Volume 23, No 4 | International Journal of Radiation Research, October 2025

Enhancing radiotherapy for gastric cancer: A systematic review on the role of predictive models in clinical decision-making

Sh. Chen[1], Sh. Wei[1], Y. Wang[1], L. Chen[1], Zh. Lin[2*]

[1]Department of Gastric Surgery, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fujian Province, China
[2]Department of Urological Oncology Surgery, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fujian Province, China

► ***Review article***

Corresponding author:
Zhitao Lin, M.D.,
E-mail: *hxsf803@163.com*

Received: April 2025
Final revised: June 2025
Accepted: June 2025

Int. J. Radiat. Res., October 2025; 23(4): 1101-1114

DOI: 10.61186/ijrr.23.4.38

Keywords: *Stomach neoplasms, radiotherapy, predictive value of tests, machine learning, clinical decision-making.*

ABSTRACT

Predictive models have become essential tools in modern oncology, significantly advancing gastric cancer management—one of the most prevalent and lethal cancers worldwide. This systematic review investigates the application of predictive models derived from public databases, emphasizing their role in improving radiotherapy outcomes. Various modeling techniques are explored, including statistical methods like logistic regression and Cox proportional hazards models, machine learning approaches such as random forests and support vector machines, and deep learning models like convolutional and recurrent neural networks. These models contribute to early detection, prognosis estimation, treatment response prediction, and tumor classification. Notably, in the context of radiotherapy, predictive models enhance tumor delineation, assist in selecting optimal radiation doses, and forecast individual treatment responses, reducing toxicity and improving precision. Other key clinical applications include molecular subtyping, biomarker discovery, and image-based diagnostics, especially through endoscopic and histopathological image analysis. These applications support the development of personalized treatment regimens and improve long-term patient outcomes. Despite their promise, several challenges remain, including inconsistent or imbalanced data, limited interpretability of complex algorithms, and concerns regarding clinical trust and AI transparency. Addressing these issues requires the development of high-quality standardized databases, stronger data-sharing frameworks, the adoption of federated learning methods, and the integration of explainable AI models into clinical workflows. This review concludes that predictive models, when properly validated and implemented, hold substantial potential to transform gastric cancer radiotherapy by enabling more tailored, data-driven treatment strategies, ultimately improving survival rates and quality of life for patients.

INTRODUCTION

Gastric cancer remains one of the most prevalent and lethal malignancies globally, ranking as the fifth most commonly diagnosed cancer and the fourth leading cause of cancer-related mortality [1]. In 2022, over 968,000 new cases and 660,000 deaths were reported, underscoring the urgent need for improved management strategies [2]. Its global burden is especially concerning due to late-stage diagnoses, tumor heterogeneity, and limited treatment options, particularly in younger populations where incidence rates are rising [3].

Radiotherapy is a key component of treatment for advanced or inoperable gastric cancer. However, its effectiveness is constrained by the difficulty in precisely targeting tumors while minimizing damage to adjacent healthy tissues [4]. Advances in artificial intelligence (AI) and data-driven predictive modeling are opening new possibilities in oncology, particularly in enhancing the safety and efficacy of radiotherapy [5].

Numerous studies have demonstrated the promise of predictive models-ranging from traditional statistical methods (e.g., Cox regression) to advanced machine learning (e.g., random forests, support vector machines) and deep learning algorithms (e.g., convolutional neural networks)-in improving early diagnosis, risk stratification, and survival prediction in gastric cancer [6, 7]. These models can analyze high-dimensional clinical, genomic, and imaging data from public databases such as TCGA, GEO, and SEER, offering potential in guiding personalized treatment decisions. Radiomics-based models, in particular, have demonstrated superior performance in tumor delineation,

[DOI: 10.61186/ijrr.23.4.38]

treatment response prediction, and recurrence surveillance [8].

Despite growing interest, most previous reviews have focused either broadly on predictive modeling in oncology or narrowly on single cancer types or techniques, without integrating insights specific to gastric cancer radiotherapy [9] Moreover, limited attention has been given to the clinical translation of these predictive tools, their validation using real-world datasets, or their implementation challenges in radiotherapy workflows.

This study provides a comprehensive and focused systematic review of predictive models specifically applied to radiotherapy in gastric cancer. The novelty lies in synthesizing diverse modeling approaches and their clinical applications, highlighting how predictive tools can optimize radiotherapy planning, enhance treatment personalization, and ultimately improve patient outcomes. By addressing gaps in the literature and discussing implementation barriers, this review offers actionable insights for future research and clinical integration.

Public databases for gastric cancer research

The increasing concern about the gastric cancer have led to development of public databases that are focused on providing research data for gastrointestinal tumors. The databases act as an important resource for the for developing modes for detection, prognosis estimation as well as the treatment processes [10].

Among the most common database is The Cancer Genome Atlas (TCGA). TCGA focuses on provision of extensive genomic data, such as the gene expression, mutation, and methylation profiles. These metrics are critical in the gastric cancer analysis process [11]. Another important database is the Gene Expression Omnibus (GEO), which provides data used for the differential gene analysis in gastric cancer research. It also contains the microarray and RNA sequencing datasets, which are applicable in the biomarker discovery and transcriptomic analysis [12]. Additionally, the Gastric Cancer Proteomics Database (GCPDB) contains the mass spectrometry-based proteomics data, which helps in the development of the clinical utility and insights of the protein biomarkers [13]. There is also the Surveillance, Epidemiology, and End Results (SEER) program, which functions as a cancer registry. It provides data on cancer incidences, survival rates as well as treatment patterns. The UK Biobank provides cancer related data on aspects such as genetics, and lifestyles supporting large-scale epidemiological research [14].

From these databases, there are various types of data available that are applicable in gastric cancer research. These include clinical data which provides patient's characteristics such as treatment outcomes and gastrointestinal tumor characteristic. Data on treatment type and effectiveness, as suggested by Wang (2024) presents the real-world case of gastric cancer situation. Others include the genomic datasets which provides gene mutations and number variations. The proteomic data gives statistics on cancer progression and treatment effects. As suggested by Almeda *et al*. (2022), the imaging data such as the endoscopic and histopathological images helps in development of artificial intelligence diagnostic models [15].

From literature review, these databases facilitate robust model training and validation, due to its large size availability. The diverse modalities of such databases are critical in facilitating the multi-omics and multi-modal analysis and precision medicine [16]. However, such data may be affected by data quality issues, such as missing or inconsistent data [17]. Another critical consideration suggested by McGrail, *et al.* (2024) is the ethical and privacy concerns, since patients' health information requires strict adherence to the confidentiality and privacy requirements and policies.

Research methodology

The study adopted a qualitative research design which was based on systematic review of literature and cases studies. Due to the evolving nature of predictive models, the study focused on comprehensively synthesizing available knowledge and clinical applications in gastric cancer research. The study adopted the PRISMA guidelines (figure 1) to ensure that the adopted review process is structured and replicable.

The systematic review followed a structured methodology to identify relevant studies on predictive models applied to gastric cancer, specifically focusing on radiotherapy techniques. The search strategy aimed to gather peer-reviewed articles from multiple academic databases, ensuring that the most recent and relevant literature was included.

Search strategy and data sources

The literature search for this study was carried out using several major academic databases to ensure comprehensive and high-quality coverage of relevant research. PubMed was utilized to identify studies with a strong focus on the biological and clinical dimensions of gastric cancer, while Scopus provided access to a broad range of multidisciplinary research articles. Additionally, Web of Science was included to capture studies published in high-impact journals, ensuring that the reviewed literature met rigorous academic standards.

To retrieve the most relevant studies, a systematic search strategy was employed, combining specific keywords with Boolean operators to optimize precision and inclusivity. The search terms were carefully selected to reflect the scope of the review and included combinations such as "predictive

[DOI: 10.61186/ijrr.23.4.38]

models" and "gastric cancer", "predictive models" and "public databases", as well as terms like "gastric cancer" or "stomach cancer", "predictive modeling" or "predictive analytics", and "clinical application" in conjunction with "predictive models". This approach ensured that the search captured a wide array of studies addressing the use of predictive modeling in the clinical context of gastric cancer.

Time interval for data search

The search was limited to studies published within a defined time interval of 2020 to 2025. This five-year period was chosen to ensure that only the most recent research on predictive models and radiotherapy techniques for gastric cancer was reviewed. Given the rapid advancement in machine learning and AI-driven predictive models, this timeframe helped to capture the most up-to-date methodologies and clinical applications.

