The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 4, No 4, April 2026
PP 52–60

Machine Learning-Based Gastric Cancer Risk Prediction Using Demographic, Lifestyle, and Symptom Features

A.Nagabhushana Rao[1], Dushmanta Pandhy[2], R. Bhagya Lakshmi[3]

[1,2,3]Department of MCA, NSRIT, Visakhapatnam, Andhra Pradesh, India

Corresponding Author *: routhubhagyalakshmi4@gmail.com

Abstract

Gastric cancer is one of the major health concerns worldwide, and early prediction can improve treatment outcomes and reduce mortality. This study presents a machine learning-based approach for predicting gastric cancer risk using demographic information, lifestyle factors, and clinical symptoms. The collected dataset is preprocessed through data cleaning, feature encoding, and feature selection to improve data quality. Three machine learning algorithms, namely Logistic Regression, Random Forest, and Support Vector Machine (SVM), are implemented and evaluated using standard performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The best-performing model is integrated into a Flask-based web application that allows users to enter patient details and obtain gastric cancer risk prediction. The developed system provides a simple and efficient approach for identifying individuals at potential risk, supporting early medical intervention and assisting healthcare professionals in clinical decision-making.

Keywords: Gastric Cancer Risk Prediction, Machine Learning, Logistic Regression, Random Forest, Support Vector Machine (SVM), Flask.

1. Introduction

Gastric cancer is one of the most common and life-threatening cancers worldwide. Early prediction of the disease is important because timely diagnosis can improve treatment outcomes and increase the survival rate of patients. However, the symptoms of gastric cancer are often similar to those of common gastric disorders, making early diagnosis difficult. Conventional diagnostic methods are effective but can be invasive, expensive, and time-consuming.

Machine learning has become an important tool in healthcare for predicting diseases using patient data. By analyzing demographic information, lifestyle factors, and clinical symptoms, machine learning models can identify patterns associated with gastric cancer risk. In this study, Logistic Regression, Random Forest, and Support Vector Machine (SVM) are used to develop

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 4, No 4, April 2026
PP 52–60

a gastric cancer risk prediction system. The performance of these models is evaluated, and the best-performing model is integrated into a Flask-based web application to provide an efficient and user-friendly prediction system.

2. Literature Survey

Recent studies have shown that machine learning plays an important role in the early prediction and diagnosis of gastric cancer. Park et al. [1] developed a machine learning-based risk prediction model using SHAP to improve both prediction accuracy and model interpretability. Du et al. [2] proposed explainable machine learning models for early gastric cancer diagnosis, while Fan et al. [3] reviewed existing prediction models and highlighted the importance of selecting appropriate risk factors and machine learning algorithms.

Several researchers have compared different machine learning techniques for gastric cancer prediction. Briggs et al. [4] reported that machine learning models provide better prediction performance than conventional risk assessment methods. Gu et al. [5] and Xu et al. [6] presented systematic and scoping reviews, emphasizing the need for reliable and clinically applicable prediction models. These studies demonstrate the growing importance of machine learning in supporting early detection of gastric cancer.

Recent advancements in artificial intelligence have further improved gastric cancer diagnosis and prognosis. Horiuchi et al. [8] discussed the application of artificial intelligence in the early detection of gastric cancer, while Niu et al. [9] reviewed the current and future applications of AI in gastric cancer management. Wang et al. [10] also demonstrated the effectiveness of machine learning techniques in gastric cancer prediction. Based on these studies, the proposed work develops a gastric cancer risk prediction system using Logistic Regression, Random Forest, and Support Vector Machine (SVM) to provide an accurate and user-friendly prediction system.

3. Methodology

The proposed methodology follows a systematic approach for predicting gastric cancer risk using machine learning techniques. Initially, the dataset is collected and preprocessed through data cleaning, handling missing values, encoding categorical features, and data normalization to improve data quality. Relevant features are then selected using correlation analysis and feature importance techniques. Three machine learning algorithms, namely Logistic Regression, Random Forest, and Support Vector Machine (SVM), are trained and evaluated using standard performance metrics such as accuracy, precision, recall, F1-score, and ROC-

AUC. The best-performing model is selected and integrated into a Flask-based web application, where users can enter patient information and receive gastric cancer risk predictions. This methodology provides an accurate, efficient, and user-friendly solution for the early prediction of gastric cancer.

4. Experimental Setup and Implementation

The proposed gastric cancer prediction system was implemented using Python and machine learning techniques to predict gastric cancer risk based on patient information. The implementation process included dataset preparation, data preprocessing, feature selection, model development, model evaluation, and deployment through a Flask-based web application. The complete implementation process is described below.

