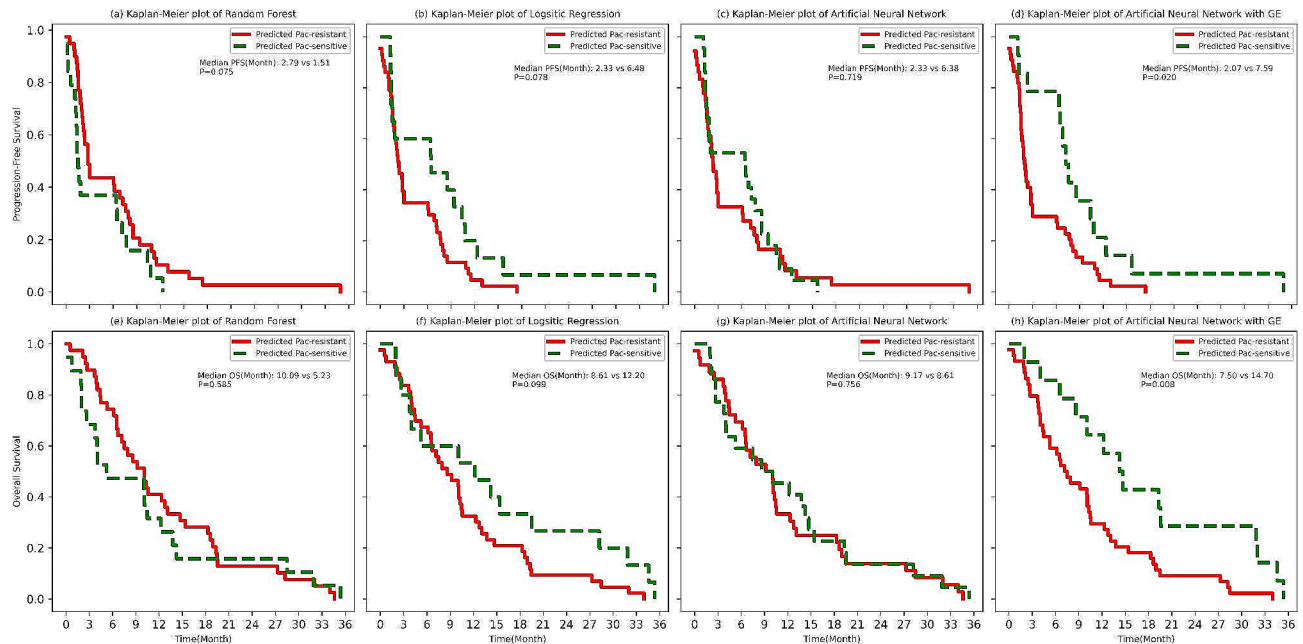
Regarding OS, no significant differences were noted between paclitaxel-sensitive and paclitaxel-resistant patients in both the RF and ANN models (Fig. 4e and g). The LR model, however, showed a trend towards longer OS for paclitaxel-sensitive patients (median OS 12.20 vs. 8.61 months,  $P=0.099$ ) (Fig. 4f). Consistently, the ANN with GE model predicted a significant extension in OS

for paclitaxel-sensitive patients compared to their resistant counterparts (median OS 14.70 vs. 7.50 months,  $P=0.008$ ) (Fig. 4h).

## Discussion

This study showed that integrated clinical and genomic models could predict which patients with AGC are more likely to benefit from second-line paclitaxel-based chemotherapy. Among the four ML-based models, the best model was the ANN with GE model, which significantly predicted paclitaxel-sensitive or paclitaxel-resistant patients with AGC. Our ANN with the GE model aggregated the embedded genetic variants with clinical features, followed by a feed-forward neural network. The main strength of this study was that the results arose from a prospectively collected database of the K-MAS-TER project, a nationwide program that has maintained high-quality genomic profiling, and the relatively large size of comprehensive datasets, including both clinical and NGS data [23].