Criteria for selection of manuscripts

The selection of manuscripts for this review was guided by well-defined inclusion criteria aimed at ensuring the relevance, quality, and scientific rigor of the studies analyzed. Only original research articles were considered appropriate for inclusion. These encompassed clinical trials, cohort studies, and other observational designs that provided primary data relevant to predictive modeling in the context of gastric cancer, specifically in relation to radiotherapy. Secondary literature such as systematic reviews, meta-analyses, opinion pieces, and case reports was excluded, as such publications do not present original empirical findings or model development.

To maintain alignment with current advancements in technology and clinical practice, the review was limited to studies published between the years 2020 and 2025. This temporal scope was chosen to reflect the latest innovations in computational modeling and radiotherapy strategies, which are evolving rapidly within oncological research. Earlier studies were not considered, as they may rely on outdated methodologies or lack integration with contemporary data infrastructures.

Given the linguistic capacity of the research team and the predominance of English in the indexed academic literature, only studies published in English were included. This ensured consistency in data extraction and interpretation, and minimized the risk of miscommunication due to translation inaccuracies.

Priority was given to studies that made use of large-scale, publicly available databases such as The Cancer Genome Atlas (TCGA), the Gene Expression Omnibus (GEO), the Surveillance, Epidemiology, and End Results (SEER) program, and the UK Biobank. These repositories are recognized for their comprehensive clinical, genomic, and imaging datasets, which are essential for building and validating robust predictive models. Studies relying on proprietary or limited-access datasets were generally excluded, especially in cases where methodological transparency could not be adequately verified.

The population of interest was strictly limited to individuals diagnosed with gastric cancer, with an emphasis on those undergoing radiotherapy. Studies involving other cancer types or non-human subjects were excluded to maintain a clear and consistent focus on the target patient population.

The methodological foundation of each study was also a critical factor in the selection process. Only those investigations that implemented predictive modeling techniques-ranging from traditional statistical models to more advanced machine learning and deep learning approaches-were considered suitable for inclusion. These models had to be applied in ways that directly related to clinical endpoints such as treatment response, survival prediction, radiotherapy effectiveness, or tumor progression.

Furthermore, selected studies were required to demonstrate a high level of methodological rigor. This included clearly described processes for model development and validation, sufficient sample sizes, and reproducibility of results. Studies that lacked methodological transparency or exhibited potential biases were excluded to uphold the quality and reliability of the review.

Exclusion criteria

This study adopted a systematic approach to review and synthesize original research focused on predictive modeling in gastric cancer, specifically in relation to radiotherapy outcomes and treatment predictions. The review was confined to peer-reviewed articles that presented primary empirical data, including both clinical trials and observational studies. Only original research was included to ensure that all findings were derived from primary data sources with clearly defined methodologies. Consequently, secondary sources such as reviews, opinion articles, conference abstracts, case reports, and meta-analyses were excluded due to their lack of original empirical data and predictive modeling content.

To maintain relevance with current advancements in computational modeling and clinical oncology, the review was limited to studies published between 2020 and 2025. This period was selected to capture recent innovations in predictive analytics, including artificial intelligence and machine learning applications in the field of gastric cancer. Older publications, particularly those predating 2020, were excluded as they often reflected outdated methodologies and lacked integration with contemporary datasets and tools.

The language of publication served as another inclusion criterion. Only studies published in English were included in the review to ensure clarity,

[DOI: 10.61186/ijrr.23.4.38]

accessibility, and consistency of interpretation by the reviewers. Articles published in other languages were excluded due to limitations in language proficiency and the potential risk of misinterpretation or oversight of methodological details.

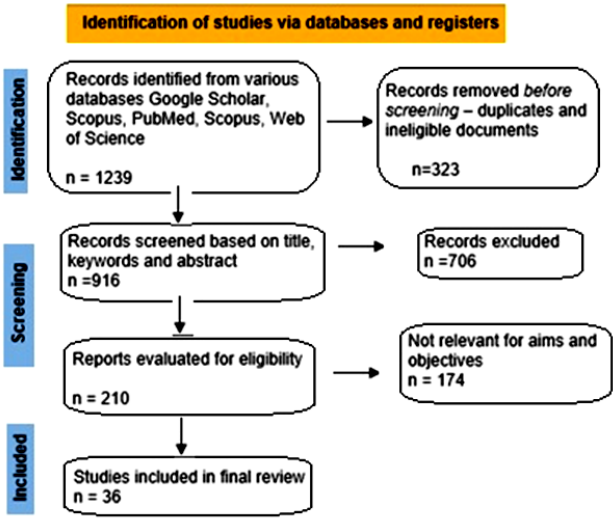

Figure 1. PRISMA studies identification process. This figure illustrates the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart, depicting the process of identifying studies for inclusion in the systematic review. The chart shows the stages of screening, eligibility assessment, and final inclusion of studies, outlining the number of articles excluded at each step and providing transparency on the review process.

In terms of data sources, the review prioritized studies that utilized publicly available, reputable databases such as The Cancer Genome Atlas (TCGA), the Gene Expression Omnibus (GEO), the Surveillance, Epidemiology, and End Results (SEER) program, and the UK Biobank. These databases are known for their comprehensive clinical, genomic, and imaging datasets, which are particularly valuable for developing and validating predictive models in oncology. Studies that relied on proprietary, inaccessible, or insufficiently described datasets were excluded to preserve transparency and reproducibility, both critical for scientific validation.

The population of interest in the reviewed studies consisted of patients diagnosed with gastric cancer, especially those with advanced or inoperable disease undergoing radiotherapy. To ensure clinical relevance, studies were only included if they investigated predictive models applicable to real-world decision-making in gastric cancer treatment. Articles involving non-human subjects or addressing cancers other than gastric were excluded.

A central focus of this review was the type of predictive model employed. Included studies featured models designed to forecast clinically meaningful outcomes such as overall survival, tumor recurrence, radiotherapy response, or disease progression. These models encompassed a range of methodologies, including traditional statistical approaches, machine learning algorithms, and deep learning frameworks. In contrast, studies limited to descriptive statistics or basic univariate analyses without predictive capabilities were not considered for inclusion.

The review also emphasized the necessity for studies to report clinically significant outcomes. Accepted endpoints included radiotherapy response, survival prediction, treatment efficacy, and tumor characterization, provided these were analyzed through validated predictive models. Studies that were purely theoretical in nature or did not report relevant clinical outcomes were excluded, as they did not contribute directly to practical improvements in gastric cancer care.

The methodological rigor of each study was evaluated using established appraisal tools such as the Cochrane Risk of Bias tool for randomized controlled trials and the Newcastle-Ottawa Scale for observational studies. Studies were only included if they demonstrated adequate methodological quality, including transparency in model development, validation processes, sufficient sample sizes, and reproducibility of results. Those with poorly defined methods or obvious biases were excluded.

Finally, only studies that reported quantitative performance metrics of the predictive models were included. These metrics typically encompassed measures such as the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and accuracy, allowing for a reliable assessment of model robustness. Studies lacking these measurable outcomes were excluded to maintain a consistent and evaluative focus on model performance.

Data extraction and limitations

A structure approach was followed in the data extraction and synthesis process. The approach focused on ensuring consistency and reproducibility. The extraction followed a standardized approach. The aspects considered were study details such as authors, year, title and sources; research focus of the predictive models; data sources of public databases; and clinical applications such as risk predictions, survival analysis or treatment response predictions. A quality assessment for the study design, sample size, reproducibility and transparency were used. However, some limitations are highlighted. These include that some studies may not have been indexed in the selected databases. Additionally, only English language articles were considered, which may have excluded some relevant studies.

Predictive modeling in gastric cancer

Due to their ability to enhance early detection of the gastrointestinal tumor and improve the treatment efforts, the predictive models for gastric cancer are

[DOI: 10.61186/ijrr.23.4.38]

gaining significant attention. As such, various models have been developed for the assessment of the risk factors, diagnostics, prediction of survival outcomes, and treatment procedures. These models could be categorized and evaluated based on their purposes as discussed in this section.

From the review of literature, these predictive models could be categorized in to three categories; first is the statistical models, second is the machine learning models, and finally the deep learning models.

Statistical predictive models

Statistical predictive models have been widely adopted in the gastric cancer research. These models include the logistic regression models, Cox proportional hazard models, as well as the Kaplan-Meier Survival Analysis [18, 19]. Logistic regression models are generally used in classification gastric cancer research problems. For instance, predicting the presence of absence of gastric cancer, or evaluating the impact of a gastric cancer risk factor such as smoking, alcohol consumption or age, on chances of the disease occurring.