Dataset Collection: The dataset used in this study contains demographic information, lifestyle factors, and clinical symptoms related to gastric cancer. The collected data was organized in a structured format and prepared for further analysis and model development.

Data Preprocessing: Data preprocessing was performed to improve the quality of the dataset. This process included data cleaning, handling missing values, encoding categorical variables, and feature scaling. These steps helped in preparing the dataset for machine learning model training.

Feature Selection: Relevant features were selected using correlation analysis and feature importance techniques. Feature selection reduced unnecessary attributes and improved the efficiency and performance of the prediction models.

Model Development: Three machine learning algorithms, namely Logistic Regression, Random Forest, and Support Vector Machine (SVM), were used to develop the gastric cancer prediction system. Each model was trained using the preprocessed dataset to identify patterns associated with gastric cancer risk.

Model Evaluation: The developed models were evaluated using standard performance metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. Based on the evaluation results, the best-performing model was selected for deployment.

Flask Web Application: The selected machine learning model was integrated into a Flask-based web application. The application allows users to enter demographic information, lifestyle factors, and clinical symptoms to obtain gastric cancer risk prediction. The prediction result is displayed as Low Risk or High Risk, along with the prediction confidence.

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 4, No 4, April 2026
PP 52–60

5. Result Analysis

The experimental results demonstrate that the proposed machine learning models effectively predict gastric cancer risk using patient information. The performance of Logistic Regression, Random Forest, and Support Vector Machine (SVM) was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC. The results indicate that each model performed well, while the selected model achieved better overall prediction performance. The confusion matrix and ROC curve further confirmed the reliability of the proposed system. In addition, the Flask-based web application successfully generated gastric cancer risk predictions based on the user input, providing a simple and efficient interface for prediction.

Table 1. Performance Comparison of Machine Learning Models

Model	Accuracy (%)	Precision	Recall	F1-Score	AUC
Logistic Regression	86.51%	0.92	0.87	0.88	0.9492
Random Forest	90.19%	0.91	0.90	0.87	0.9891
SVM	94.25%	0.94	0.94	0.94	0.9703

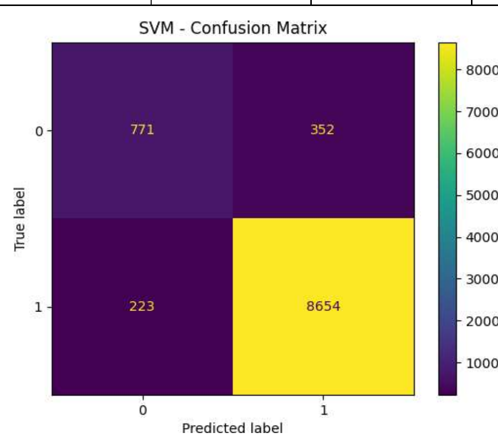

Figure 1. Confusion Matrix of the Selected SVM Model

Table 1 presents the performance comparison of Logistic Regression, Random Forest, and Support Vector Machine (SVM). The results show that all three models performed effectively in predicting gastric cancer risk. Although Random Forest achieved a higher ROC-AUC score, SVM demonstrated better overall performance across accuracy, precision, recall, and F1-score. Based on the overall evaluation, SVM was selected as the final model for deployment in the

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 4, No 4, April 2026
PP 52–60

Flask-based web application.

Figure 1 shows the confusion matrix of the selected SVM model. The matrix indicates that the model correctly classified most samples, with a high number of true positive and true negative predictions. The relatively low number of misclassified cases demonstrates the effectiveness of the proposed model for gastric cancer risk prediction.

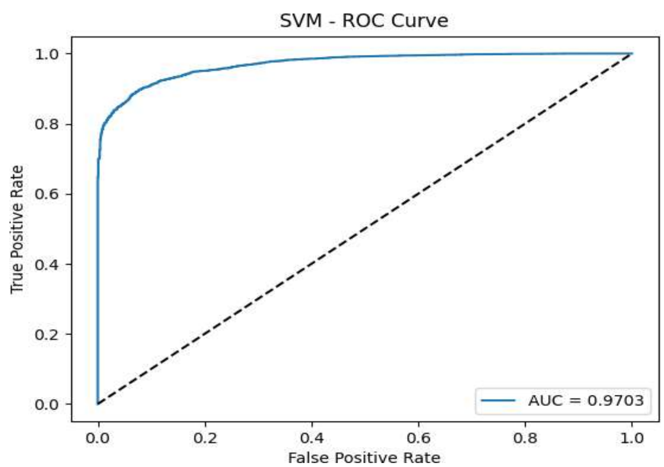

Figure 2. ROC-AUC Curve of the Selected SVM Model

Figure 2 shows the ROC curve of the selected SVM model. The high ROC-AUC value indicates that the model effectively distinguishes between low-risk and high-risk cases. This demonstrates the reliability of the proposed model for gastric cancer risk prediction.