Clinical decisions to proceed with further treatment and to choose the optimal chemotherapy regimen are



**Fig. 4** Kaplan–Meier curves of progression-free survival and overall survival in validation sets according to machine-learning methods

always challenging. Most patients with AGC become more fragile, especially after the failure of first-line chemotherapy. Additionally, chemotherapy occasionally worsens the clinical condition owing to toxicity without providing benefits. Clinical deterioration with ineffective chemotherapy may lead to a loss of chance for subsequent treatment. Currently, there are no predictive models or systems capable of determining the potential benefits of palliative chemotherapy for patients with cancer or identifying the most effective chemotherapy regimens. In the era of AI, ML-based models can be used as clinical decision-support systems [26]. Owing to technological advances in genomic profiling, NGS testing is a routine workup in oncology, and the incorporation of complicated NGS results into clinical decisions is important for individualized therapy. In addition to genomic data, numerous clinical factors must be considered when making clinical decisions. It can easily organize and interpret data from clinical practice. Thus, our ML models can serve as important backbones for future clinical decision support systems.

It is interesting to note that recently, using ML methods, the first and largest performed study identified a gene signature predictive of paclitaxel benefit in GC from the phase 3 SAMIT trial evaluating adjuvant chemotherapy [18]. A custom-designed NanoString panel, including genes involved in chromosomal stability or immunogenic cell death, was used, and an ML model identified a gene signature for predicting paclitaxel benefit [18]. Similarly, our study predicted the benefits of paclitaxel in patients with AGC using ML methods. However, an important point of consideration is that our study focused on

palliative settings, which use paclitaxel as standard chemotherapy, and more caution is warranted to balance the benefits and toxicity in vulnerable patients. Our ML models integrated clinical factors and genomic data, where NGS tests can be routinely performed rather than gene signature analyses in real-world practice.

Previously, prognostic factor analysis using pooled data from two pivotal phase 3 trials evaluating second-line ramucirumab alone or ramucirumab+paclitaxel chemotherapy found 12 independent factors for poor survival, including several clinical and laboratory findings [27]. Another retrospective study similarly reported the prognostic significance of clinicolaboratory factors associated with second-line chemotherapy efficacy [12]. Good performance status and a long duration of prior first-line chemotherapy are commonly associated with better survival. However, genomic data reflecting the underlying tumor biology were not analyzed in either study. Although our study did not evaluate laboratory findings, genomic alterations were comprehensively and in-depth considered when developing the prediction models. Future research is required to expand and update our models based on various factors, including laboratory findings.

This study had few limitations. First, as a retrospectively designed study using an already established dataset, additional factors could not be analyzed. Second, there may be selection bias between the training and validation sets, despite being randomly assigned and having no statistically significant differences between them. Third, although internal validation was performed, as the small size of the validation set may not guarantee the

generalizability of models, our ML models must be validated using an external independent dataset. Fourth, a prospective clinical trial is required to confirm the clinical utility of these prediction models. Finally, ML models should be incorporated into real-world practices and advanced independently if true AI-based models are used.

## Conclusions

Our ML models integrated clinical and genomic factors and identified patients with AGC with a greater likelihood of benefit from second-line paclitaxel chemotherapy. This study provides the foundation for future advanced prediction ML models.

## Abbreviations

AGC	advanced gastric cancer
AUROC	area under the receiver operating characteristic curve
ML	machine learning
RF	random forest
LR	logistic regression
ANN	artificial neural network
ANN	artificial neural network with genetic embedding
HER2	human epidermal growth factor receptor 2
PFS	progression-free survival
OS	overall survival
NGS	next-generation sequencing
CNV	copy-number variation
SNV	single nucleotide variant

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-024-12268-9>.

Supplementary Material 1

## Acknowledgements

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HR22C1302).

## Author contributions

Study concepts and design: JHK; Data acquisition: JHK, JWL, SL, YJC, KWP; Quality control of data and algorithms: JHK, YC, JL, KS, SL, YJC, KWP; Data analysis and interpretation: JHK, YC, JL; Statistical analysis: JHK, YC, JL, KS; Manuscript preparation: JHK, YC; Manuscript review: JHK, YC, JL, KS, JWK, JWK, SL, YJC, KWP.

## Funding

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HR22C1302).

## Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author (JHK) upon request. The data are not publicly available because they contain information that can compromise the participants' privacy/consent.

## Declarations

### Ethics approval and consent to participation

The study protocol was approved by the Institutional Review Board of the Korea University Anam Hospital. Informed consent was obtained from all the patients. This study was conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines.

### Consent for publication

Not applicable.

### Competing interests

The authors have no conflicts of interest to declare.