Similarly, the Cox proportional; hazard model is used in survival analysis, such as estimation until the time of occurrence of an event, such as death. The model is also used in identification of survival factors [20]. For instance, the research by Sabbagh, *et al.*, (2023) used the model and highlighted that neoadjuvant chemoradiotherapy demonstrated worse survival outcomes compared to perioperative chemotherapy, with a hazard ratio (HR) of 1.57. as well, the model could be adopted in analyzing the effect of various covariates, such as treatment type, tumor stage and genetic mutations on survival rates. In such research, the following model could be specified;

$$h(t) = h_0(t)e^{(\beta_1 X_1 + \beta_2 + \ldots + \beta_n X_n)} \quad (1)$$

In this model, $h(t)$ implies the hazard rate, the $h_0(t)$ implies the baseline hazard function, while the $X_1 \ldots X_n$ implies the covariate factors such as tumor size. The summary of various authors application of these models in research is summarized in table 1.

Table 1. Statistical models in gastric cancer research

Author(s)	Objectives	Model(s) Used	Data Used	Key Findings
Karamoozian *et al.* (2021) (18)	Evaluating the factors that influence short-term and long-term survival of patients with gastric cancer	Bayesian mixture cure rate frailty model	Risk Factor Data - recorded in hospitals of Kerman province	Chemotherapy, morphology, metastasis was identified as effective factors in short-term and long-term survival of patients.
Allen *et al.* (2021) (20)	Comparison of neoadjuvant chemotherapy to neoadjuvant chemotherapy plus chemoradiation for patients with gastric adenocarcinoma	A Multi-institutional Analysis of Cox Proportional Hazards	data from 2 high-volume cancer centers	CRT recorded higher rates of completed perioperative therapy, pathologic response; lower pathologic stage, and improved survival
Sabbagh *et al.* (2023) (21)	Evaluated survival outcomes based on treatment type - neoadjuvant chemoradiotherapy versus perioperative chemotherapy	Kaplan Meier (KM) and Cox proportional hazards analysis	The National Cancer Database Treatment Data	Found worse survival outcomes with neoadjuvant chemoradiotherapy (HR = 1.57) as compared to PC
Darang *et al.* (2023) (22)	to identify genes, biomarkers, and metabolic pathways affecting gastric cancer	Gene Expression Profiling and Interactive Analyses (GEPIA) and the Kaplan-Meier method	Patient Data - gene expression profiles of tumor lesions and adjacent non-tumor mucosa sample	important pathway was enriched in ECM-receptor interaction;
Shamsi *et al.* (2024) (23)	Evaluating 5-year survival rates for gastric cancer	Kaplan-Meier analysis and log-rank test.	Patient Data - patients with GAC referred to Afzalipour, Bahonar, and Shafa Hospitals	Found that total gastrectomy improved survival compared to subtotal gastrectomy

This table summarizes various statistical predictive models applied in gastric cancer research. The table includes the authors, objectives, models used, the data utilized, and key findings. Models such as Cox proportional hazards, Kaplan-Meier survival analysis, and Bayesian mixture models are highlighted in their use for survival analysis and risk factor evaluation. The data sources span clinical databases and cancer registries, contributing to survival prediction and risk assessment.

The Kaplan-Meier Survival analysis model is a non-parametric technique, majorly applicable predicting survival probabilities over time. For instance, the comparison of different survival curves for different patients, such as effectiveness of chemotherapy versus surgery [22]. A study by Shamsi, *et al.,* (2024) adopted the model to evaluate the 5-year survival rates for gastric cancer. The study results showed that total gastrectomy significantly improved survival compared to subtotal gastrectomy [23].

Machine learning predictive models

Machine learning models are enhanced predictive techniques, which use clinical and genomic data for gastric cancer detection, prognosis and treatment research. They are considered complex models with capabilities to capture non-linear relationship [10].

Example of these is the random forest predictive model, which is widely used in researching gastric cancer re-occurrence and treatment responses. It utilizes varying decision trees, which could be trained on different datasets. By aggregating results from all the trees, this model is able to provide more accurate prediction. Park (2025) [24] used this model to evaluate mortality after curative gastrectomy for gastric cancer. Another research by Zhou *et al.*,

[DOI: 10.61186/ijrr.23.4.38]

(2025) adopted random forest to predict presentation delays in gastric cancer patients, achieving an AUC of 0.893-0.925.

Another applicable model is the Extreme Gradient Boosting (XGBoost) model, which is considered effective in high structural applications. For instance, the prediction of chemotherapy responses and survival rates. Xu & Guo (2024) [25] harnessed the power of improved XGBoost model to research on the class imbalance in gastric cancer survival predictions. The results showed enhanced performance and interpretability. Another study adopted the model for validation of three gastric cancer subtypes. The cancer categories were identified through similarity network fusion and consensus clustering, which, enhancing clinical prediction performance. The summary of predictive models' application in research by various researchers is summarized in table 2.

Table 2. Machine learning predictive models' applications in gastric cancer research.

Author(s)	Objectives	Model(s) Used	Data Used	Key Findings
Park (2025) [24]	Mortality prediction after curative gastrectomy for gastric cancer	CatBoost, Gradient Boosting, Light GBM, Random Forest, XGBoost	Post-Gastrectomy Data - Korean Gastric Cancer Association	Provided accurate risk stratification - CatBoost model demonstrated robust and consistent performance in predicting PM risk
Zhou *et al.* (2025) [26]	Predicting presentation delays in gastric cancer patients	support vector machine (SVM), random forest (RF), gradient boosted trees (GBDT), extremely gradient boosting (XGBoost)	Patient Delay Data - gastric cancer patients admitted to a tertiary oncology hospital	Achieved an AUC of 0.893-0.925 - RF based model has favorable performance for the prediction of presentation delay in gastric cancer patients
Xu & Guo (2024) [25]	Addressing class imbalance in survival predictions	XGBoost model	Survival Data	Improved performance and interpretability
Li *et al.* (2022) (8)	Validation of three gastric cancer subtypes	XGBoost, Cox regression model	Genomic Data - mRNA, microRNA, and DNA methylation datamRNA, microRNA, and DNA methylation data	subtype 1 - favorable prognosis and high ARID1A and PIK3CA mutations; subtype 2 - poor prognosis and harbored high recurrent TP53 mutations; subtype 3 - high CHD1, APOA1

This table presents a summary of machine learning models used in gastric cancer research, highlighting the specific models (e.g., Random Forest, XGBoost, SVM) and their application areas such as survival prediction, chemotherapy response, and tumor classification. The table also includes key findings from the respective studies, providing insight into the models' performance (e.g., AUC values) and the datasets used.

This model is considered powerful and significantly applicable due to its ability to optimize loss function by using gradient boosting. Considering the mathematical formulation, the XGBoost model could be applied in the prediction of chemotherapy response in gastric cancer treatment. The model equation would use boosting framework, where each trees uses the previous trees errors (equation 2):

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta \cdot f_t(Xi) \quad (2)$$

Where; $\hat{y}_i^{(t)}$ implies the intended predictions, η is the learning rate of the model, while $f_t(Xi)$ refers to the model weak learning. This model would be minimized by application of the gradient descent, in equation 3.

$$L = i = \sum_{i=1}^{N} l(yi,\hat{y}i) + \sum_{i=1}^{N} \Omega(ft) \quad (3)$$

Where; l(yi,ŷi) implies the loss function of the model, while Ω(ft) implies the regularization term which controls the complexity of the tree. This model could use datasets such as the genomic and clinical data (TCGA) and features such as gene expression and treatment type, to predict the probability of a successful chemotherapy response in gastric cancer patients' treatment.

Deep learning models

These are the most advanced models that could be adopted in gastric cancer research, utilizing public databases. They have a powerful capacity to process information resembling that of a human being. Through enhanced and interconnected layers of neurons, the data is transformed and analyzed using series of activation functions and weighted connections [27]. In gastric cancer research, these models are applicable in early case detection using imaging technologies, for instance the endoscopic or histopathological imaging. They are also useful in survival prediction and treatment response modeling.

A nother deep learning model considered is the Convolutional neural networks (CNNs) which is applicable in gastric cancer research, particularly in imaging analysis (Maheswari et al., 2024). CNN processes images using conventional filters to detect patterns in them. In this application, they are highly suitable for endoscopic and histopathological image analysis.

