

AI-Enhanced Non-Invasive Diagnostic System for Early Gastric Ulcer Detection using Salivary pH Monitoring

Srinath A.V

Assistant Professor

Department of Biomedical Engineering
Sri Manakula Vinayagar Engineering College Puducherry.

Syed Feroz Ahamed N

Department of Biomedical Engineering

Sri Manakula Vinayagar Engineering College Puducherry.

Avinash P

Department of Biomedical Engineering

Sri Manakula Vinayagar Engineering College Puducherry.

Mahesh Kumar A

Department of Biomedical Engineering

Sri Manakula Vinayagar Engineering College Puducherry.

Abstract - Gastric ulcer disease is a prevalent gastrointestinal condition that can lead to serious and potentially life-threatening complications if not detected at an early stage. Conventional diagnostic methods, such as endoscopy, are invasive, costly, and impractical for routine or continuous screening. This study proposes an artificial intelligence (AI)-enabled, non-invasive system for early gastric ulcer risk prediction based on salivary pH monitoring. Salivary pH levels are acquired using an analog pH sensor interfaced with an ESP32 microcontroller, allowing real-time and wireless data collection. Machine learning and deep learning approaches, including Random Forest and neural network models, are employed to analyse salivary pH measurements in combination with demographic and lifestyle parameters to classify ulcer risk levels. Experimental results demonstrate a classification accuracy exceeding 90%, indicating the potential of salivary pH as a viable predictive biomarker for gastric ulcer risk. The proposed system provides a low-cost, portable, and patient-friendly solution for early screening, making it particularly suitable for home-based monitoring and healthcare settings with limited resources.

Keywords: Gastric Ulcer, Salivary pH, Non-Invasive Diagnosis, Artificial Intelligence, Machine Learning, ESP32

I. INTRODUCTION

Gastric ulcer is a common digestive disorder and may result in serious complications if it is not identified at an early stage. Clinical diagnosis is most commonly performed using endoscopy, which provides reliable results but involves invasive procedures and higher costs. Due to these limitations, endoscopy is not ideal for regular screening or early-stage assessment, especially in large populations. Recent studies have highlighted the need for alternative diagnostic methods that are simple, patient-friendly, and suitable for preliminary evaluation of gastric conditions [4], [12].

Saliva has gained attention as a non-invasive diagnostic medium because it can be collected easily and reflects physiological changes related to gastrointestinal function. Previous research has

reported that salivary characteristics, including pH variation, are associated with gastric acid imbalance and digestive disorders [13], [14]. In parallel, artificial intelligence techniques have been increasingly applied in gastrointestinal disease analysis to support automated detection and decision-making [1], [6]. By integrating salivary pH monitoring with AI-based analysis, this study aims to develop a non-invasive system for early gastric ulcer detection, with the objective of assisting preliminary risk assessment in a practical and patient-friendly manner.

II. LITERATURE SURVEY

Recent studies have extensively explored the use of artificial intelligence for automated analysis of gastrointestinal diseases, particularly through medical imaging and clinical data interpretation. In the context of gastric pathology, deep learning techniques have been applied to identify ulcerative and precancerous conditions with improved diagnostic consistency. Varadhaganapathy et al. demonstrated the feasibility of detecting gastrointestinal ulcers using convolutional neural networks trained on endoscopic images, highlighting the role of image-based learning in supporting clinical diagnosis [4].

Similarly, Ahmad et al. employed an attention-based YOLOv7 framework for gastric lesion detection, reporting enhanced localization accuracy in complex endoscopic scenes [10]. Advancements in model robustness and feature learning have further strengthened AI-assisted gastroscopy analysis. Kim et al. introduced an adversarial augmentation strategy based on Poisson-Gaussian noise to improve the resilience of deep learning models against real-world image distortions encountered during gastroscopic examinations [2].

Tran et al. proposed a distance-transform-based detection framework capable of identifying gastrointestinal tract lesions in a single-stage process, reducing dependency on multi-step image processing pipelines [6]. In addition, Jhang et al. developed a gastric section correlation network that captured spatial relationships across

different stomach regions, enabling more reliable detection of precancerous lesions [7]. These approaches collectively demonstrate the effectiveness of deep neural networks in minimizing inter-observer variability and improving diagnostic reliability.

