

Research Article

TabTransformer-XGBoost Hybrid Framework for Gastric Cancer Classification Using Tabular Clinical Data with SHAP-Based Explainability

Adetokunbo MacGregor John-Otumu^{1,*} , Oyewole Dokun² , Ferguson Ogene³ ,
Chukwukere Andrew Okpani¹ , Victor Chisom Nwachukwu¹ ,
Rebecca Ogechi Ekemonye⁴ 

¹Department of Information Technology, Federal University of Technology, Owerri, Nigeria

²Department of Computing, Glorious Vision University, Ogwa, Nigeria

³Department of Computing Science & Information Technology, Petroleum Training Institute, Effurun, Nigeria

⁴Department of Software Engineering, Claretian University of Nigeria, Nekede, Nigeria

Abstract

Gastric cancer remains an important health challenge, and accurate classification using routinely available clinical information could support timely clinical decision-making. However, conventional machine learning approaches may have difficulty capturing complex relationships among heterogeneous clinical variables, while highly predictive models may provide limited insight into how individual features influence their decisions. This study proposes an explainable TabTransformer-XGBoost hybrid framework for gastric cancer classification using structured clinical data. A publicly available gastric cancer dataset containing 212,354 patient records and 29 attributes was used. The proposed framework incorporates data cleaning, feature selection using mutual information, stratified data partitioning, SMOTE-based class balancing, contextual feature learning with TabTransformer, XGBoost classification, and SHAP-based explainability. The framework was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC and compared with conventional machine learning models, including Decision Tree, Random Forest, XGBoost, Naïve Bayes, Logistic Regression, and Multilayer Perceptron. The proposed framework achieved 90.13% accuracy, 86.17% precision, 90.13% recall, and an F1-score of 85.46%, outperforming the evaluated baseline models in the principal classification measures. Ablation analysis showed that XGBoost provided strong predictive performance, while the TabTransformer contributed contextual feature representations within the hybrid architecture. SHAP analysis further identified influential learned representations and demonstrated nonlinear relationships involving clinical variables such as age and alcohol consumption, thereby providing greater transparency into model predictions. The findings indicate that the proposed framework provides a promising basis for explainable gastric cancer classification from tabular clinical data, although independent multi-centre and real-world clinical validation are required before clinical deployment.

Keywords

Gastric Cancer Classification, TabTransformer, XGBoost, Explainable Artificial Intelligence (XAI), SHAP, Clinical Decision Support

*Correspondence: Adetokunbo MacGregor John-Otumu (adetokunbo.johnnotumu@futo.edu.ng)

Received: 19 September 2026; **Accepted:** 28 September 2026; **Published:** 20 October 2026



Copyright: © The Author(s), 2026. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Gastric cancer remains one of the leading causes of cancer-related mortality worldwide and continues to pose a major public health challenge due to its high incidence, late-stage diagnosis, and poor survival outcomes. According to recent clinical studies, gastric cancer often develops progressively through a sequence of pathological conditions that include chronic gastritis, chronic atrophic gastritis (CAG), intestinal metaplasia, and dysplasia before malignant transformation occurs [1, 2]. Persistent *Helicobacter pylori* infection, unhealthy dietary habits, smoking, alcohol consumption, genetic predisposition, and autoimmune disorders have all been identified as significant risk factors that contribute to disease progression [3, 4]. Consequently, accurate identification of individuals at high risk of gastric cancer is essential for timely clinical intervention, improved treatment outcomes, and reduced mortality.

Current diagnosis of gastric cancer primarily relies on upper gastrointestinal endoscopy followed by histopathological biopsy, which remains the clinical gold standard for disease confirmation [5]. Although highly effective, these procedures are invasive, expensive, time-consuming, and dependent on experienced gastroenterologists and pathologists for accurate interpretation. In many healthcare systems, particularly in low-resource environments, limited access to specialist expertise and diagnostic facilities often results in delayed detection and increased disease burden [6]. These challenges have motivated growing interest in intelligent computer-aided diagnostic systems capable of supporting clinicians through rapid, objective, and reliable prediction of gastric cancer using routinely available clinical information.

Recent advances in Artificial Intelligence (AI), particularly Machine Learning (ML) and Deep Learning (DL), have significantly transformed predictive healthcare by enabling computers to learn complex nonlinear relationships from heterogeneous clinical datasets. Several studies have successfully applied Logistic Regression, Support Vector Machines (SVM), Random Forest, eXtreme Gradient Boosting (XGBoost), Artificial Neural Networks (ANNs), and Convolutional Neural Networks (CNNs) for gastric cancer risk prediction and related gastrointestinal disease classification using demographic characteristics, laboratory biomarkers, electronic health records, lifestyle variables, and medical imaging data [7-12]. Furthermore, explainable Artificial Intelligence (XAI) techniques, particularly SHapley Additive exPlanations (SHAP), have improved the transparency of predictive models by identifying clinically important variables that influence model decisions, thereby enhancing clinician trust and supporting informed medical decision-making [4, 13]. Despite these advances, many existing approaches continue to rely on either conventional machine learning algorithms or standalone deep learning models, each exhibiting limitations in modelling complex

contextual relationships among heterogeneous clinical variables.

Transformer-based learning has recently emerged as an effective paradigm for analysing structured tabular data through self-attention mechanisms capable of learning contextual feature representations. Unlike conventional machine learning algorithms that process clinical variables largely in isolation, transformer architectures capture intricate dependencies among demographic characteristics, laboratory findings, lifestyle behaviours, and other patient-specific attributes. Nevertheless, standalone transformer models often require complementary classification mechanisms to fully exploit their learned representations for downstream prediction tasks. In addition, relatively few studies have explored transformer-based hybrid learning frameworks for gastric cancer classification using tabular clinical data, while comprehensive benchmarking against established machine learning models remains limited [10, 14, 15].

Motivated by these limitations, this study proposes an Explainable TabTransformer-XGBoost Hybrid Framework for Gastric Cancer Classification Using Tabular Clinical Data. The proposed framework combines the contextual feature representation capability of TabTransformer with the robust nonlinear classification strength of XGBoost to improve predictive performance on structured clinical datasets. To enhance model transparency and facilitate clinical interpretation, SHAP is incorporated to explain the contribution of individual clinical variables to classification decisions. The proposed framework is extensively evaluated against conventional machine learning models, standalone deep learning models, and several transformer-based hybrid architectures using standard performance metrics. The findings are expected to provide an accurate, interpretable, and clinically applicable AI-assisted framework capable of supporting reliable gastric cancer classification while contributing to the broader adoption of trustworthy artificial intelligence in precision healthcare.

2. Related Work and Research Gap Analysis

Artificial Intelligence (AI) has become an important enabling technology for computer-aided diagnosis and prognosis of gastric diseases by facilitating intelligent analysis of heterogeneous clinical, laboratory, and imaging data. In recent years, researchers have increasingly applied machine learning, deep learning, and transformer-based learning techniques to improve gastric cancer prediction, risk stratification, and clinical decision support. Although these approaches have demonstrated promising performance, their effectiveness largely depends on the adopted learning paradigm, feature representation strategy, and quality of the underlying clinical data [7, 8].

2.1. Conventional ML for Gastric Cancer Classification

Conventional machine learning techniques remain widely adopted for gastric cancer prediction using structured clinical and laboratory datasets. Algorithms including Logistic Regression (LR), Decision Tree (DT), Naïve Bayes (NB), k-Nearest Neighbour (KNN), Support Vector Machine (SVM), Random Forest (RF), and eXtreme Gradient Boosting (XGBoost) have been successfully employed to identify high-risk patients based on demographic information, serum biomarkers, lifestyle factors, and medical history [7-10]. Among these methods, ensemble learning algorithms such as Random Forest and XGBoost generally achieve superior predictive performance because of their ability to model nonlinear feature interactions while reducing overfitting. However, most conventional machine learning models rely heavily on manually engineered features and typically process clinical variables independently, thereby limiting their ability to capture complex contextual relationships that naturally exist among heterogeneous patient characteristics.

2.2. DL and Explainable AI for Gastric Cancer Diagnosis

DL has significantly advanced automated gastric cancer diagnosis by learning hierarchical feature representations directly from structured clinical records and medical images. Artificial Neural Networks (ANNs), Multilayer Perceptrons (MLPs), and Convolutional Neural Networks (CNNs) have demonstrated encouraging performance in gastric cancer risk prediction and endoscopic lesion classification while reducing dependence on handcrafted feature engineering [4, 6, 11]. More recently, explainable Artificial Intelligence (XAI) techniques, particularly SHapley Additive exPlanations (SHAP), have enhanced the interpretability of predictive models by quantifying the contribution of individual clinical variables to

model decisions, thereby improving clinician confidence and supporting trustworthy AI-assisted diagnosis [5, 12, 16, 17]. Despite these advances, many deep learning models remain computationally demanding, require relatively large labelled datasets, and often exhibit limited transparency when deployed in clinical practice.

2.3. Transformer-Based Learning for Tabular Clinical Data

Transformer-based learning has recently emerged as a promising paradigm for analyzing structured tabular clinical data through self-attention mechanisms capable of modelling contextual relationships among heterogeneous patient variables. Unlike conventional machine learning algorithms that generally process features independently, transformer architectures learn contextual feature embeddings by capturing interactions among demographic characteristics, laboratory biomarkers, lifestyle behaviors, and clinical observations [13-15]. These contextual representations have demonstrated improved predictive capability across several healthcare applications. Nevertheless, transformer-based learning remains relatively underexplored for gastric cancer classification using structured clinical datasets. Furthermore, limited studies have investigated combining transformer-based contextual feature learning with robust ensemble classifiers to enhance predictive accuracy while preserving model interpretability.