Bhardwaj, *et al.*, (2024) [31] adopted the advanced CNN models conducting gastric cancer diagnosis. The analysis depicted high accuracy rate of up to 99.88% in analysis of endoscopic images. As well, Wu, *et al.*, (2025) [29] adopted the CNNs in the prediction of HER2 status gastric adenocarcinoma. The prediction achieved an accuracy AUC of 0.847 which was considered critical in the determination of treatment efficacy for the trastuzumab.

[DOI: 10.61186/ijrr.23.4.38]

Table 3. Deep learning models in gastric cancer research.

Author(s)	Objectives	Model used	Data Used	Key Findings
Maheswari *et al.* (2024) [28]	Image processing for gastric cancer diagnosis	CNN "Gastronet," deep learning system integrating three algorithms-Perform Multiple Tasks Net, Mix Net, and Overall Net	Endoscopic Images	Achieved 99.88% accuracy in endoscopic image analysis
Wu *et al.* (2025) [29]	Predicting HER2 trastuzumab treatment efficacy status in gastric adenocarcinoma	convolutional neural network (CNN) model	Histopathological Data - cohort of 300 consecutive surgical specimens and 101 biopsy specimens	Achieved an AUC of 0.847 for treatment efficacy prediction
Thota *et al.* (2025) [30]	Mortality prediction for upper gastrointestinal cancer	long short-term memory (LSTM) neural networks	mortality data from the Center for Disease Control and Prevention (CDC) Wide-Ranging Online Data for Epidemiological Research (WONDER) Database	mortality rates for upper GI tract cancers are notably higher among elderly individuals, particularly those aged 80 and above

This table outlines the deep learning models applied to gastric cancer research, focusing on Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. The table includes the objectives of each study, the data types (e.g., endoscopic images, histopathological data), and the key findings, such as the high accuracy of CNNs in image classification and LSTMs' application for mortality prediction. The table demonstrates the strength of deep learning in handling high-dimensional data like medical images.

Another model considered is the recurrent neural networks (RNNs) and its improved version referred to as the long short-term memory (LSTM). These models are applicable for cancer progression prediction. These models process sequential data such as the patient follow up records, through maintenance of memory across time steps. For instance, Thota, *et al.*, (2025) [30] adopted LSTM in the prediction of mortality as a result of the upper gastrointestinal tract cancer. The results showed quite a high level of forecasting reliability. Considering the mathematical representation of application of RNN, the hidden states are updated as follows;

$$h_t = \tanh(W_h h_{t-1} + W_x X_t + b) \quad (4)$$

Where; h_t implies the hidden state, W_h, W_x implies the weight matrices, the X_t implies the input factors such as the tumor size change. To counter the weakness of RNN vanishing gradient, the LSTM introduces the gates as shown in the following specifications.

$$f_t = \sigma(W_f \cdot [h_{t-1} X_t] + b_f)$$
$$i_t = \sigma(W_i \cdot [h_{t-1} X_t] + b_i)$$
$$C_t = f_t \times C_{t-1} + i_t \times \tanh(W_C \cdot [h_{t-1} X_t] + b_C)$$
$$o_t = \sigma(W_o \cdot [h_{t-1} X_t] + b_o)$$
$$h_t = o_t \times \tanh(C_t) \quad (5)$$

Where; f_t i_t and o_t implies the forget, input and the output gates, while the C_t implies memory cell which stores the long-term patient history. The RNN and $C_t = f_t \times C_{t-1} + i_t \times \tanh(W_C \cdot [h_{t-1} X_t] + b_C)$ LSTM model are applicable in predicting gastric cancer recurrency using longitudinal electronic health records. The applicable features include factors like the tumor size progression and biomarker levels, with the output of future cancer recurrence risk. Table 4 and figure 2 summarize the three categories of models, the specific models and the areas of application in gastric cancer research.

Table 2. Statistical data of the radiographic parameters (kVp and mAs values) and patient anthropometric data for selected X-ray examinations.

Model Type	Specific Models	Application in Gastric Cancer
Statistical Models	Logistic Regression	Risk prediction, survival analysis
	Cox Proportional Hazards Model	Survival analysis, hazard estimation
	Generalized Linear Models (GLM)	Association between risk factors and outcomes
	ARIMA/Time Series Models	Forecasting incidence rates
Machine Learning Models	Random Forest (RF)	Predicting recurrence, patient classification
	Support Vector Machine (SVM)	Classifying histopathological images
	XGBoost	Chemotherapy response prediction
	K-Means Clustering	Molecular subtyping of gastric cancer
Deep Learning Models	The Convolutional Neural Network (CNN) model	Tumor detection in endoscopic images
	The Recurrent Neural Network (RNN)	Predicting cancer progression
	The Long Short-Term Memory (LSTM) model	Time-series analysis of patient data
	Autoencoder	Data augmentation, feature extraction

This table provides a comprehensive summary of the different predictive models used in gastric cancer research. It categorizes models into three types-statistical models, machine learning models, and deep learning models-and lists their applications in gastric cancer, such as risk prediction, chemotherapy response prediction, and tumor detection. The table also serves as a quick reference to understand the specific use of each model in clinical and research settings.

Radiotherapy in gastric cancer: role of predictive models

Radiotherapy remains a cornerstone in the treatment of gastric cancer, especially in cases where surgery is not an option or tumors are inoperable. The integration of predictive models into radiotherapy is transforming the way treatment is planned and delivered. These models improve the precision of radiotherapy, allowing for more personalized treatment plans that enhance outcomes

[DOI: 10.61186/ijrr.23.4.38]

and minimize side effects.

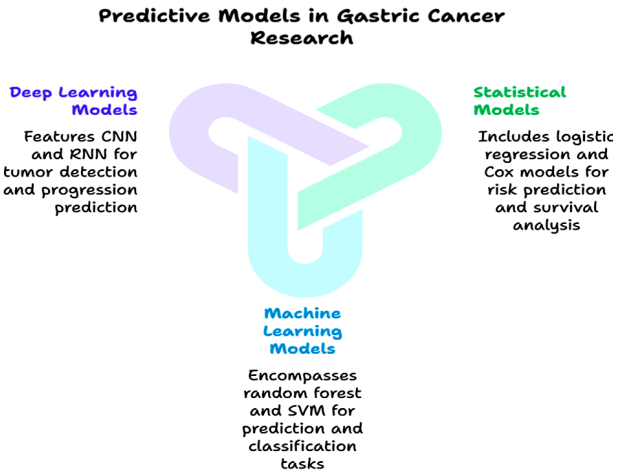

Figure 2. Summary of model application in gastric cancer research. This figure presents a visual summary of the applications of predictive models in gastric cancer research. It categorizes the models into statistical, machine learning, and deep learning types, and highlights their clinical applications such as tumor detection, survival analysis, and chemotherapy response prediction. Each model is represented in a distinct section to visually compare their respective roles in gastric cancer research and treatment.

Radiotherapy treatment planning and optimization

Predictive models play a crucial role in optimizing radiotherapy treatment plans. One of the primary challenges in gastric cancer radiotherapy is delivering the appropriate dose to the tumor while minimizing damage to surrounding healthy tissues. This is especially important in gastric cancer, where the tumor location and surrounding organs, such as the stomach, liver, and intestines, can complicate treatment.

Radiomics-based predictive models have been shown to improve tumor delineation by analyzing data from CT, MRI, and PET scans. For instance, convolutional neural networks (CNNs) are widely used to process medical imaging and assist in the segmentation of gastric tumors, making it easier to define the precise area that needs to be treated with radiation [32]. This enhanced tumor delineation is crucial for accurately planning the radiation dose and preventing excessive radiation exposure to adjacent organs.

Predicting treatment response to radiotherapy

Another critical application of predictive models in radiotherapy is the ability to predict individual patient responses to treatment. Since gastric cancer tumors can vary significantly in their sensitivity to radiation, predicting how a tumor will respond to treatment is key in personalizing care. Machine learning models, such as random forests and support vector machines, have been effectively used to predict treatment outcomes based on factors such as tumor size, stage, and molecular characteristics. By analyzing a combination of clinical, genomic, and imaging data, these models can predict which patients are most likely to benefit from radiotherapy, allowing clinicians to adjust the treatment accordingly. Models can also predict the likelihood of radiation resistance, enabling the adjustment of radiation doses or switching to alternative therapies when necessary [33].