Beyond conventional endoscopic image analysis, research has also extended toward multi-organ pathology interpretation and report-level automation. Bui et al. presented a bi-graph interaction network that modelled both spatially constrained and unconstrained relationships in pathology images, offering improved classification across multiple organ systems, including gastric tissues [3]. Complementing image-based methods, Wang et al. applied a three-branch BERT-based architecture to analyse gastroscopy diagnosis text, enhancing consistency in clinical reporting and reducing interpretation errors [5].

Such text-based frameworks highlight the expanding role of natural language processing in gastrointestinal healthcare analytics. In parallel, AI-driven clinical decision support systems have been developed to assist physicians in gastrointestinal disease management. Zhang et al. reviewed machine learning-based decision support platforms and emphasized their contribution to early diagnosis and treatment planning in digestive disorders [12]. Prunella et al. further demonstrated the value of interpretable AI by identifying prognostic biomarkers associated with angiogenesis and immune profiles in gastric and colon cancers, reinforcing the importance of explainable models in clinical adoption [1].

Foundational concepts supporting these developments are grounded in established deep learning principles outlined by Goodfellow et al. [15]. More recently, attention has shifted toward non-invasive and point-of-care diagnostic strategies aimed at improving accessibility and long-term monitoring. Ravenscroft and Occhipinti introduced a predictive point-of-care platform that utilized minimally invasive biomarkers for early disease detection, underscoring the potential of portable diagnostic systems [8]. Saliva, in particular, has been widely recognized as a valuable diagnostic fluid due to its safety, ease of collection, and physiological relevance.

Humphrey and Wong highlighted the diagnostic applicability of salivary parameters in systemic disease assessment [14], while Kaufmann et al. reviewed non-invasive biomarkers linked to gastrointestinal disorders and emphasized their clinical relevance [13]. Supporting infrastructure for such approaches has been enabled by IoT-based healthcare monitoring systems, which facilitate real-time data acquisition and remote analysis [11]. Overall, existing literature demonstrates significant progress in AI-assisted gastric ulcer detection, primarily through image-based and text-based diagnostic frameworks. However, most current methods remain dependent on invasive clinical procedures or hospital-based imaging systems.

The integration of non-invasive biomarkers such as salivary pH with intelligent data analysis remains relatively underexplored. This research gap motivates the present study, which aims to combine salivary pH monitoring with AI-based analysis to support early gastric ulcer risk assessment in a cost-effective and patient-friendly manner.

III. PROPOSED SYSTEM

The proposed system presents a low-cost and non-invasive diagnostic platform designed for early gastric ulcer risk assessment through salivary pH monitoring. The system employs a glass-electrode pH sensor (ELC1057) interfaced with an ESP32 microcontroller to acquire real-time salivary pH measurements before and after food intake. The analog signal obtained from the sensor is appropriately conditioned and digitized using the onboard analog-to-digital converter of the microcontroller. The processed pH data are then integrated with dietary patterns and lifestyle-related parameters to form a comprehensive input dataset. Machine learning-based classification models analyse these multimodal inputs to determine gastric ulcer risk levels, enabling portable, accurate, and home-based screening as illustrated in Fig 1.1.

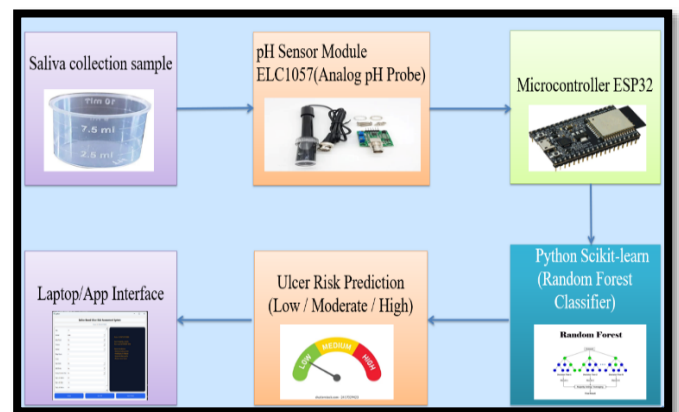


Fig 1.1 System architecture

The Fig1.2a demonstrates that a large proportion of participants consume spicy or fried foods either regularly or on an occasional basis. Such dietary habits are commonly associated with increased gastric acid secretion and irritation of the gastric mucosa. Similar observations on the influence of dietary patterns in gastrointestinal disorders have been reported in earlier studies [4], [13].