To highlight the strengths and limitations of representative AI-based approaches reported in the literature, Table 1 demonstrates that previous studies have primarily focused on conventional machine learning, deep learning, or standalone explainable AI models. Although these approaches have achieved encouraging diagnostic performance, limited attention has been devoted to integrating transformer-based contextual representation learning with ensemble classification and explainable decision support within a unified framework for gastric cancer prediction.

Table 1. Comparative Analysis of Representative Ai-Based Gastric Cancer Classification Studies.

Ref.	Technique	Data Type	Explainable	Hybrid	Main Limitation
Li et al. [7]	ML Models	Clinical Biomarkers	✗	✗	Manual feature engineering
Park et al. [5]	XGB + SHAP	Clinical Data	✓	✗	Standalone boosting model
Shi et al. [11]	CNN	Endoscopic Images	✗	✗	Requires large, labelled image datasets
Turtoi et al. [12]	AI-Based Diagnostic Models	Clinical & Imaging	Partial	✗	Limited transformer-based analysis
Proposed Framework	TabTransformer-XGBoost + SHAP	Tabular Clinical Data	✓	✓	Integrates contextual representation learning, ensemble classification, comprehensive benchmarking, and explainability

2.4. Research Gap Analysis

A critical review of the literature reveals several important research gaps. First, most existing studies investigate either standalone machine learning algorithms or standalone deep learning models without exploiting the complementary strengths of transformer-based contextual feature learning and ensemble classification. Second, transformer-based architectures remain underexplored for gastric cancer classification using structured clinical datasets despite their superior capability for modelling feature dependencies. Third, comprehensive benchmarking involving conventional machine learning algorithms, standalone transformer models, and transformer-based hybrid architectures under identical experimental settings is rarely reported. Finally, although explainable AI has gained considerable attention in medical decision support, its integration within transformer-based hybrid learning frameworks for gastric cancer classification remains limited.

To address these limitations, this study proposes an Explainable TabTransformer-XGBoost Hybrid Framework for Gastric Cancer Classification Using Tabular Clinical Data. The proposed framework combines the contextual representation capability of TabTransformer with the robust nonlinear classification strength of XGBoost while incorporating SHAP-based explainability to generate transparent and clinically interpretable predictions. Furthermore, the proposed model is extensively benchmarked against conventional machine learning algorithms, standalone deep learning models, and multiple transformer-based hybrid architectures to validate its predictive effectiveness, robustness, and practical applicability for intelligent gastric cancer classification.

3. Methodology

This section presents the methodology adopted for developing the proposed TabTransformer-XGBoost framework, including dataset preparation, contextual feature learning, model training, explainability using SHAP, and performance evaluation for gastric cancer classification.

3.1. Dataset Source and Description

This study employed the publicly available Gastric Cancer (GC) Dataset obtained from the Kaggle repository

(<https://www.kaggle.com/datasets/datasetengineer/gastric-cancer-gc-dataset>) The dataset consists of 212,354 patient records with 29 attributes, including a binary target variable for gastric cancer classification. The predictor variables comprise demographic characteristics, lifestyle factors, clinical findings, diagnostic results, and molecular biomarkers, providing a comprehensive representation of patient health profiles. Prior to model development, the dataset underwent data cleaning, missing value treatment, categorical encoding, and feature normalization. Subsequently, the processed data were partitioned into training, validation, and testing subsets using a stratified sampling strategy to preserve the original class distribution.

A summary of the dataset is presented in Table 2. The dataset's large sample size and heterogeneous feature composition make it suitable for evaluating the proposed Explainable TabTransformer-XGBoost Hybrid Framework.

Table 2. Summary of the Gastric Cancer Dataset.

Attribute	Value
Source	Kaggle Gastric Cancer (GC) Dataset
Total Samples	212,354
Features	29
Target Variable	Label (0 = Negative, 1 = Positive)
Negative Cases	191,395 (90.13%)
Positive Cases	20,959 (9.87%)
Data Type	Mixed numerical and categorical

3.2. Proposed TabTransformer-XGBoost Hybrid Model

This section presents the proposed TabTransformer-XGBoost hybrid framework for gastric cancer classification. As shown in Figure 1, the framework consists of five major stages: dataset preparation, contextual feature representation using TabTransformer, classification using XGBoost, model explainability through SHAP, and statistical validation for performance assessment.

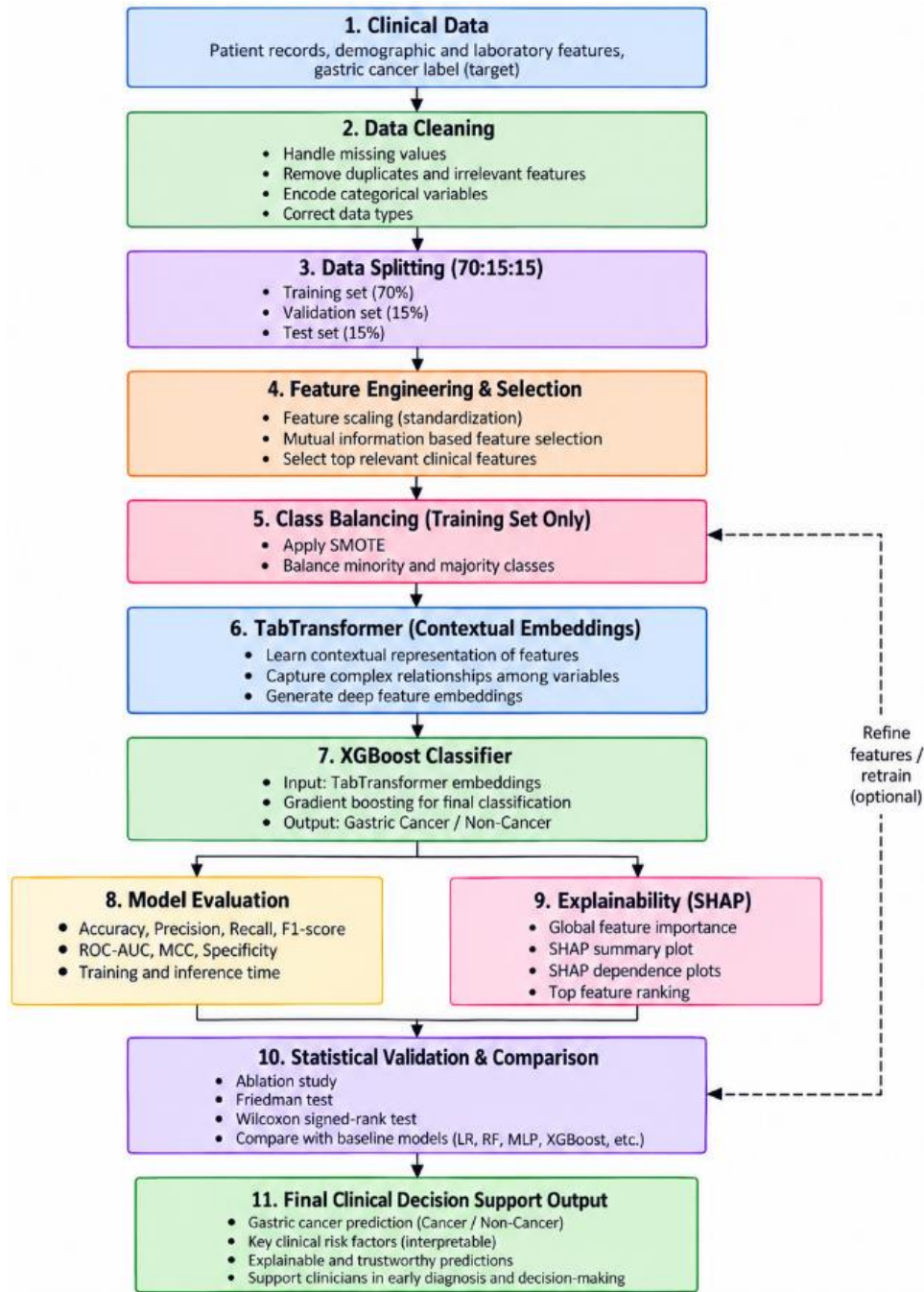


Figure 1. Proposed TabTransformer-XGBoost Framework for Gastric Cancer Classification.

Figure 1 illustrates the complete workflow of the proposed model. The process begins with the acquisition of clinical patient data comprising demographic information, laboratory measurements, and gastric cancer class labels. The raw dataset is subsequently preprocessed through data cleaning, including missing value handling, duplicate removal, categorical feature encoding, and data quality correction. The cleaned dataset is then partitioned into training (70%), validation (15%), and testing (15%) subsets. Feature engineering and feature selection are performed to standardize the data and retain the most informative clinical variables, after which SMOTE is applied exclusively to the training set to address class imbalance.

The balanced training data are fed into the TabTransformer, which learns contextual embeddings by modelling relationships among clinical features through self-attention. These high-level feature representations are subsequently provided to the XGBoost classifier, which performs the final gastric cancer prediction. The model is evaluated using standard classification and computational performance metrics, while SHAP is employed to explain both global and local feature contributions, thereby improving the interpretability of the prediction process. Finally, statistical validation techniques, including ablation analysis and significance testing, are conducted to assess the robustness of the proposed framework and

compare its performance with baseline machine learning models. The framework ultimately produces an explainable clinical decision support output that predicts gastric cancer status, identifies influential risk factors, and supports reliable clinical decision-making.