### Author details

<sup>1</sup>Department of Computer Science and Engineering, Korea University, Seoul, Korea

<sup>2</sup>OncoMASTER Inc., Seoul, Korea

<sup>3</sup>Institute of Human Behavior & Genetic, Korea University College of Medicine, Seoul, Korea

<sup>4</sup>Biomedical Research Center, Korea University Anam Hospital, Seoul, Korea

<sup>5</sup>Division of Medical Oncology, Department of Internal Medicine, Korea University College of Medicine, Korea University Anam Hospital, 73, Goryeodae-ro, Seongbuk-gu, Seoul 02841, Republic of Korea

Received: 19 January 2024 / Accepted: 16 April 2024

Published online: 20 April 2024

## References

- Kim TH, Kim IH, Kang SJ, Choi M, Kim BH, Eom BW, Kim BJ, Min BH, Choi CI, Shin CM, et al. Korean practice guidelines for gastric Cancer 2022: an Evidence-based, Multidisciplinary Approach. *J Gastric Cancer*. 2023;23(1):3–106.
- Fuchs CS, Tomasek J, Yong CJ, Dumitru F, Passalacqua R, Goswami C, Safran H, Dos Santos LV, Aprile G, Ferry DR, et al. Ramucirumab monotherapy for previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (REGARD): an international, randomised, multicentre, placebo-controlled, phase 3 trial. *Lancet*. 2014;383(9911):31–9.
- Wilke H, Muro K, Van Cutsem E, Oh SC, Bodoky G, Shimada Y, Hironaka S, Sugimoto N, Lipatov O, Kim TY, et al. Ramucirumab plus Paclitaxel versus placebo plus paclitaxel in patients with previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (RAINBOW): a double-blind, randomised phase 3 trial. *Lancet Oncol*. 2014;15(11):1224–35.
- Thuss-Patience PC, Kretschmar A, Bichev D, Deist T, Hinke A, Breithaupt K, Dogan Y, Gebauer B, Schumacher G, Reichardt P. Survival advantage for irinotecan versus best supportive care as second-line chemotherapy in gastric cancer—a randomised phase III study of the Arbeitsgemeinschaft Internistische Onkologie (AIO). *Eur J Cancer*. 2011;47(15):2306–14.
- Kang JH, Lee SJ, Lim DH, Park KW, Oh SY, Kwon HC, Hwang IG, Lee SC, Nam E, Shin DB, et al. Salvage chemotherapy for pretreated gastric cancer: a randomized phase III trial comparing chemotherapy plus best supportive care with best supportive care alone. *J Clin Oncol*. 2012;30(13):1513–8.
- Kang YK, Boku N, Satoh T, Ryu MH, Chao Y, Kato K, Chung HC, Chen JS, Muro K, Kang WK, et al. Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2017;390(10111):2461–71.
- Kang YK, Chen LT, Ryu MH, Oh DY, Oh SC, Chung HC, Lee KW, Omori T, Shitara K, Sakuramoto S, et al. Nivolumab plus chemotherapy versus placebo plus chemotherapy in patients with HER2-negative, untreated, unresectable advanced or recurrent gastric or gastro-oesophageal junction cancer (ATTRACTION-4): a randomised, multicentre, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol*. 2022;23(2):234–47.
- Janjigian YY, Shitara K, Moehler M, Garrido M, Salman P, Shen L, Wyrwicz L, Yamaguchi K, Skocyzylas T, Campos Bragagnoli A, et al. First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. *Lancet*. 2021;398(10294):27–40.