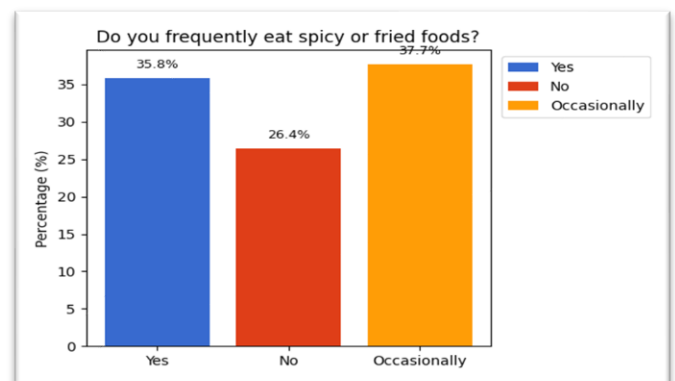


Fig 1.2a Datasets Classification

The distribution of reported digestive symptoms indicates that while most individuals do not experience discomfort, a notable group reports acid reflux, burning sensation, or nausea. Such symptoms are commonly considered indicators of early gastric

disturbances and are often used as supportive clinical inputs as mentioned in Fig 1.2b. Earlier studies have emphasized the usefulness of symptom-related information in gastrointestinal disease assessment and decision support systems [9], [12].

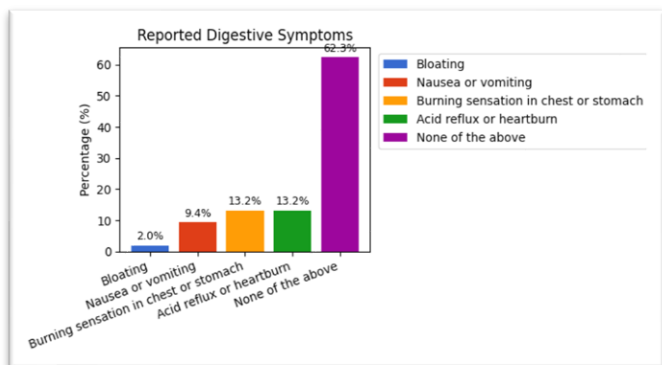


Fig 1.2b Datasets Classification

Serial No	Feature Name	Description
1	Patient ID	Unique identifier for each subject
2	Age	Age of the individual (years)
3	Gender	Male / Female
4	Saliva pH	Measured salivary pH value
5	Spicy Food Intake	Yes / No
6	Tobacco Use	Yes / No
7	Alcohol Consumption	Yes / No
8	Sleep Hours	Average sleep duration per day
9	Stress Level	Low / Moderate / High
10	Skip Meals	Yes / No
11	Soft Drinks Consumption	Yes / No
12	Empty Stomach Pain	Yes / No
13	Ulcer Risk Label	0 = No Ulcer, 1 = Ulcer Risk

i) Datasets analysis

Real-time data were collected directly from patients, capturing their daily lifestyle habits, food intake, and health symptoms. Salivary pH levels were measured non-invasively and analysed alongside these factors, helping to assess individual risk levels and identify early signs of gastric ulcer susceptibility in a practical, patient-friendly manner as labelled in Table 1.1

Table 1.1 Datasets from Patients

ii) Overall System Design

The proposed system follows a layered, modular architecture that integrates non-invasive biosensing, embedded processing, wireless communication, and artificial intelligence to support early gastric ulcer risk prediction. The overall framework is organized into three primary functional layers: the Data Acquisition Layer, the Processing and Communication Layer, and the Intelligence Layer, as depicted in Fig. 1.1. This layered structure ensures scalability, reliability, and flexibility, while also allowing seamless integration of future enhancements such as additional biomarkers or cloud-enabled healthcare platforms.

iii) Data Acquisition Layer

The Data Acquisition Layer is responsible for non-invasive physiological signal collection, specifically salivary pH measurement. Saliva samples of approximately 2–3 ml are collected from subjects under standardized conditions to minimize variability due to external factors such as food intake, hydration level. Sample collection is preferably performed before meals and after a resting period to ensure consistency. An analog pH sensor (ELC1057) is employed to measure the hydrogen ion concentration in saliva. The sensor operates based on electrochemical principles, generating an analog voltage proportional to the pH value of the sample. Prior to measurement, the sensor is calibrated using standard buffer solutions (pH 4.0, 7.0, and 10.0) to ensure measurement accuracy and reliability. The use of a salivary biomarker enables a patient-friendly, painless, and repeatable diagnostic approach, making the system suitable for frequent monitoring.