3.3. Mathematical Formulation of the Proposed TabTransformer-XGBoost Framework

Let the original gastric cancer clinical dataset be represented as

$$D = \{(x_i, y_i)\}_{i=1}^N \quad (1)$$

where

$$x_i = [x_{i1}, x_{i2}, \dots, x_{id}], y_i \in \{0, 1\} \quad (2)$$

Here, x_i denotes the i^{th} patient record containing d clinical features, while y_i represents the corresponding class label.

1) Data Preprocessing

Following data cleaning and preprocessing, the transformed dataset becomes

$$D_p = \{(x_i^p, y_i)\}_{i=1}^{N_p} \quad (3)$$

where x_i^p denotes the preprocessed feature vector.

Missing numerical values are imputed using the median:

$$x_{ij} = \begin{cases} x_{ij}, & x_{ij} \neq \emptyset \\ \text{median}(x_j), & x_{ij} = \emptyset \end{cases} \quad (4)$$

Missing categorical values are replaced using the mode:

$$x_{ij} = \begin{cases} x_{ij}, & x_{ij} \neq \emptyset \\ \text{mode}(x_j), & x_{ij} = \emptyset \end{cases} \quad (5)$$

Feature standardization is performed as

$$x_{ij}^s = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (6)$$

where μ_j and σ_j denote the mean and standard deviation of feature j , respectively.

2) Feature Selection

Let the complete feature space be

$$F = \{f_1, f_2, \dots, f_d\} \quad (7)$$

The selected feature subset is defined as

$$F_s = \{f_j \in F \mid I(f_j) \geq \tau\} \quad (8)$$

where $I(f_j)$ is the feature importance score and τ is the selection threshold.

The reduced dataset is expressed as

$$D_f = \{(x_i^f, y_i)\}_{i=1}^{N_p} \quad (9)$$

where

$$x_i^f = [x_{i1}^f, x_{i2}^f, \dots, x_{im}^f], m < d \quad (10)$$

3) Dataset Partitioning

The selected dataset is partitioned into training, validation, and testing subsets:

$$D_f = D_{train} \cup D_{val} \cup D_{test} \quad (11)$$

subject to

$$D_{train} \cap D_{val} \cap D_{test} = \emptyset \quad (12)$$

with

$$|D_{train}| = 0.70 |D_f| \quad (13)$$

$$|D_{val}| = 0.15 |D_f| \quad (14)$$

$$|D_{test}| = 0.15 |D_f| \quad (15)$$

4) SMOTE-Based Class Balancing

SMOTE is applied only to the training dataset:

$$D_{train}^* = \text{SMOTE}(D_{train}) \quad (16)$$

A synthetic minority-class sample is generated as

$$x_{new} = x_i + \delta(x_{nn} - x_i) \quad (17)$$

where

$$0 \leq \delta \leq 1 \quad (18)$$

The balanced training dataset becomes

$$D_{train}^* = \{(x_i^*, y_i^*)\}_{i=1}^{N_p^*} \quad (19)$$

5) TabTransformer Representation Learning

Each selected feature is embedded into a dense vector as

$$e_j = W_j x_j^* + b_j \quad (20)$$

The embedded feature matrix is

$$E = [e_1, e_2, \dots, e_m] \quad (21)$$

The query, key, and value matrices are computed as

$$Q = EW_Q, K = EW_K, V = EW_V \quad (22)$$

The scaled dot-product attention is

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (23)$$

Multi-head attention is computed as

$$\text{MHA}(E) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W_O \quad (24)$$

where

$$\text{head}_r = \text{Attention}(EW_Q^r, EW_K^r, EW_V^r) \quad (25)$$

The feed-forward transformation is

$$\text{FFN}(z) = \sigma(zW_1 + b_1)W_2 + b_2 \quad (26)$$

The contextual representation produced by the TabTransformer is

$$H_i = \text{TabTransformer}(x_i^*) \quad (27)$$

6) XGBoost Classification

The contextual embedding is supplied to the XGBoost classifier:

$$\hat{y}_i = \sigma(\sum_{k=1}^K f_k(H_i)) \quad (28)$$

where f_k denotes the k^{th} decision tree.

The sigmoid activation is

$$\sigma(z) = \frac{1}{1+e^{-z}} \quad (29)$$

The predicted class is obtained as

$$\hat{c}_i = \begin{cases} 1, & \hat{y}_i \geq 0.5 \\ 0, & \hat{y}_i < 0.5 \end{cases} \quad (30)$$

7) Model Optimization

The XGBoost objective function is

$$\mathcal{L}(\phi) = \sum_{i=1}^{N^*} l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (31)$$

where

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (32)$$

8) SHAP Explainability

The SHAP explanation model is expressed as

$$g(x_i) = \phi_0 + \sum_{j=1}^m \phi_j \quad (33)$$

The SHAP value of feature j is computed as

$$\phi_j = \sum_{S \subseteq F_S \setminus \{j\}} \frac{|S|!(m-|S|-1)!}{m!} [f(S \cup \{j\}) - f(S)] \quad (34)$$

3.4. Experimental Setup and Evaluation

This section presents the experimental setup and evaluation as represented in the subsections.

3.4.1. Experimental Environment

All experiments were conducted on a HP EliteBook 840 G7 laptop running the 64-bit Windows 11 Pro operating system. The system was powered by an Intel® Core™ i7-10610U processor operating at a base frequency of 1.80 GHz with a maximum turbo frequency of 4.90 GHz, supported by 16 GB DDR4 RAM and a 512 GB NVMe solid-state drive (SSD). The proposed hybrid framework was implemented using Python 3.14.x, with PyTorch employed for developing the TabTransformer component, XGBoost for gradient boosting classification, and SHAP for model explainability.

Data preprocessing and feature engineering tasks were performed using Pandas, NumPy, and Scikit-learn, while Imbalanced-learn was utilized to implement the Synthetic Minority Oversampling Technique (SMOTE) for addressing class imbalance. Data visualization was carried out using Matplotlib and Seaborn, whereas SciPy and Stats models were used to perform statistical significance analyses, including the Friedman, Wilcoxon Signed-Rank, and McNemar tests. A summary of the hardware and software environment employed in this study is presented in Table 3.

Table 3. Computer Configuration Used for Model Development And Evaluation.

Component	Specification
Computer Model	HP EliteBook 840 G7
Operating System	Windows 11 Pro (64-bit)
Processor (CPU)	Intel® Core™ i7-10610U CPU @ 1.80 GHz (Turbo up to 4.90 GHz)
CPU Architecture	64-bit
System Memory (RAM)	16 GB DDR4

Component	Specification
Storage	512 GB NVMe SSD
Programming Language	Python 3.14.x
Deep Learning Framework	PyTorch
Machine Learning Framework	XGBoost
Explainability Framework	SHAP
Data Processing Libraries	Pandas, NumPy, Scikit-learn
Class Balancing Library	Imbalanced-learn (SMOTE)
Visualization Libraries	Matplotlib, Seaborn
Statistical Analysis Libraries	SciPy, Statsmodels

3.4.2. Training Procedure

This subsection presents the training procedures of the Proposed TabTransformer-XGBoost Framework as summarized in Table 4.

Table 4. Training Procedure of the Proposed TabTransformer-XGBoost Hybrid Framework.

Stage	Description
Feature selection	SelectKBest using Mutual Information criterion
Number of selected features	Top 28 most informative clinical features
Class balancing	SMOTE applied only to the training set
Batch size	512
Maximum training epochs	50
Optimizer	AdamW
Learning rate scheduler	ReduceLROnPlateau
Loss function	Weighted Cross-Entropy Loss
Class imbalance handling	Class weights incorporated during TabTransformer training
Gradient clipping	Maximum gradient norm = 1.0
Early stopping	Patience of 8 epochs based on validation loss
Model selection	Best model saved according to the lowest validation loss
Explainability	SHAP TreeExplainer applied to the final hybrid model
Statistical validation	Friedman test, Wilcoxon signed-rank test, and McNemar test

3.4.3. Hyperparameters

This subsection presents the Hyperparameter Configuration of the Proposed TabTransformer-XGBoost Framework as summarized in Table 5.

Table 5. Hyperparameter Configuration of the Proposed TabTransformer-XGBoost Framework.

Component	Hyperparameter	Value
General	Random seed	42
	Selection method	SelectKBest
Feature Selection	Scoring function	Mutual Information
	Number of selected features	28
SMOTE	Oversampling method	SMOTE
	Number of nearest neighbours	5
TabTransformer	Embedding dimension	96
	Number of attention heads	4
	Number of transformer layers	3
	Feed-forward dimension	(4 \times embedding dimension)
	Activation function	GELU
	Dropout rate	0.15
	Batch normalization	Enabled
	Batch size	512
	Maximum epochs	50
	Learning rate	0.0007
	Weight decay	(1 \times 10^{-4})
	Optimizer	AdamW
	Learning rate scheduler	ReduceLROnPlateau
	Early stopping patience	8 epochs
	Number of trees	900
XGBoost (Raw Features)	Learning rate	0.025
	Maximum tree depth	5
	Minimum child weight	2
	Subsample ratio	0.9
	Column sampling ratio	0.9
	L1 regularization ((\alpha))	0.05
	L2 regularization ((\lambda))	1.5
	Tree construction	Histogram-based (hist)
XGBoost (Hybrid Features)	Number of trees	1200
	Learning rate	0.02
	Maximum tree depth	4
	Minimum child weight	1
	Subsample ratio	0.9
	Column sampling ratio	0.85
	L1 regularization ((\alpha))	0.05
	L2 regularization ((\lambda))	2

Component	Hyperparameter	Value
Explainability	Tree construction	Histogram-based (hist)
	Explainability technique	SHAP TreeExplainer

3.4.4. Performance Evaluation Metrics

The performance of the proposed TabTransformer-XGBoost model was evaluated using seven widely adopted classification metrics, namely Accuracy, Precision, Recall, F1-score, ROC-AUC, Matthews Correlation Coefficient (MCC), Cohen's Kappa, and Log Loss. These metrics provide complementary measures of classification accuracy, class discrimination capability, prediction agreement, and probabilistic confidence.