Conclusion

This study presented a machine learning-based gastric cancer risk prediction system using Logistic Regression, Random Forest, and Support Vector Machine (SVM). The developed models were evaluated using standard performance metrics, and SVM demonstrated the best overall performance for predicting gastric cancer risk. The selected model was successfully integrated into a Flask-based web application, providing a simple and user-friendly interface for risk prediction. The proposed system can support the early identification of individuals at risk and assist healthcare professionals in making timely decisions. In the future, the system can be enhanced by using larger real-world clinical datasets and advanced machine learning techniques to further improve prediction performance and reliability.

References

1. Park, B., Kim, C. H., Jun, J. K., Suh, M., Choi, K. S., Choi, I. J., & Oh, H. J. (2025). A machine learning risk prediction model for gastric cancer with SHapley additive explanations. Cancer Research and Treatment, 57(3), 821–829. https://doi.org/10.4143/crt.2024.843

2. Du, H., Yang, Q., Ge, A., Zhao, C., Ma, Y., & Wang, S. (2024). Explainable machine learning models for early gastric cancer diagnosis. Scientific Reports, 14, 17457. https://doi.org/10.1038/s41598-024-67892-z

3. Fan, Z., He, Z., Miao, W., & Huang, R. (2023). Critical analysis of risk factors and machine-learning-based gastric cancer risk prediction models: A systematic review. Processes, 11(8), 2324. https://doi.org/10.3390/pr11082324

4. Briggs, E., de Kamps, M., Hamilton, W., Johnson, O., McInerney, C. D., & Neal, R. D. (2022). Machine learning for risk prediction of oesophago-gastric cancer in primary care: Comparison with existing risk-assessment tools. Cancers, 14(20), 5023. https://doi.org/10.3390/cancers14205023

5. Gu, J., He, F., Clifford, G. M., Li, M., Fan, Z., Li, X., et al. (2022). Prediction models for gastric cancer risk in the general population: A systematic review. Cancer Prevention Research, 15(5), 309–322. https://doi.org/10.1158/1940-6207.CAPR-21-0426

6. Xu, L., Lyu, J., Zheng, X., & Wang, A. (2024). Risk prediction models for gastric cancer: A scoping review. Journal of Multidisciplinary Healthcare, 17, 4337–4352. https://doi.org/10.2147/JMDH.S479699

7. Kather, J. N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C. A., Gaiser, T., Marx, A., & Halama, N. (2019). Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nature Medicine, 25(7), 1054–1056. https://doi.org/10.1038/s41591-019-0462-y

8. Horiuchi, Y., Aoyama, K., Tokai, Y., & Fujisaki, J. (2024). Application of artificial intelligence for diagnosis of early gastric cancer. Clinical Endoscopy, 57(2), 163–172. https://doi.org/10.5946/ce.2023.173

9. Niu, P. H., Zhao, L. L., Wu, H. L., Zhao, D. B., & Chen, Y. T. (2020). Artificial intelligence in gastric cancer: Application and future perspectives. World Journal of Gastroenterology, 26(36), 5408–5419. https://doi.org/10.3748/wjg.v26.i36.5408

10. Wang, H. N., An, J. H., Wang, F. Q., Hu, W. Q., & Zong, L. (2025). Predicting gastric cancer survival using machine learning. World Journal of Gastrointestinal Oncology, 17(5). https://pmc.ncbi.nlm.nih.gov/articles/PMC12142261

11. Zhou, H., Gu, Q., Bao, R., Qiu, L., Zhang, Y., Wang, F., … Yang, Q. (2025). Machine learning based models for predicting presentation delay risk among gastric cancer patients. Frontiers in Oncology, 14, 1503047. https://doi.org/10.3389/fonc.2024.1503047

12. Demirhan, D. B., Yılmaz, H., Erol, H., Kayili, H. M., & Salih, B. (2023). Prediction of gastric cancer by machine learning integrated with mass spectrometry-based N-glycomics. Analyst, 148(9), 2073–2080. https://doi.org/10.1039/D2AN02057B

13. Liatsou, E., Driva, T. S., Vergadis, C., Sakellariou, S., Lykoudis, P., Apostolou, K. G., Tsapralis, D., & Schizas, D. (2025). Current role of artificial intelligence in the management of gastric cancer. Biomedicines, 13(12), 2939. https://doi.org/10.3390/biomedicines13122939

14. 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. Annals of Medicine, 57(1), 2461679.

https://doi.org/10.1080/07853890.2025.2461679

15. Choi, S., & Kim, S. (2023). Artificial intelligence in the pathology of gastric cancer. Journal of Gastric Cancer, 23(3), 410–427.