Monitoring treatment progress and detecting treatment failure

Monitoring treatment progress is vital in radiotherapy, and predictive models are increasingly being used for this purpose. After a patient undergoes radiotherapy, follow-up imaging and clinical assessments are necessary to track tumor shrinkage and detect any signs of recurrence. Predictive models can help by analyzing follow-up imaging and clinical data to identify early signs of tumor regrowth or treatment failure, allowing clinicians to adjust treatment plans promptly.

For example, deep learning models, such as CNNs, can process post-treatment CT or MRI scans to detect subtle changes in the tumor that may not be visible to the human eye. These models can detect microstructural changes in the tumor and surrounding tissues, helping clinicians assess the effectiveness of radiotherapy early in the treatment process and reduce the risk of overlooking early treatment failures.

Reducing radiation toxicity and side effects

Reducing radiation toxicity is a significant challenge in the treatment of gastric cancer due to the proximity of the stomach and other critical organs. Predictive models can help mitigate this risk by optimizing radiation dosing. For example, models based on radiomics or AI-driven tools can predict how a tumor's radiation exposure will affect surrounding tissues, helping clinicians tailor the treatment to minimize healthy tissue damage.

Models that incorporate genomic data, such as immune signatures or genetic mutations, also play an essential role in predicting how a tumor will respond to radiation. This helps avoid over-radiating patients who might be more sensitive to radiation-induced side effects, such as esophagitis or bowel complications, and ensures that the treatment is as effective and safe as possible [34].

While predictive models hold significant promise for advancing radiotherapy in gastric cancer, there are several challenges that need to be addressed. First, the integration of diverse data sources-clinical, imaging, genomic, and radiological-into a single predictive model requires sophisticated algorithms and comprehensive data quality control. Furthermore, the interpretability of these models remains a key concern. Clinicians must be able to trust the predictions made by AI models, and understanding how a model arrived at its decision is

[DOI: 10.61186/ijrr.23.4.38]

crucial for its clinical adoption. Additionally, ensuring that predictive models are validated and generalizable across different patient populations is essential for their widespread implementation. Models need to be tested and refined using large-scale, diverse datasets to ensure that they are robust and reliable in various clinical settings.

Clinical application case discussions

There is various way of clinical application of predictive models from public databased in gastric cancer research. The areas of applications range from early detection, treatment and patient management. From the analysis of literature and case studies, this section discusses the various specific application cases for these models.

Risk prediction and early detection

Risk prediction is among the most critical and vital application of predictive models in gastric cancer research. The risk prediction models entail integration of demographic, clinical and genetic data from public databases, to predict the probability of developing gastric cancer. Statistical models such as the Cox proportional hazard models and logistic models have been applied in the in analyzing gastric cancer risk factor. For instance, the generic predisposition, diet and smoking risk factors. The study by Yu *et al*. (2025) [35] utilized logistic regression and random forest techniques, while predicting gastric cancer risk. Machine learning models such as the random forest and support vector machines utilize lifestyle information to capture the gastric cancer risk factor relationships. Park *et al.* (2024) [24] conducted a cohort analysis in gastric cancer research utilizing the SHapley Additive exPlanations (SHAP) model. The CNNs are the most common used deep learning models used in the endoscopic image analysis, which helps in detecting early stages of cancer [12], through gastroscopy imaging. Koushik, *et al*. (2024) adopted the deep belief network (DBM) model in efficient prediction of gastric cancer research, achieving an accuracy of 99.88%.

Gastric cancer treatment response prediction

Different patients have varying response to cancer treatments such as chemotherapy, immunotherapy, or targeted therapies. As such, predictive models utilizing predictive models have proved critical to predict and optimize treatment selection. The XGBoost and support vector machines learning models have proved relevant to evaluate and predict how patients respond to chemotherapy regiments (figure 2). These predictions utilize the histopathological and molecular market data. Sasagawa, *et al.* (2024) [33] demonstrated how the chemotherapy responsiveness in gastric cancer could be evaluated using the genomic and immune signatures. Chen, *et al.,* (2024) adopted the dynamic aware model and the longitudinal liquid biopsy data to predict treatment response of gastric cancer treatments.

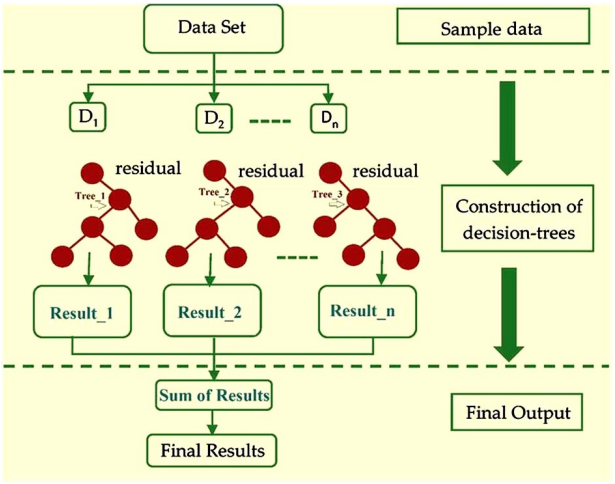

Figure 3. XGBoost Model Architecture. Source: Ahmed, *et al.*, (2023) with permission. This figure displays the architecture of the XGBoost model, showing how gradient boosting is applied to predict chemotherapy responses in gastric cancer. The diagram illustrates the decision tree structure and the boosting process where each tree corrects the errors made by the previous one, optimizing prediction accuracy. The figure emphasizes the model's iterative improvement through error correction, resulting in enhanced clinical prediction performance.

A Multi-Modal data integration model is a deep learning model utilized by Chen *et al*. (2024) in predicting the response to the anti-HER2 therapy and the anti-HER2 combined immunotherapy. Another treatment prediction was done by Eweje, *et al*., (2024) [36] whose study utilized the artificial intelligence based digital pathology model, for prediction of response and outcome of advanced gastro-esophageal cancer, using the immune checkpoint inhibitor therapy [36]. It is therefore evident from these research that these predictive models are critical in treatment predictions.

Molecular subtyping and biomarker discovery

There has been advancement in the molecular subtyping and biomarker discovery in gastric cancer by the application of advanced models. Among the applications include molecular subtyping, such as the multi-omics integration and the Disulfidptosis-Related lncRNAs [37]. Advanced models such as the K-means clustering on RNA sequencing data can utilized in the identification of novel gastric cancer subtypes that correlate with treatment response patterns. For instance, Bai et al. (2024) [37] utilized the multi-omics data including the transcriptomic, epigenetic, and somatic mutations, in the process of classifying gastric cancer into various molecular subtypes. Additionally, the biomarker discovery is another application of the machine learning models, in the prediction of chemotherapy responses [33]. The prediction took advantage of the genomic and immune signatures. It is also observed that the

[DOI: 10.61186/ijrr.23.4.38]

advanced deep learning models such as the auto-encoders help extraction of the latest features from the high dimensional genomic data, and reveal important and hidden patterns.

Imaging-based diagnosis and tumor classification

Another clinical application of the predictive models in gastric cancer research is the imaging-based diagnosis and tumor classification. Machine learning models have proved relevant in histopathology image analysis and feature fusion strategies [38]. Zubair *et al*., (2024) conducted an enhanced gastric cancer classification research, using the gastric histology classification and segmentation (GHCS) and achieved an accuracy of 98.87%. while, using the expectation-maximizing Naïve Bayes classifier, they achieved an approximate accuracy of 97.28% [38]. On their research, Loddo *et al*., (2024) demonstrated the application of the feature fusion in improving diagnostic precision. Their comparative analysis utilized both the handcrafted and deep features achieved, and resulted to an accuracy of 95%.

Deep learning models are quite relevant due to their improved accuracy predictions. Applicable deep learning models in gastric cancer research include CNN-based architectures like ResNet, VGG16, and EfficientNet models. Their application includes automating feature extraction and early-stage gastric cancer in endoscopic images. For instance, Bhardwaj *et al*., (2024) adopted the advanced CNN model for the detection of detecting gastric cancer from endoscopic images. Their study demonstrated the effectiveness of deep transfer learning in gastric cancer research. Additionally, Xu *et al*., (2024) adopted the GastroNet FusionAI algorithm to model and diagnose gastric cancer tumors. It is evident that these advanced models present promising future in clinical gastric cancer research.

Recurrence prediction

Recurrence prediction of gastric cancer is another critical application of the predictive models derived from public databases in gastric cancer research. Among the data leveraged on to predict recurrence include imaging, clinical and genomic information. The most common deep learning models include Multi-layer Perceptron (MLP) Models, an artificial neural network model with high capacity of learning complex non-linear relationships within data through its hidden layers (figure 3). For instance, a study by Guo *et al*., (2024) adopted the MLP model in the prediction of early recurrence in local advanced gastric cancer, after the patient have been subjected to gastrectomy.