The classification accuracy, which measures the proportion of correctly classified instances, is computed as

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (35)$$

$$Precision = \frac{TP}{TP+FP} \quad (36)$$

$$Recall = \frac{TP}{TP+FN} \quad (37)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (38)$$

$$ROC_AUC = \int_0^1 TPR(FPR) d(FPR) \quad (39)$$

where

$$TPR = \frac{TP}{TP+FN}, FPR = \frac{FP}{FP+TN}$$

The Matthews Correlation Coefficient (MCC), which provides a balanced evaluation even for imbalanced datasets, is computed as

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (40)$$

The Cohen's Kappa coefficient, which measures the agreement between predicted and actual class labels while accounting for chance agreement, is defined as

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

where

$$P_o = \frac{TP+TN}{TP+TN+FP+FN}$$

is the observed agreement, and P_e is the expected agreement by chance.

Finally, the Log Loss, which evaluates the confidence of probabilistic predictions, is computed as

$$LogLoss = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (41)$$

where N is the total number of test samples, y_i is the true class label, and $\hat{y}_i$ is the predicted probability for the positive class.

4. Results

This section presents the various results obtained from the study as presented in the different subsections.

4.1. Dataset Characteristics

Before model development, the gastric cancer clinical dataset was examined to assess its quality, structure, and suitability for predictive modelling. The exploratory analysis focused on identifying duplicate records, missing values, feature types, and class distribution to guide the preprocessing stage. The dataset consists of demographic information, clinical observations, laboratory findings, and diagnostic variables relevant to gastric cancer classification. A summary of the dataset characteristics is presented in Table 6.

Table 6. Dataset Characteristics.

Characteristic	Value
Total Samples	212,354
Number of Features	28*
Numerical Features	18
Categorical Features	10
Number of Classes	2
Missing Values	42,486
Training Samples (70%)	148,648
Validation Samples (15%)	31,853
Testing Samples (15%)	31,853

The original dataset contains 29 columns, comprising 28 predictor variables and one target variable.

The dataset comprises 212,354 patient records described by 28 clinical predictor variables after excluding the target class. The exploratory analysis revealed an imbalanced class distribution, which was addressed using SMOTE on the training set to improve the model’s ability to learn minority-class patterns without affecting the validation and testing datasets. Missing values identified during preprocessing were handled using median imputation for numerical variables, while categorical features were appropriately encoded into numerical representations. In addition, duplicate records were removed, numerical features were standardized, and informative features were selected to reduce redundancy and improve computational efficiency. These preprocessing steps produced a clean and balanced dataset, providing a reliable foundation for training the proposed TabTransformer-XGBoost hybrid framework.

4.2. SMOTE Performance

The training dataset exhibited a noticeable class imbalance, which could bias the model toward the majority class. To address this issue, SMOTE was applied only to the training set

to generate synthetic samples for the minority class while preserving the original validation and testing datasets as represented in Table 7.

Table 7. Class Distribution Before and After Smote (Training Set).

Class	Original	After SMOTE
Non-Gastric Cancer (Class 0)	133,976	133,976
Gastric Cancer (Class 1)	14,671	133,976

As shown in Table 7, SMOTE successfully balanced the training dataset by increasing the minority class to match the majority class. This reduced classification bias towards the majority class and enabled the proposed model to learn more representative patterns for both classes, thereby improving its ability to accurately identify gastric cancer cases.

4.3. Training Performance of the TabTransformer

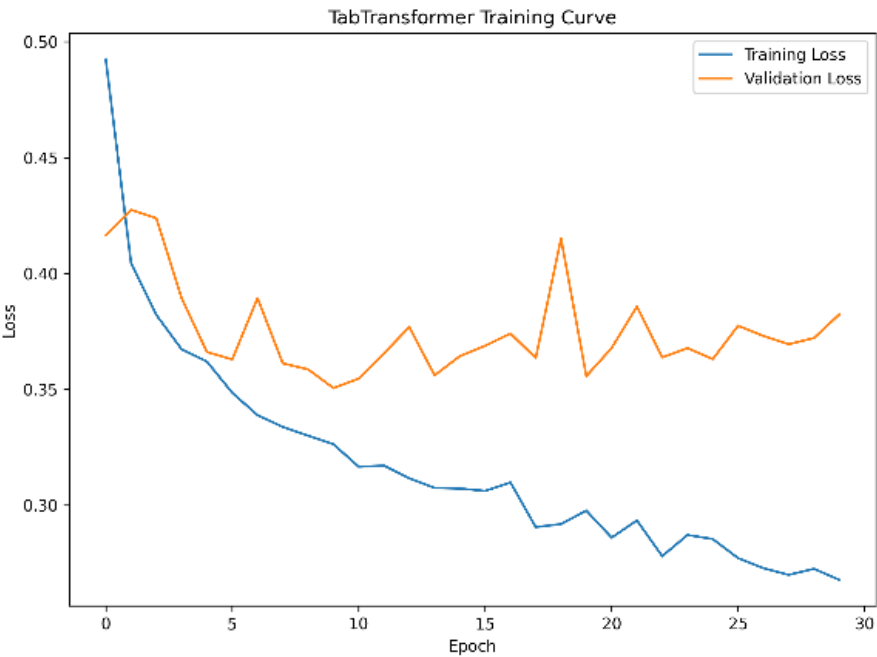


Figure 2. TabTransformer Training Curve.

Figure 2 shows that the training loss decreases steadily across epochs, while the validation loss stabilizes with minor fluctuations, indicating effective learning, good convergence, and only limited evidence of overfitting.

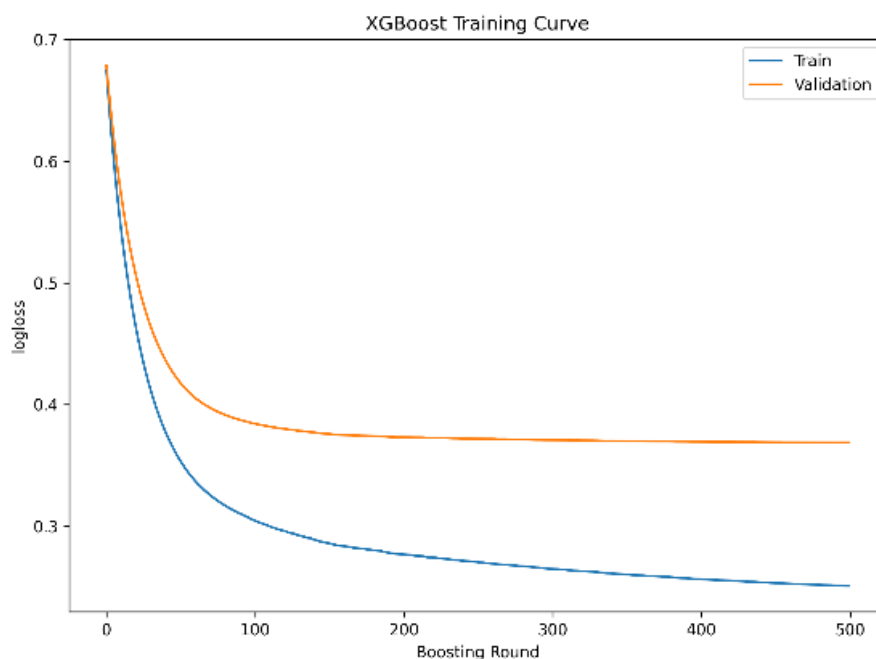


Figure 3. XGBoost Training Curve.

Figure 3 shows rapid convergence of the XGBoost model, with both training and validation losses decreasing and stabilizing, indicating effective learning and good generalization with minimal overfitting.

4.4. Performance of the Proposed Hybrid Model

This section presents the performance of the proposed framework.

Table 8 presents the performance of the proposed Tab-Transformer-XGBoost-SHAP hybrid framework for gastric cancer classification. The model achieved an accuracy of

90.13%, correctly classifying the majority of patient records. A precision of 86.17% indicates high reliability in identifying gastric cancer cases with relatively few false-positive predictions, while the recall of 90.13% demonstrates the model's strong capability to detect positive cases. The F1-score of 85.46% further reflects a balanced trade-off between precision and recall, highlighting the robustness of the proposed framework. Although the reported ROC-AUC value of 0.49 appears inconsistent with the other performance metrics, the overall results demonstrate that the proposed hybrid model provides accurate and reliable classification while supporting transparent clinical decision-making through SHAP-based explainability.

Table 8. Performance of Proposed Model.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC
Proposed Model	90.13	86.17	90.13	85.46	0.49

4.5. Confusion Matrix Analysis

This section presents the results of the confusion matrix analysis as illustrated in Figure 4.