9. Janjigian YY, Kawazoe A, Yanez P, Li N, Lonardi S, Kolesnik O, Barajas O, Bai Y, Shen L, Tang Y, et al. The KEYNOTE-811 trial of dual PD-1 and HER2 blockade in HER2-positive gastric cancer. *Nature*. 2021;600(7890):727–30.
10. Rha SY, Oh DY, Yanez P, Bai Y, Ryu MH, Lee J, Rivera F, Alves GV, Garrido M, Shiu KK, et al. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for HER2-negative advanced gastric cancer (KEYNOTE-859): a multicentre, randomised, double-blind, phase 3 trial. *Lancet Oncol*. 2023;24(11):1181–95.
11. Ji SH, Lim DH, Yi SY, Kim HS, Jun HJ, Kim KH, Chang MH, Park MJ, Uhm JE, Lee J, et al. A retrospective analysis of second-line chemotherapy in patients with advanced gastric cancer. *BMC Cancer*. 2009;9:110.
12. Kanagavel D, Pokataev IA, Fedyanin MY, Tryakin AA, Bazin IS, Narimanov MN, Yakovleva ES, Garin AM, Tjulandin SA. A prognostic model in patients treated for metastatic gastric cancer with second-line chemotherapy. *Ann Oncol*. 2010;21(9):1779–85.
13. Catalano V, Graziano F, Santini D, D'Emidio S, Baldelli AM, Rossi D, Vincenzi B, Giordani P, Alessandroni P, Testa E, et al. Second-line chemotherapy for patients with advanced gastric cancer: who may benefit? *Br J Cancer*. 2008;99(9):1402–7.
14. Hasegawa H, Fujitani K, Nakazuru S, Hirao M, Mita E, Tsujinaka T. Optimal indications for second-line chemotherapy in advanced gastric cancer. *Anticancer Drugs*. 2012;23(4):465–70.
15. Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, Liu J, Yue YG, Wang J, Yu K, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med*. 2015;21(5):449–56.
16. Weaver BA. How Taxol/paclitaxel kills cancer cells. *Mol Biol Cell*. 2014;25(18):2677–81.
17. Swanton C, Marani M, Pardo O, Warne PH, Kelly G, Sahai E, Elustondo F, Chang J, Temple J, Ahmed AA, et al. Regulators of mitotic arrest and ceramide metabolism are determinants of sensitivity to paclitaxel and other chemotherapeutic drugs. *Cancer Cell*. 2007;11(6):498–512.
18. Sundar R, Barr Kumarakulasinghe N, Huak Chan Y, Yoshida K, Yoshikawa T, Miyagi Y, Rino Y, Masuda M, Guan J, Sakamoto J, et al. Machine-learning model derived gene signature predictive of paclitaxel survival benefit in gastric cancer: results from the randomised phase III SAMIT trial. *Gut*. 2022;71(4):676–85.
19. Li X, Zhai Z, Ding W, Chen L, Zhao Y, Xiong W, Zhang Y, Lin D, Chen Z, Wang W, et al. An artificial intelligence model to predict survival and chemotherapy benefits for gastric cancer patients after gastrectomy development and validation in international multicenter cohorts. *Int J Surg*. 2022;105:106889.
20. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. *arXiv Preprint arXiv:13013781* 2013.
21. Asgari E, Mofrad MR. Continuous distributed representation of biological sequences for Deep Proteomics and Genomics. *PLoS ONE*. 2015;10(11):e0141287.
22. Kim S, Lee H, Kim K, Kang J. Mut2Vec: distributed representation of cancerous mutations. *BMC Med Genomics*. 2018;11(Suppl 2):33.
23. Park KH, Choi JY, Lim AR, Kim JW, Choi YJ, Lee S, Sung JS, Chung HJ, Jang B, Yoon D, et al. Genomic Landscape and Clinical Utility in Korean Advanced Pan-cancer patients from prospective clinical sequencing: K-MASTER program. *Cancer Discov*. 2022;12(4):938–48.
24. Lee Y, Lee S, Sung JS, Chung HJ, Lim AR, Kim JW, Choi YJ, Park KH, Kim YH. Clinical application of targeted deep sequencing in metastatic colorectal Cancer patients: actionable genomic alteration in K-MASTER Project. *Cancer Res Treat*. 2021;53(1):123–30.
25. Choi YJ, Choi JY, Kim JW, Lim AR, Lee Y, Chang WJ, Lee S, Sung JS, Chung HJ, Lee JW, et al. Comparison of the data of a next-generation sequencing panel from K-MASTER Project with that of orthogonal methods for detecting Targetable genetic alterations. *Cancer Res Treat*. 2022;54(1):30–9.
26. Adlung L, Cohen Y, Mor U, Elinav E. Machine learning in clinical decision making. *Med*. 2021;2(6):642–65.
27. Fuchs CS, Muro K, Tomasek J, Van Cutsem E, Cho JY, Oh SC, Safran H, Bodoky G, Chau I, Shimada Y, et al. Prognostic Factor Analysis of Overall Survival in gastric Cancer from two phase III studies of second-line Ramucirumab (REGARD and RAINBOW) using pooled Patient Data. *J Gastric Cancer*. 2017;17(2):132–44.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.