Multi DeepSurv model is another deep learning approach applicable in the analysis and prediction of survival likelihood. The model utilizes multiple data modalities, such as the clinical data and genetic information. Research conducted by Mao & Liu (2024) used the Multi DeepSurv model in carrying out survival analysis of the gastric cancer cases. Other models include the TLS-WSI Model, which was applied in the recurrence after postoperative, and demonstrated superior performance [39].

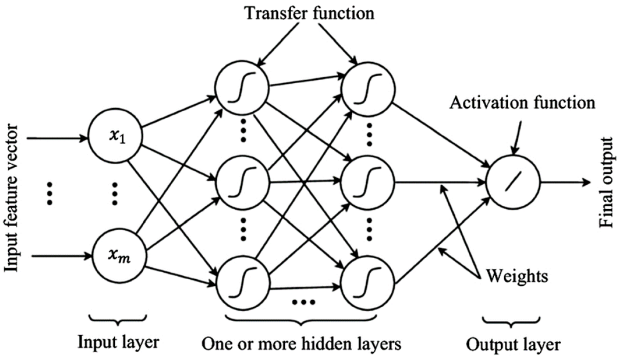

Figure 4. An architecture of multi-layer perceptron (MLP) model. This figure demonstrates the architecture of the Multi-Layer Perceptron (MLP) model, a deep learning approach used in gastric cancer recurrence prediction. The diagram shows the multiple layers of neurons in the MLP, highlighting how data flows through the network and how hidden layers learn complex patterns in the data. The figure is key in illustrating how MLP models handle non-linear relationships between clinical, genomic, and imaging data for gastric cancer prediction.

Radiotherapy in gastric cancer: Predictive models and clinical applications

Radiotherapy is a cornerstone of treatment for advanced gastric cancer, especially in cases where surgery is not feasible or tumors are inoperable. However, the effectiveness of radiotherapy is heavily dependent on the ability to precisely target tumors while minimizing damage to surrounding healthy tissue. Predictive models derived from public databases, leveraging artificial intelligence (AI), machine learning (ML), and deep learning (DL), are playing a pivotal role in optimizing radiotherapy treatment planning and enhancing clinical outcomes for gastric cancer patients.

Role of predictive models in radiotherapy planning

One of the most significant applications of predictive models in radiotherapy is the optimization of treatment planning. Radiotherapy planning involves determining the optimal dose of radiation and the best approach to deliver this dose to the tumor while avoiding critical structures. Predictive models, particularly radiomics-based models, analyze data from imaging modalities like CT, MRI, and PET scans to improve tumor delineation and reduce radiation exposure to healthy tissues.

These models use advanced AI techniques such as convolutional neural networks (CNNs) and random forests to process medical imaging data and enhance the accuracy of tumor segmentation. By accurately identifying tumor boundaries and assessing their characteristics, these models help clinicians plan

[DOI: 10.61186/ijrr.23.4.38]

more effective radiotherapy treatments. For instance, CNNs can detect and classify gastric tumors from endoscopic images, offering precise guidance for radiation treatment [31]. This not only ensures better targeting of tumors but also minimizes the risks associated with unnecessary radiation exposure to non-cancerous tissues.

Predicting treatment response to radiotherapy

Predictive models also play a crucial role in forecasting how individual patients will respond to radiotherapy. Since gastric cancer shows varying sensitivity to radiation, it is essential to personalize treatment to maximize effectiveness and minimize adverse effects. Machine learning models, such as support vector machines (SVMs), XGBoost, and deep learning algorithms, can predict how different tumor types will respond to radiation therapy based on clinical data, histopathological information, and imaging features.

For example, models like the Multi DeepSurv model integrate clinical, genomic, and imaging data to predict treatment outcomes and assess the likelihood of tumor regression after radiotherapy [27]. These predictive models not only help in identifying which patients will respond best to radiotherapy but also assist in adjusting radiation doses and treatment schedules, ensuring more personalized care.

Early detection of radiotherapy treatment failure

Monitoring the effectiveness of radiotherapy over time is crucial to ensure successful treatment of gastric cancer. Predictive models have emerged as valuable tools for tracking treatment progress by analyzing follow-up imaging data to detect early signs of tumor recurrence or treatment failure. These models can identify subtle changes in tumor size, shape, or metabolic activity that may not be readily apparent to clinicians. Machine learning algorithms, for example, evaluate longitudinal changes in tumor morphology or metabolic patterns and predict the likelihood of resistance to ongoing radiation therapy. Early detection of treatment failure through such predictive analytics enables clinicians to promptly modify therapeutic strategies, potentially switching to alternative treatments such as chemotherapy or immunotherapy. This proactive approach helps prevent unnecessary exposure to ineffective radiation cycles and improves patient outcomes.

Optimizing radiation dose and minimizing toxicity

Delivering an optimal radiation dose that effectively targets the tumor while minimizing damage to surrounding healthy tissues is a significant challenge in radiotherapy. Predictive models, particularly those based on radiomics, analyze imaging data to capture tumor heterogeneity and forecast dose distributions that maximize tumor control while sparing normal tissue. Additionally, integrating genomic data-such as immune-related signatures and genetic biomarkers-enhances these predictions by identifying patients' individual radiosensitivity or resistance. This personalized approach allows clinicians to adjust radiation doses to reduce the risk of radiation-induced toxicities, such as esophagitis or bowel complications, while maintaining treatment efficacy. For example, predictive models that assess genetic susceptibility to side effects can inform radiation planning, ensuring safer and more precise treatment tailored to each patient's biological profile.

Future directions and challenges

While predictive models have shown promise in optimizing radiotherapy for gastric cancer, several challenges remain before widespread clinical implementation. A primary hurdle is the integration of heterogeneous data sources—including clinical records, imaging, and genomic profiles—into cohesive, robust predictive algorithms. Additionally, the interpretability of AI and machine learning models is vital; clinicians must understand the rationale behind model predictions to build trust and facilitate decision-making. Future research must focus on improving model generalization across diverse patient populations, enhancing data quality, and addressing ethical considerations such as patient data privacy and consent. Establishing standardized protocols and frameworks for integrating AI-driven tools into clinical workflows will be essential to ensure consistent, reliable, and ethical use of these technologies in routine practice.

Radiotherapy techniques and predictive model applications

The effectiveness of radiotherapy and patient response vary significantly based on the radiotherapy technique used. Predictive models can provide critical support in tailoring these approaches:

- **Conventional Radiotherapy (CRT):** CRT typically delivers high doses of radiation directly to the tumor but can lack precision, potentially exposing surrounding healthy tissues to collateral damage. Predictive models optimize CRT by determining the most effective radiation dose personalized to each patient, reducing toxicity while maintaining tumor control.
- **Intensity-Modulated Radiotherapy (IMRT):** IMRT allows modulation of radiation beam intensity, providing more precise targeting of tumors. Predictive analytics guide dose distribution and beam shaping to maximize tumor irradiation while sparing nearby critical structures, such as the stomach and adjacent organs. This precision is particularly beneficial for tumors located near sensitive tissues.
- **Stereotactic Body Radiotherapy (SBRT):** SBRT

[DOI: 10.61186/ijrr.23.4.38]

delivers highly focused, high-dose radiation in fewer treatment sessions. Predictive models incorporating tumor geometry and spatial relationships to critical anatomy assist in optimizing dose delivery, minimizing side effects, and enhancing treatment efficacy by concentrating radiation on the tumor while protecting surrounding healthy tissues.

- **Proton Therapy:** Utilizing protons instead of X-rays, proton therapy offers superior dose conformity with reduced exit dose beyond the tumor. Predictive models help determine the ideal use of proton therapy by analyzing tumor size, location, and biological response, ensuring precise delivery and minimizing damage to normal tissue.
- **Chemoradiation Therapy:** Combining chemotherapy with radiotherapy enhances the overall therapeutic effect by sensitizing tumors to radiation. Predictive models evaluate the synergistic effects of these combined treatments, allowing clinicians to tailor regimens based on individual patient responses and optimize timing, dosage, and sequencing to maximize treatment success.

DISCUSSION

The application of predictive models in gastric cancer radiotherapy has evolved significantly, yet the choice and justification of specific models remain highly context-dependent. This discussion synthesizes the comparative strengths, limitations, and clinical implications of statistical, machine learning, and deep learning models as explored in this review.