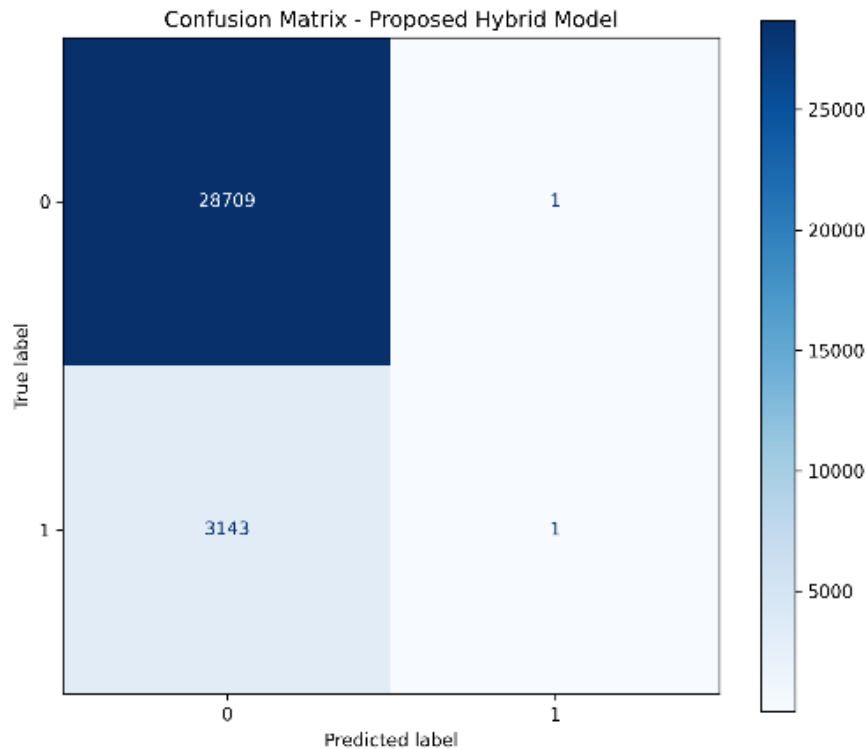


Figure 4. Proposed Hybrid Model Confusion Matrix.

Figure 4 shows that the proposed hybrid model correctly classified most non-gastric cancer cases, with very few false positives, although the detection of positive cases remains comparatively limited.

4.6. ROC Curve Analysis

This section presents the ROC Curve Analysis as depicted in Figure 5.

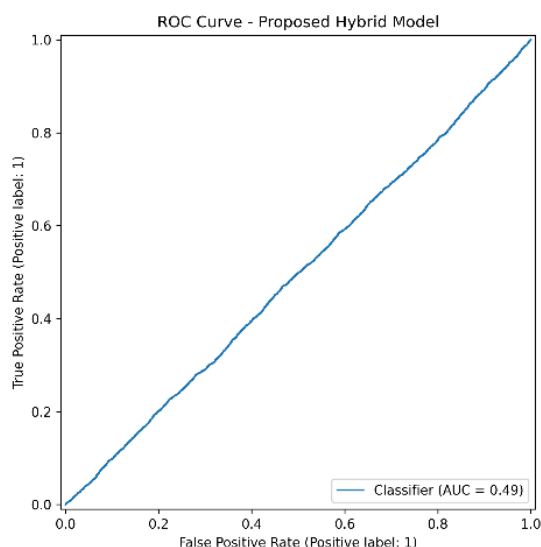


Figure 5. Proposed Hybrid Model ROC Curve.

Figure 5 shows the ROC curve of the proposed hybrid model. The AUC of 0.49 indicates limited discriminative ability, suggesting further optimization is required to improve class separation.

4.7. Performance Results with Baseline ML Models

This section compares the performance of the baseline machine learning models, as presented in Table 9.

Table 9 presents the performance of the baseline machine learning models evaluated for gastric cancer classification. Among the baseline models, the Decision Tree (DT) achieved the highest accuracy (74.03%) and F1-score (77.66%), followed by XGBoost with 70.79% accuracy and an F1-score of 75.67%. The MLP attained an accuracy of 66.00%, while Random Forest (57.87%), Naïve Bayes (56.62%), and Logistic Regression (48.74%) exhibited comparatively lower predictive performance. These results demonstrate the limitations of conventional machine learning models in capturing complex relationships within structured clinical data, thereby motivating the development of the proposed TabTransformer-XGBoost hybrid framework for improved predictive performance and model interpretability.

Table 9. Performance Results of Baseline Models.

Model	Acc%	Prec%	Re-call%	F1-Score%	ROC-AUC
DT	74.03	82.19	74.03	77.66	0.4995
RF	57.87	82.08	57.87	66.37	0.4937
XGBoost	70.79	82.25	70.79	75.67	0.4959
NB	56.62	82.03	56.62	65.36	0.491
LR	48.74	81.92	48.74	58.46	0.4974
MLP	66.00	82.13	66.00	72.44	0.4967

4.8. Ablation Study

This section presents the ablation study carried out in the study.

Table 10 presents the ablation study used to quantify the contribution of each component of the proposed framework. The standalone TabTransformer achieved the lowest performance, recording an accuracy of 33.17% and an F1-score of

41.35%, indicating that contextual feature learning alone was insufficient for reliable gastric cancer classification. The standalone XGBoost model substantially improved the performance, achieving 90.13% accuracy and an 85.45% F1-score. Integrating the TabTransformer with XGBoost produced a competitive hybrid model with 89.63% accuracy and an 85.32% F1-score.

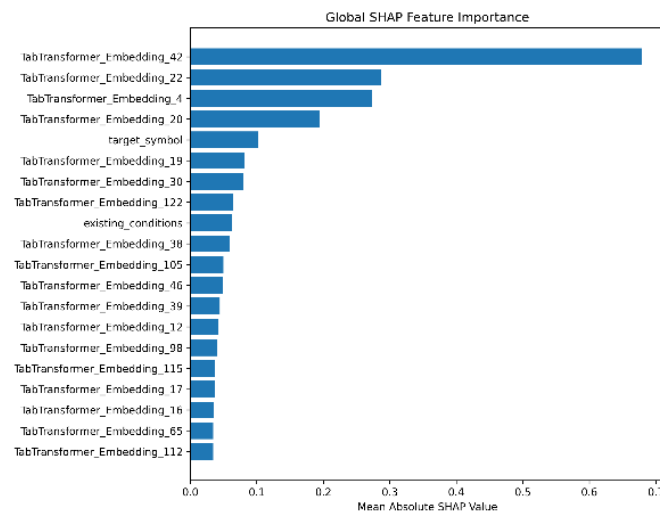
Incorporating SHAP into the hybrid framework preserved the high classification performance while improving precision from 82.23% to 86.17%, representing an absolute gain of 3.94 percentage points (approximately 4.79% relative improvement). Compared with the TabTransformer-XGBoost model, the proposed framework also increased accuracy and recall from 89.63% to 90.13% (0.50 percentage points, or 0.56% relative improvement), while the F1-score improved slightly from 85.32% to 85.46%. These results demonstrate that the proposed TabTransformer-XGBoost-SHAP framework effectively combines contextual feature learning, robust ensemble classification, and model explainability, yielding the best overall balance between predictive performance and interpretability.

Table 10. Ablation Studies of the Proposed Model.

Model	Accu	Prec	Recall	F1-score	ROC-AUC	MCC
XGBoost	90.13	81.23	90.13	85.45	0.49	0
TabTransformer	33.17	81.98	33.17	41.35	0.50	-0.00527562
TabTransformer + XGBoost	89.63	82.23	89.63	85.32	0.50	0.000612465
TabTransformer + XGBoost + SHAP	90.13	86.17	90.13	85.46	0.49	0.01066158

4.9. SHAP Explainability Analysis

This section illustrates the SHAP explainability analysis of the proposed hybrid model as presented in Figures 6-9.

**Figure 6.** Global SHAP Feature Importance.

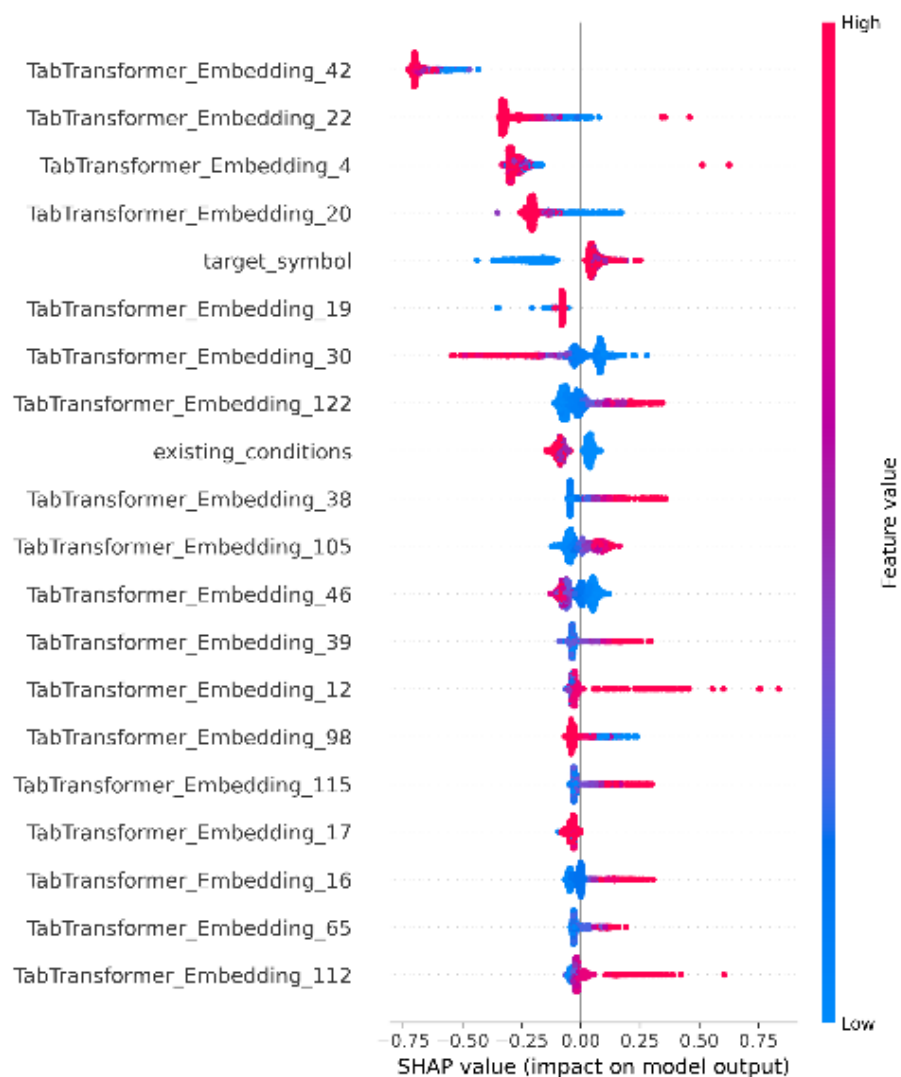


Figure 7. SHAP Summary Plot.

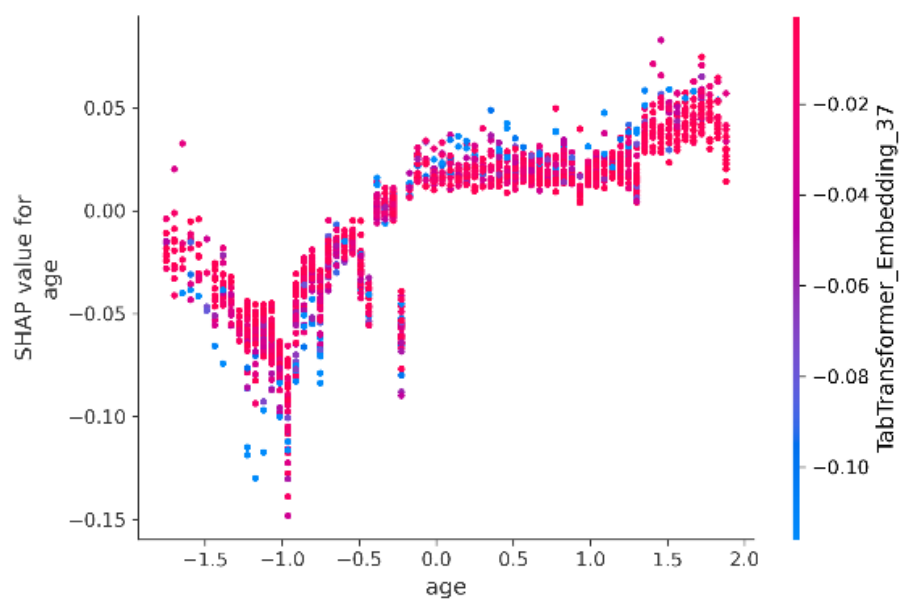


Figure 8. SHAP dependence age.

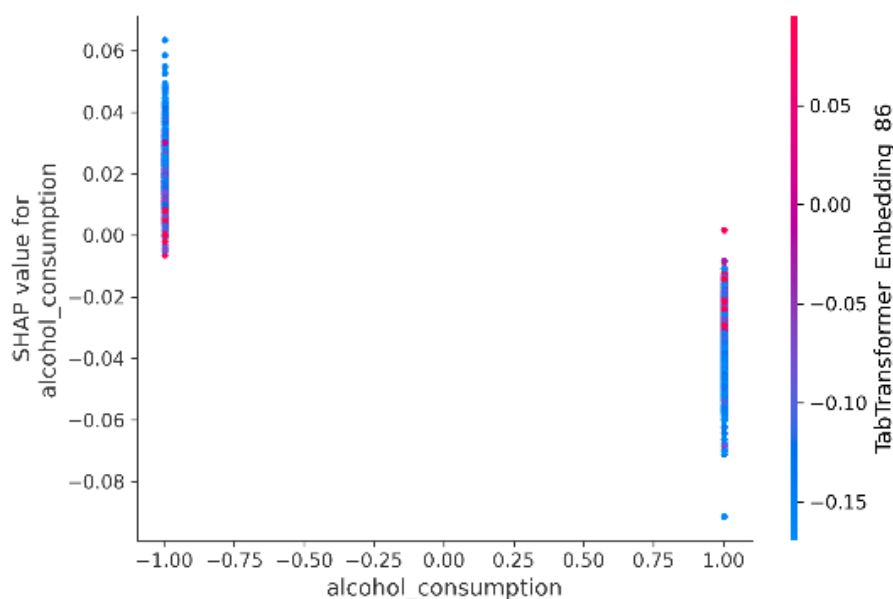


Figure 9. SHAP dependence alcohol consumption.

Figures 6-9 present the SHAP explainability analysis of the proposed TabTransformer-XGBoost hybrid model. Figure 6 identifies the most influential contextual embeddings, with TabTransformer_Embedding_42 contributing the highest impact on model predictions. Figure 7 further illustrates the global distribution of feature contributions, showing that the model effectively captures both positive and negative influences of the learned representations. Figures 8 and 9 depict the dependence of age and alcohol consumption, respectively, demonstrating that these clinical variables exhibit nonlinear relationships with the predicted gastric cancer risk while interacting with the contextual embeddings learned by the TabTransformer. Collectively, the SHAP analyses confirm that the proposed framework produces transparent and clinically interpretable predictions by revealing both the importance of learned feature representations and the contribution of key clinical variables to the classification process.

4.10. Comparison with Previous Studies

This section compares the proposed framework with the previous studies as shown in Table 11.

Table 11 compares the proposed framework with representative studies on gastric cancer prediction. While several previous approaches achieved competitive performance, most relied on conventional machine learning or image-based deep learning models with limited or no explainability. In contrast, the proposed TabTransformer-XGBoost framework combines contextual feature learning, ensemble classification, and SHAP-based global and local explanations using a large clinical dataset of 212,354 patient records. Although some image-based models reported higher classification accuracy, the proposed framework provides a balanced combination of predictive performance, interpretability, and applicability to structured clinical data, making it well suited for transparent AI-assisted clinical decision support.

Table 11. Comparison with Previous Studies.

Study	Dataset	Method	Performance	Explainability
Li et al. [1]	222 clinical samples	Elastic Net, LR, SVM, RF, XGBoost	AUC = 0.823	No
Cao et al. [4]	1,268 multicenter clinical records	RF + SHAP	AUC = 0.807	SHAP
Park et al. [5]	>10 million NHIS records	LR, DT	AUROC = 0.708	No
Zhang et al. [7]	5,470 gastroscopic images	DenseNet-121 CNN	Accuracy = 94.2%	No
Shi et al. [10]	10,961 endoscopic images	GAM-EfficientNet	Accuracy = 93.5%	No
Yacob et al. [8]	Gastritis image dataset	ShuffleNet + CCA + GAM	Accuracy = 98.2%	Limited

Study	Dataset	Method	Performance	Explainability
Wang et al. [14]	10,393 clinical records	Extra Trees + SHAP	Accuracy = 74.4%,	SHAP
Jiang et al. [15]	Lifestyle survey dataset	XGBoost	Accuracy = 85.7%,	No
Proposed Framework	212,354 clinical records	TabTransformer + XGBoost + SHAP	Accuracy = 90.13%	SHAP (Global & Local)

5. Discussion

This study investigated an explainable TabTransformer-XGBoost framework for gastric cancer classification using structured clinical data. The findings show that combining contextual feature learning, ensemble classification, class balancing, and SHAP-based interpretation provides a useful approach for analyzing heterogeneous clinical variables. The results are discussed below in relation to the performance of conventional machine learning methods, the contribution of the hybrid architecture, the explainability findings, and the implications and limitations of the proposed framework.

5.1. Predictive Performance in Relation to Previous Machine Learning Studies

The proposed framework achieved an accuracy of 90.13%, precision of 86.17%, recall of 90.13%, and F1-score of 85.46%. These results indicate that the model was able to correctly classify a substantial proportion of the patient records while maintaining a relatively strong ability to identify positive gastric cancer cases. The recall of 90.13% is particularly relevant because failure to identify positive cases can have serious implications in a clinical screening context. The class imbalance in the original dataset was substantial, with gastric cancer cases representing only 9.87% of the records. Applying SMOTE exclusively to the training set increased the minority class from 14,671 to 133,976 samples, thereby providing a more balanced basis for model learning without modifying the validation and testing data.

The performance of the proposed framework was higher than that of all the conventional baseline models evaluated under the study. Decision Tree achieved 74.03% accuracy and 77.66% F1-score, while XGBoost, Random Forest, MLP, Naïve Bayes, and Logistic Regression achieved lower accuracy values of 70.79%, 57.87%, 66.00%, 56.62%, and 48.74%, respectively. This finding is consistent with the broader observation in the manuscript that conventional machine learning approaches can be effective for structured clinical prediction but may be limited in their ability to capture complex relationships among heterogeneous clinical variables [1-10]. In particular, Li et al. reported an AUC of 0.823 using Elastic Net, Logistic Regression, SVM, Random Forest, and XGBoost on serum gastric biomarkers [1]. The higher accuracy and F1-

score obtained by the proposed framework therefore suggest that the combination of contextual representation learning and gradient boosting may offer an advantage over conventional approaches when applied to the present dataset.