Traditional statistical models, such as logistic regression and Cox proportional hazards models, remain widely used due to their interpretability and ease of implementation. These models are effective when datasets are well-structured and relatively small. However, their predictive performance may be limited in capturing complex, nonlinear interactions within multi-modal data. For example, while logistic regression is sufficient for basic classification tasks such as evaluating risk factors [18], it falls short in comparison to more sophisticated techniques in predictive accuracy and adaptability to high-dimensional data.

Machine learning models, particularly random forests and XGBoost, have demonstrated enhanced performance in several clinical prediction tasks. Park *et al.* (2025) and Zhou *et al*. (2025) showed these models achieving AUCs approaching 0.93, which is notably higher than traditional methods. Their ensemble nature allows for the capture of non-linear dependencies and interactions, improving risk stratification, treatment response prediction, and survival analysis.

Deep learning models, particularly CNNs and LSTMs, offer unparalleled accuracy in image analysis and time-series predictions. CNNs have shown high effectiveness in endoscopic and histopathological image analysis, achieving accuracy levels exceeding 99% [28, 31]. However, these models are typically data-hungry, computationally intensive, and less interpretable than their traditional counterparts, which may limit their routine clinical adoption without robust interpretability tools.

Model selection should align with the type and quality of available data, the clinical question at hand, and the requirement for interpretability. For instance, statistical models are more appropriate in settings where model transparency is essential, such as identifying survival factors or risk stratification based on limited clinical variables.

Machine learning models are suited for complex, moderately sized datasets involving a mix of clinical and genomic features, especially when predictive accuracy is a priority. They are increasingly used for radiotherapy treatment optimization and predicting treatment response.

Deep learning models are the most effective for tasks requiring the analysis of high-dimensional unstructured data such as imaging. Their ability to automate feature extraction and handle large volumes of image or time-series data makes them ideal for tumor classification and recurrence prediction, although they require careful consideration of explainability.

The integration of predictive models into radiotherapy workflows is most advanced with radiomics-based and deep learning models. CNNs have proven effective in segmenting gastric tumors from imaging data, enabling precise tumor delineation (Wu *et al*., 2025). Machine learning models such as random forests and XGBoost contribute significantly to treatment personalization by predicting radiotherapy response based on clinical and genomic profiles [5, 33]. Furthermore, models like Multi DeepSurv, which integrate multi-modal data, enable individualized survival prediction and treatment outcome modeling, thereby enhancing evidence-based radiation planning.

Despite notable advancements, the literature reveals several gaps. Few studies directly compare model performances across diverse datasets, and there is limited external validation of many proposed models. Additionally, studies often focus on single modalities (e.g., only imaging or only genomics), despite evidence suggesting that multi-modal integration improves predictive robustness.

A recurring challenge is the interpretability of complex models, which can hinder clinical trust and adoption. The “black box” nature of many AI-based tools necessitates explainability frameworks for safe and responsible integration into clinical decision-making. Additionally, the heterogeneity and quality of public database inputs, combined with data privacy and security concerns, must be addressed through

[DOI: 10.61186/ijrr.23.4.38]

standardized data governance and compliance with regulatory guidelines.

Moreover, while predictive models offer valuable clinical support, they must be viewed as complementary tools rather than replacements for clinical judgment. Human oversight remains essential to ensure context-appropriate and ethically sound application of AI in oncology.

CONCLUSION

Predictive models, particularly those derived from public databases, offer significant potential in improving the management of gastric cancer, especially in the context of radiotherapy. By enabling early diagnosis, personalized treatment planning, and accurate prediction of treatment responses and recurrence, these models support more effective and individualized care. This study highlights the importance of integrating predictive tools into clinical workflows to enhance radiotherapy outcomes and improve patient survival and quality of life.

Challenges and implications

Despite the fact that predictive models are promising in gastric cancer research, several challenges could be highlighted. For an effective and efficient adoption of application of these models, these challenges need to be addressed, as they have critical implication on the gastric cancer patient care as well as the future research. The first challenge is the data availability, quality and standardization. This is a common challenge of public databases, where they suffer from incomplete, imbalanced, or inconsistent data, which has severe limitations to models' generalizability (Sasagawa *et al*., 2024). Another challenge identified is the heterogenous data sources, where there are databases have different formats for data such as clinical records, imaging, genomics, and molecular data.

The second challenge highlighted is the model interpretability and clinical trust issues. These models have demonstrated high prediction accuracies. However, their 'black box nature' makes them quite difficult in their clinical applications and interpretations by healthcare professionals (Mao & Liu, 2024). Additionally, there are concerns raised regarding the clinical decision making and liability, due to the limited transparency in the AI driven predictions. Another critical challenge moted was the ethical and regulatory concerns. There are various regulatory and ethical aspects and concerns that must be adhered to, in regards health data, information and healthcare models. Data privacy and patient confidentiality is critical to consider. As well, medical models developed through AI should observe and comply with set regulatory guidelines. This is because biasness in predictive models may lead to disparities in gastric cancer diagnosis and treatment, negatively affecting the patients and clinical outcomes.

In addressing these challenging and looking to the future, several implication recommendations are developed. First, the researchers utilizing public database prediction models should focus on developing standardized, high-quality, and publicly available databases for gastric cancer. These databases should be accompanied with improved data-sharing policies and federated learning techniques, can help overcome these barriers. This research recommends the need for development of explainable AI (XAI) frameworks, aimed to improve transparency and understanding to machine learning models. These frameworks could include Shapley Additive Explanations (SAE). There is need for a comprehensive and active bias detection framework, and patient data and AI guidelines. These frameworks should be adopted and incorporated in predictive models' development and application, to fair, unbiased, and regulatory-compliant AI applications. Notably, while these models are quire relevant, they should serve as complementary tools in clinical research and never replace clinical judgement.

***Acknowledgments*:** The authors would like to thank the Clinical Oncology School of Fujian Medical University and Fujian Cancer Hospital for their support and access to resources during the preparation of this manuscript.

***Conflict of Interest*:** The authors declare no conflicts of interest related to this work.

***Funding*:** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

***Ethical Consideration*:** This study is a systematic review and did not involve direct research on human or animal subjects. Therefore, ethical approval was not required.

***Authors' Contributions*:** S.C.: Conceptualized the study, conducted data analysis, and drafted the manuscript. S.W.: Conducted literature review and contributed to data interpretation. Y.W.: Provided technical support and critically reviewed the manuscript. L.C.: Supervised the project and participated in final revisions. Z.L.: Coordinated the correspondence, formatted the manuscript, and finalized submission.

***AI Usage Declaration*:** During the preparation of this manuscript, AI tools were used to only assist with language editing, grammar correction, and rephrasing for clarity. All content was reviewed and verified by the authors to ensure accuracy and originality.

***Written permission for use of figure*:** I, Saghir Ahmed, hereby grant permission to Zhitao Lin (email: hxsf803@163.com) to reproduce Figure 3: XGBoost Model Architecture, originally published by others, in

[DOI: 10.61186/ijrr.23.4.38]

the manuscript currently under preparation. This permission is granted for academic and non-commercial use, with the understanding that full credit and proper citation will be given to the original source of the figure. This permission is subject to any conditions or requirements specified by the original publisher, and I trust that all necessary acknowledgments and licensing terms will be properly observed.