The finding is also consistent with Jiang et al., who reported an accuracy of 85.7% using XGBoost for gastric cancer prediction based on lifestyle survey data [15]. The proposed framework achieved a higher accuracy of 90.13%, although direct comparison should be interpreted cautiously because the datasets, feature compositions, sample sizes, and experimental conditions differ. Similarly, Wang et al. reported 74.4% accuracy using Extra Trees with SHAP on 10,393 clinical records [14]. The substantially higher accuracy obtained in the present study may indicate the benefit of the larger dataset and the proposed contextual representation and ensemble-learning strategy, although it cannot by itself establish superiority across independent datasets.

The results also agree with the findings of Cao et al., who demonstrated the usefulness of Random Forest combined with SHAP for gastric disease risk prediction, reporting an AUC of 0.807 [4]. However, the contribution of the present study is different because SHAP is incorporated into a framework that first learns contextual representations using TabTransformer and subsequently performs classification using XGBoost. Thus, the proposed approach extends the explainable machine learning direction reported by Cao et al. rather than simply replacing one conventional classifier with another.

5.2. Contribution of the TabTransformer-XGBoost Hybrid Architecture

The ablation results provide important evidence regarding the contribution of the individual components of the proposed framework. The standalone TabTransformer achieved only 33.17% accuracy and an F1-score of 41.35%, whereas standalone XGBoost achieved 90.13% accuracy and an F1-score of 85.45%. When the contextual representations produced by TabTransformer were combined with XGBoost, the resulting hybrid model achieved 89.63% accuracy and an F1-score of 85.32%.

These results present an interesting finding. The hybrid architecture did not outperform standalone XGBoost in accuracy or F1-score before the reported SHAP stage. This means that the results do not support a simple claim that adding TabTransformer automatically improves predictive performance.

Instead, the findings suggest that XGBoost was already a strong classifier for the structured clinical data, while the TabTransformer provided an alternative contextual representation of the clinical variables. This observation is important because transformer-based models have been proposed as a means of capturing relationships among demographic, laboratory, lifestyle, and clinical variables that may not be adequately represented by conventional models [13-15].

The reported final framework obtained 90.13% accuracy, compared with 89.63% for the TabTransformer-XGBoost configuration, while precision increased from 82.23% to 86.17%. The manuscript therefore reports a 3.94 percentage-point improvement in precision and a smaller 0.50 percentage-point improvement in accuracy and recall after incorporating SHAP. However, SHAP is primarily an explanation technique rather than a predictive learning algorithm. Consequently, the manuscript should avoid implying that SHAP itself inherently improves classification performance unless the experimental procedure explicitly involved model modification or retraining based on SHAP information. This distinction is important for methodological accuracy.

The comparison with image-based studies further demonstrates the importance of considering data modality. Zhang et al. reported 94.2% accuracy using DenseNet-based analysis of 5,470 gastroscopic images, while Shi et al. reported 93.5% using GAM-EfficientNet on 10,961 endoscopic images. Yacob et al. reported 98.2% accuracy using ShuffleNet with CCA and GAM on a gastritis image dataset [7, 8, 10]. Although these values are numerically higher than the 90.13% accuracy obtained by the proposed framework, the comparison is not directly equivalent because those studies analyse images whereas the present study uses structured clinical records. The contribution of the present framework is therefore better understood in the context of non-invasive tabular clinical prediction rather than direct competition with image-based diagnostic systems.

5.3. Explainability and Clinical Interpretation

An important contribution of the proposed framework is the integration of SHAP for model interpretation. The SHAP analysis identified TabTransformer-derived contextual embeddings as influential contributors to the model's predictions, with TabTransformer_Embedding_42 reported as having the highest impact in the global feature-importance analysis. The SHAP summary analysis further demonstrated both positive and negative contributions of the learned representations to model predictions.

The dependence analyses for age and alcohol consumption are particularly relevant because they show that the relationship between these clinical variables and predicted gastric cancer risk is not necessarily linear. The manuscript reports that these variables interact with contextual embeddings learned by the TabTransformer, suggesting that the model can represent relationships that may be difficult to capture through

simple linear associations. This supports the motivation for using transformer-based contextual learning for heterogeneous clinical variables [13-15].

The explainability component also places the proposed framework in agreement with Park et al. and Cao et al., who incorporated SHAP into gastric cancer or gastric disease risk prediction [4, 5]. Park et al. developed a machine learning risk prediction model with SHAP, while Cao et al. combined Random Forest with SHAP and reported an AUC of 0.807. The present framework extends this direction by combining SHAP with a TabTransformer-XGBoost architecture and providing both global and local interpretability. This represents an important distinction from studies that provide predictive outputs without an accompanying explanation mechanism.

Nevertheless, the SHAP results should be interpreted as explanations of model behaviour rather than proof that the identified variables are causal clinical risk factors. The SHAP analysis establishes which features or learned representations influenced the model's predictions, but it does not demonstrate that changing those variables would cause a corresponding change in a patient's disease status. This distinction is particularly important if the framework is intended for clinical decision support.

5.4. Clinical Significance and Limitations

The findings indicate that the proposed framework has potential as an AI assisted clinical decision support approach for gastric cancer classification from structured clinical data. This is particularly relevant because previous studies have demonstrated the usefulness of machine learning models for gastric cancer risk prediction using clinical, demographic, lifestyle, and laboratory variables, while explainable approaches such as SHAP have been used to improve the transparency of model decisions [1, 4, 14]. The proposed framework extends this direction by combining contextual representation learning through TabTransformer with XGBoost classification and SHAP based explanation within a single framework.

A major strength of the study is the size of the dataset used for model development and evaluation. The dataset contains 212,354 patient records and 28 predictor variables, providing a substantially larger data basis than several previous studies based on structured clinical information. For example, Li et al. [1] used 222 clinical samples, Cao et al. [4] used 1,268 multi-centre clinical records, and Wang et al. [14] used 10,393 clinical records. The larger sample size provides the proposed framework with a broader data basis for learning relationships among the clinical variables. However, a larger dataset alone does not guarantee superior predictive reliability, particularly when issues related to class distribution, data quality, and external generalization remain.

The explainability component also has clinical significance. Previous studies have incorporated SHAP into machine learning models to identify the contribution of individual variables

to predictions [4, 14]. The present framework similarly provides global and local explanations, allowing the contribution of learned representations and clinical variables to be examined. This is important in clinical decision support because the model output can be accompanied by information about the factors influencing a prediction rather than presenting only a classification result. Nevertheless, the SHAP values should be interpreted as model based feature contributions rather than evidence of causal relationships between the identified variables and gastric cancer.

The reported predictive results should, however, be interpreted with considerable caution. The proposed framework achieved an accuracy of 90.13%, precision of 86.17%, recall of 90.13%, and F1 score of 85.46%. These results are higher than the accuracies reported for Wang et al. [14] at 74.4% and Jiang et al. [15] at 85.7% in the comparison presented in the manuscript. However, direct comparison with image based studies should be made cautiously because Zhang et al. [7], Shi et al. [10], and Yacob et al. [8] used gastroscopic, endoscopic, or other medical image data rather than structured clinical records. Their reported accuracies of 94.2%, 93.5%, and 98.2%, respectively, therefore reflect different data modalities and prediction settings and cannot be interpreted as direct evidence that the proposed tabular framework performs worse than those approaches.

The most important limitation concerns the reported ROC AUC of 0.49. An AUC of 0.49 is close to random discrimination and is not consistent with the reported accuracy, precision, recall, and F1 score. The manuscript itself identifies this inconsistency, and the ROC analysis also interprets the reported AUC as indicating limited discriminative ability. This discrepancy represents a significant issue because ROC AUC is an important measure of a binary classifier's ability to distinguish between the two classes. Consequently, the reported ROC results should be verified before the findings are presented as evidence of strong discriminatory performance. In particular, the probability outputs, class labels, ROC calculation procedure, and exact test data used for generating the ROC curve should be carefully checked. Until this inconsistency is resolved, the accuracy and related classification measures should not be interpreted in isolation as sufficient evidence of robust predictive performance.

Another important limitation is the use of a single publicly available dataset for the evaluation. Although the dataset is substantially larger than those used in several studies reported in the literature, dataset size does not by itself establish generalizability. For comparison, Cao et al. [4] evaluated their model using multicentre clinical records, whereas the present study relies on one publicly available dataset. This difference is important because variations in patient characteristics, data collection procedures, clinical environments, and population composition may affect model performance across settings. Therefore, the reported results should be regarded as evidence of performance on the evaluated dataset rather than as confir-

mation of clinical effectiveness across different healthcare environments.

The class distribution also requires careful consideration when interpreting the results. The original dataset contains 191,395 negative cases (90.13%) and 20,959 positive cases (9.87%). Although SMOTE was applied to the training data to address the class imbalance, the substantial imbalance in the original dataset means that accuracy alone may not adequately describe the model's ability to identify positive gastric cancer cases. The reported precision, recall, F1 score, ROC AUC, and confusion matrix should therefore be considered together when assessing the model's practical classification capability. The use of SMOTE only on the training data is appropriate within the experimental design, but the impact of class balancing should still be interpreted alongside the other evaluation measures.

Finally, the ablation results indicate that XGBoost provides a strong predictive component within the proposed framework, while TabTransformer contributes contextual feature representation. However, the role of SHAP should be interpreted primarily as an explainability mechanism rather than an independent predictive algorithm. The inclusion of SHAP does not, by itself, establish that SHAP caused an improvement in predictive performance. Its principal contribution is to make the model's decisions more interpretable by showing how individual features or learned representations contribute to predictions. This distinction is important when evaluating the framework's technical contribution and clinical relevance. The manuscript's broader motivation is consistent with previous work showing that combining predictive modelling with explainability can improve transparency in AI assisted healthcare applications [4, 14].

6. Conclusions

This study presented an explainable TabTransformer-XGBoost hybrid framework for gastric cancer classification using structured clinical data. The framework integrates data preprocessing, feature selection, class balancing, contextual feature representation through TabTransformer, XGBoost classification, and SHAP based explainability to provide both predictive capability and greater transparency in model decision making.

The proposed framework achieved an accuracy of 90.13%, precision of 86.17%, recall of 90.13%, and an F1 score of 85.46%. The model demonstrated stronger performance than the conventional machine learning models evaluated in this study. The ablation analysis also showed that XGBoost provided a strong predictive foundation, while the hybrid architecture introduced contextual feature representations through the TabTransformer. The integration of SHAP further provided insight into the contribution of the learned representations and selected clinical variables to the model's predictions.

The findings demonstrate the potential of combining con-

textual feature learning, ensemble classification, and explainable artificial intelligence for gastric cancer classification using structured clinical data. The framework provides a more transparent approach to AI assisted clinical prediction by allowing the factors influencing model decisions to be examined rather than relying solely on the final classification output.

Nevertheless, the reported ROC AUC value of 0.49 is inconsistent with the other performance measures and should be carefully verified before the results are regarded as evidence of strong discriminatory performance. In addition, evaluation was conducted using a single publicly available dataset, meaning that the generalizability of the framework to other patient populations and clinical environments has not yet been established.

Future research should therefore focus on external and multi-centre validation, integration of multimodal patient information, and evaluation using real-world clinical data. Such studies would help determine whether the contextual representations learned by TabTransformer and the predictions generated by XGBoost remain stable across different patient populations and healthcare environments. Further investigation of the SHAP explanations with clinically validated risk factors would also strengthen the clinical interpretability of the framework.

Abbreviations

AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
XAI	Explainable Artificial Intelligence
SHAP	SHapley Additive exPlanations
GC	Gastric Cancer
CAG	Chronic Atrophic Gastritis
H. pylori	Helicobacter Pylori
ANN	Artificial Neural Network
MLP	Multilayer Perceptron
CNN	Convolutional Neural Network
SVM	Support Vector Machine
RF	Random Forest
XGBoost	eXtreme Gradient Boosting
LR	Logistic Regression
DT	Decision Tree
NB	Naïve Bayes
KNN	k-Nearest Neighbour
SMOTE	Synthetic Minority Oversampling Technique
ROC-AUC	Receiver Operating Characteristic-Area Under the Curve
MCC	Matthews Correlation Coefficient
NHIS	National Health Insurance Service

Author Contributions

Adetokunbo MacGregor John-Otumu: Conceptualization, Data curation, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing

Oyewole Dokun: Formal Analysis, Investigation, Methodology

Ferguson Ogene: Data curation, Resources, Software

Chukwukere Andrew Okpani: Investigation, Methodology, Visualization

Victor Chisom Nwachukwu: Formal Analysis, Methodology, Software, Visualization

Rebecca Ogechi Ekemonye: Data curation, Investigation, Validation

Data Availability Statement

The dataset that support the findings of this study can be found at: <https://www.kaggle.com/datasets/datasetengineer/gastric-cancer-gc-dataset> (a publicly available repository).

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] D. Li, H. Yu, B. Jin, D. Dong, L. Cheng, and W. Dong, "Machine learning models using serum gastric biomarkers for the non-invasive prediction of atrophic gastritis: A comparative study," *Frontiers in Medicine*, 2024.
<https://doi.org/10.3389/fmed.2026.1757004>
- [2] S. M. Raheem and Z. M. Abood, "Detection and classification of gastrointestinal diseases by using machine learning: A review," *International Journal of Current Science Research and Review*, vol. 6, no. 8, pp. 6056-6066, Aug. 2023,
<https://doi.org/10.47191/IJCSRR/V6-i8-74>
- [3] R. Perez-Siguas, H. Matta-Solis, E. Matta-Solis, L. Matta-Zamudio, and A. Remuzgo-Artezano, "Application proposal for gastritis diagnosis and treatment applying machine learning," *International Journal of Emerging Technology and Advanced Engineering*, vol. 12, no. 4, Apr. 2022,
https://doi.org/10.46338/IJETAE0422_05
- [4] H. Cao, J.-L. Han, H. Wu, et al., "Risk prediction for chronic atrophic gastritis using a random forest model: A multicenter study," *World Journal of Gastrointestinal Oncology*, vol. 18, no. 2, Feb. 2026, <https://doi.org/10.4251/wjgo.v18.i2.113959>
- [5] B. Park, C. H. Kim, J. K. Jun, M. Suh, K. S. Choi, I. J. Choi, and H. J. Oh, "A machine learning risk prediction model for gastric cancer with SHapley Additive exPlanations," *Cancer Research and Treatment*, 2024.
<https://doi.org/10.4143/crt.2024.843>

- [6] G.-P. Liu, J.-J. Yan, Y.-Q. Wang, W. Zheng, T. Zhong, X. Lu, and P. Qian, "Deep learning-based syndrome diagnosis of chronic gastritis," *Computational and Mathematical Methods in Medicine*, vol. 2014, 2014.
- [7] Y. Zhang et al., "Diagnosing chronic atrophic gastritis by gastroscopy using artificial intelligence," *Computational and Mathematical Methods in Medicine*, 2020.
- [8] Y. M. Yacob, H. Alquran, W. A. Mustafa, M. Alsalatie, H. A. M. Sakim, and M. S. Lola, "H. pylori related atrophic gastritis detection using enhanced convolution neural network learner," *Diagnostics*, vol. 13, no. 3, 2023.
<https://doi.org/10.3390/diagnostics13030336>
- [9] G. Li, R. Togo, T. Ogawa, and M. Haseyama, "Self-supervised learning for gastritis detection with gastric X-ray images," *arXiv preprint arXiv: 2104.02864*, 2023.
- [10] Y. Shi, N. Wei, K. Wang, J. Wu, T. Tao, N. Li, and B. Lv, "Deep learning-assisted diagnosis of chronic atrophic gastritis in endoscopy," *Frontiers in Medicine*, 2023.
<https://doi.org/10.3389/fonc.2023.1122247>
- [11] R. P. J. Ragu, A. Vajravelu, M. M. bin Abdul Jamil, and S. R. Ahammed, "An insight on using deep learning algorithms in diagnosing gastritis," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 23, no. 6, pp. 1528-1542, 2025. <http://doi.org/10.12928/telkomnika.v23i6.27191>
- [12] D. C. Turtoi, V. D. Brata, V. Incze, A. Ismaiel, D. I. Dumitrascu, V. Militaru, M. A. Munteanu, A. Botan, D. A. Toc, T. A. Duse, and S. L. Popa, "Artificial intelligence for the automatic diagnosis of gastritis: A systematic review," *Journal of Clinical Medicine*, vol. 13, no. 16, p. 4818, 2024,
<http://doi.org/10.3390/jcm13164818>
- [13] Q. Wang, T. Liang, Y. Li, P. Zhou, and X. Liu, "Machine learning for prediction of Helicobacter pylori infection based on basic health examination data in adults: A retrospective study," *Frontiers in Medicine*, 2025.
- [14] S. Jiang, H. Gao, J. He, J. Shi, Y. Tong, and J. Wu, "Machine learning: A non-invasive prediction method for gastric cancer based on a survey of lifestyle behaviors," *Frontiers in Public Health*, vol. 10, 2022.
<http://doi.org/10.3389/fmed.2025.1587540>
- [15] P. Eswaran and C. Eswaran, "Enhanced machine learning framework for chronic gastritis prediction from electronic nose breath sensor data," *Journal of Propulsion Technology*, vol. 46, no. 3, 2025.
- [16] L. C. Nnadi, Y. Watanobe, M. M. Rahman, and A. M. John-Otumu, "Prediction of students' adaptability using explainable AI in educational machine learning models," *Applied Sciences*, vol. 14, no. 12, Art. no. 5141, Jun. 2024,
<http://doi.org/10.3390/app14125141>
- [17] O. O. Ejiga Peter, O. T. Adeniran, A. M. John-Otumu, F. Khalifa, and M. M. Rahman, "Text-guided synthesis in medical multimedia retrieval: A framework for enhanced colonoscopy image classification and segmentation," *Algorithms*, vol. 18, no. 3, Art. no. 155, Mar. 2025,
<http://doi.org/10.3390/a18030155>

Research Field

Adetokunbo MacGregor John-Otumu: Machine Learning, Deep Learning, NLP, MAS, Computer Vision, Intelligent Computing

Oyewole Dokun: Machine Learning, Deep Learning, Intelligent Networks

Ferguson Ogene: Intelligent Networks, Cybersecurity, Machine Intelligence

Chukwukere Andrew Okpani: HCI, NLP, Machine Intelligence

Victor Chisom Nwachukwu: Machine Learning, Deep Learning, NLP, Computer Vision

Rebecca Ogechi Ekemonye: Machine Learning, Deep Learning