REFERENCES

1. Tan N, Wu H, Cao M, Yang F, Yan X, He S, *et al.* (2024) Global, regional, and national burden of early-onset gastric cancer. C*ancer Biol Med,* ***21****(8): 667-78*.
2. rawla p and barsouk a (2019) epidemiology of gastric cancer: global trends, risk factors and prevention. *prz gastroenterol,* ***14****(1): 26-38*.
3. Kang K, Bagaoisan Ma, Zhang Y (2024) Unveiling the younger face of gastric cancer: a comprehensive review of epidemiology, risk factors, and prevention strategies. *Cureus,* ***16****(6): e62826.*
4. Yu JI (2023) Role of adjuvant radiotherapy in gastric cancer. *j Gastric Cancer,* ***23****(1): 194-206*.
5. Chen HHW and Kuo MT (2017) improving radiotherapy in cancer treatment: Promises and challenges. *Oncotarget,* ***8****(37): 62742-58*.
6. Lei C, Sun W, Wang K, Weng R, Kan X, li R (2025) Artificial intelligence-assisted diagnosis of early gastric cancer: present practice and future prospects. *Ann Med,* ***57****(1): 2461679*.
7. Eloranta S and Boman M (2022) Predictive models for clinical decision making: deep dives in practical machine learning. *J Intern Med,* ***292****(2): 278-95*.
8. Li S and Zhou B (2022) A review of radiomics and genomics applications in cancers: the way towards precision medicine. *Radiat Oncol,* ***17****(1): 217*.
9. Isaksson LJ, Pepa M, Zaffaroni M, Marvaso G, Alterio D, Volpe S, *et al.* (2020) Machine learning-based models for prediction of toxicity outcomes in radiotherapy. *Front Oncol,* ***10****: 790*.
10. Mcgrail KM, Teng J, Bentley C, O'doherty KC, Burgess MM (2024) Research data use in a digital society: a deliberative public engagement. *Int j Popul Data Sci,* ***9****(1): 2372*.
11. Brito DMS, Lima OG, Mesquita FP, Da Silva EL, De Moraes MEA, Burbano RMR, *et al.* (2023) A shortcut from genome to drug: the employment of bioinformatic tools to find new targets for gastric cancer treatment. *Pharmaceutics,* ***15****(****9****): 2303*.
12. Zhang Z, Shao X, Wu H, Su X, Wang G, Zhu L, *et al.* (2023) Prognostic impact of aberrantly expressed protein-coding gene associated with gastric cancer's regulatory t cells, based on online databases. *Altern Ther Health Med,* ***29****(3): 160-5.*
13. Ramesh P, Nisar M, Neha, Ammankallu S, Babu S, Nandakumar R, *et al.* (2024) Delineating protein biomarkers for gastric cancers: a catalogue of mass spectrometry-based markers and assessment of their suitability for targeted proteomics applications. *J Proteomics,* ***306****: 105262*.
14. Conroy MC, Lacey B, Bešević J, Omiyale W, Feng Q, Effingham M, *et al.* (2023) UK biobank: a globally important resource for cancer research. *Br J Cancer,* ***128****(4): 519-27*.
15. Almeda AF, Grimes SM, lee H, Greer S, Shin G, Mcnamara M, *et al.* (2022) The gastric cancer registry: a genomic translational resource for multidisciplinary research in gastric cancer. *Cancer Epidemiol Biomarkers Prev,* ***31****(9): 1693-700.*
16. Martin AN, Chan NW, Cheung DC, Fong ZV (2024) A guide to large data sets for population-based cancer research: strengths, limitations, and pitfalls. *Cancer,* ***130****(22): 3802-14*.
17. Hicks-Courant K, Ko EM, Matsuo K, Melamed A, Nasioudis D, Rauh-Hain JA, *et al.* (2024) Secondary databases in gynecologic cancer research. *Int J Gynecol Cancer,* ***34****(10): 1619-29.*
18. Karamoozian A, Baneshi MR, Bahrampour A (2021) Short-term and long-term survival of patients with gastric cancer. *Gastroenterol Hepatol Bed Bench, 14(2): 115-22*.
19. Jin Q, Cao J, Wang G, He N (2025) Neoadjuvant chemotherapy can effectively avoid unnecessary extended resection for gastric cancer with clinical evidence of duodenum or pancreas head involvement. *J Cancer,* ***16****(4): 1181–8*.
20. Allen CJ, Pointer DT, Jr., Blumenthaler AN, Mehta Rj, Hoffe SE, Minsky BD, *et al.* (2021) Chemotherapy versus chemotherapy plus chemoradiation as neoadjuvant therapy for resectable gastric adenocarcinoma: a multi-institutional analysis. *Ann Surg,* ***274****(4): 544–8*.
21. Sabbagh S, Jabbal IS, Iska S, Mohanna M, Itani M, Dominguez B, *et al.* (2023) Survival analysis comparing neoadjuvant chemoradiotherapy versus perioperative chemotherapy for locoregional gastric cancer. *Journal of Clinical Oncology,* ***41****(16-suppl): e16097–e*.
22. Darang E, Pezeshkian- Z, Mirhoseini SZ, Ghovvati S (2023) Bioinformatics and pathway enrichment analysis identified hub genes and potential biomarker for gastric cancer prognosis. *Front Oncol,* ***13****: 1187521*.
23. Shamsi S, Larizadeh MH, Bahador M, Nouri M (2024) A five-year survival analysis of patients with gastric cancer in kerman province. *Frontiers in Biomedical Technologies,* ***11****(4): 607-615*
24. Park B, Kim CH, Jun JK, Suh M, Choi KS, Choi IJ, *et al.* (2024) A machine learning risk prediction model for gastric cancer with shapley additive explanations. *Cancer Res Treat,* ***57****(3): 821-829*.
25. Xu L and Guo C (2024) Imbalanced survival prediction for gastric cancer patients based on improved xgboost with cost sensitive and focal loss. *Expert Systems,* ***41****(11): e13666*.
26. Zhou H, Gu Q, Bao R, Qiu L, Zhang Y, Wang F, *et al.* (2024) Machine learning based models for predicting presentation delay risk among gastric cancer patients. *Front Oncol,* ***14****: 1503047*.
27. Mao S and Liu J (2025) Mulitdeepsurv: survival analysis of gastric cancer based on deep learning multimodal fusion models. *Biomed Opt Express,* ***16****(1): 126-41*.
28. Maheswari R, Allah IM, Mukesh B, Britto SS, Editors. Discerning and leveraging gastric cancer using deep analytics algorithms. 2024 International Conference on Signal Processing and Advance Research in Computing (SPARC); 2024 12–13 sept. 2024.
29. Wu Z, Wang T, Lan J, Wang J, Chen G, Tong T, *et al.* (2025) Deep learning-based prediction of her2 status and trastuzumab treatment efficacy of gastric adenocarcinoma based on morphological features. *j Transl Med,* ***23****(1): 13.*
30. Thota M, Mahajan S, Boppana SH, Boppana SL, Komati SSK, Mintz D (2025) Predicting mortality in upper gastrointestinal tract cancer using long short-term memory neural networks. *Journal of Clinical Oncology,* ***43****(4_Suppl): 340-340*
31. Bhardwaj P, Kim S, Koul A, Kumar Y, Changela A, Shafi J, *et al.* (2024) Advanced cnn models in gastric cancer diagnosis: enhancing endoscopic image analysis with deep transfer learning. *Front Oncol,* ***14****: 1431912.*
32. Li S, Deng YQ, Zhu ZL, Hua HL, Tao ZZ (2021) A comprehensive review on radiomics and deep learning for nasopharyngeal carcinoma imaging. *Diagnostics (Basel), 11(9): 1523.*
33. Sasagawa S, Honma Y, Peng X, Maejima K, Nagaoka K, Kobayashi Y, *et al.* (2025) Predicting chemotherapy responsiveness in gastric cancer through machine learning analysis of genome, immune, and neutrophil signatures. *Gastric Cancer,* ***28****(2): 228–44*.
34. Rubini D, Gagliardi F, Menditti VS, D'ambrosio L, Gallo P, D'onofrio I, *et al.* (2024) Genetic profiling in radiotherapy: a comprehensive review. *Front Oncol,* ***14****: 1337815*.
35. Yu S, Jiang H, Xia J, Gu J, Chen M, Wang Y, *et al.* (2025) Construction of machine learning-based models for screening the high-risk patients with gastric precancerous lesions. *Chin Med,* ***20****(1): 7*.
36. Eweje f, Li Z, Gopaulchan M, Attaranzadeh A, Kloker L, Jiang Y, *et al.* (2024) Use of artificial intelligence–based digital pathology to predict outcomes for immune checkpoint inhibitor therapy in advanced gastro-esophageal cancer. *j Clin Oncol,* ***42*** *(Suppl 16; Abstr 4013).*
37. Bai Z, Wang H, Han J, An J, Yang Z, Mo X (2024) Multiomics integration and machine learning reveal prognostic programmed cell death signatures in gastric cancer. *Sci Rep,* ***14****(1): 31060.*
38. Zubair M, Owais M, Mahmood T, Iqbal S, Usman SM, Hussain I (2024) Enhanced gastric cancer classification and quantification interpretable framework using digital histopathology images. Sci Rep, **14**(1): 22533.
39. Huang L, Feng B, Yang Z, Feng ST, Liu Y, Xue H, *et al.* (2025) A transfer learning radiomics nomogram to predict the postoperative recurrence of advanced gastric cancer. *J Gastroenterol Hepatol,* ***40*** *(4): 844-54*.

[DOI: 10.61186/ijrr.23.4.38]