iv) Processing and Communication Layer

The Processing and Communication Layer acts as the central embedded unit of the system. An ESP32 microcontroller is selected due to its low power consumption, integrated Wi-Fi capability, and high-resolution 12-bit analog-to-digital converter (ADC). The analog voltage generated by the pH sensor is first conditioned and then digitized using the ESP32 ADC. The microcontroller performs initial signal processing operations, including noise filtering and scaling, to convert raw sensor outputs into standardized pH values. These digitized readings are timestamped and formatted for data transmission. Wireless communication is accomplished through the ESP32's built-in Wi-Fi module, enabling reliable data exchange with a local device or remote server. This real-time transmission capability supports continuous monitoring and allows seamless integration with telemedicine and cloud-based diagnostic platforms.

v) Intelligence Layer

The Intelligence Layer is responsible for data-driven analysis and decision-making. Digitized salivary pH values obtained from the ESP32 are combined with supporting parameters such as age, dietary habits, smoking status, alcohol intake, NSAID usage, and stress levels to construct a meaningful feature set. Machine learning techniques implemented using the scikit-learn library, particularly the Random Forest classifier, are employed to analyse the data and estimate gastric ulcer risk. This model enables effective handling of multiple input features while also providing interpretable insights into the relative importance of physiological and lifestyle factors. The AI models classify subjects into low-risk, moderate-risk, or high-risk gastric ulcer categories as shown in Fig 1.1. The output of the Intelligence Layer is presented as both a risk label and a prediction confidence score, enabling transparent and clinically interpretable decision support. This AI-driven approach enhances diagnostic accuracy while reducing dependence on invasive procedures.

vi) System Workflow Summary

The complete workflow of the proposed system begins with non-invasive saliva sample collection, followed by real-time pH sensing and digitization using the ESP32 microcontroller. The processed data are transmitted wirelessly to the AI module, where machine learning and deep learning models analyse the input features and generate gastric ulcer risk predictions. The modular architecture ensures adaptability, cost-effectiveness, and suitability for deployment in home-based and resource-constrained healthcare environments.

vii) Hardware Implementation

The hardware implementation of the proposed system consists of an analog pH sensor (ELC1057), an ESP32 microcontroller, calibration solutions, and necessary maintenance materials. The ELC1057 analog pH sensor is used to measure salivary pH over a broad range of 0–14 with an accuracy of ± 0.1 , ensuring dependable detection of physiological pH variations. The sensor output is interfaced with the ESP32 microcontroller, which is chosen for its low power consumption, built-in Wi-Fi capability, and 12-bit analog-to-digital converter that enables accurate signal digitization.

To ensure measurement precision and sensor stability, standard buffer solutions with pH values of 4.0, 7.0, and 10.0 are utilized for calibration prior to data acquisition. In addition, appropriate electrode maintenance is carried out using potassium chloride (KCl) solution for hydration and distilled water for cleaning, which helps extend sensor lifespan and maintain consistent performance during repeated measurements. The ESP32 supports real-time monitoring and reliable data transmission, making the system portable and scalable, as illustrated in Fig. 1.3.

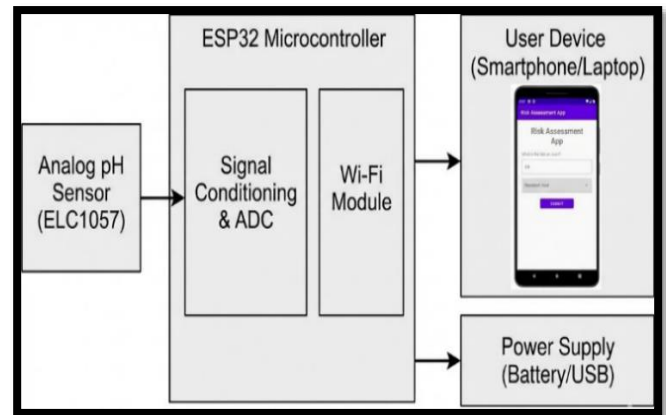


Fig 1.3 Hardware Execution

viii) Data Collection

Real-time salivary pH data were initially collected from participants and used to train the machine learning model for gastric ulcer risk classification. During the testing phase, fresh salivary samples were again obtained from individuals under similar controlled conditions to ensure consistency. Multiple pH readings were recorded for each participant, and an average salivary pH value was calculated to reduce measurement variability. Along with pH data, selected lifestyle and symptom-related information was collected through a brief questionnaire. These inputs were provided to the previously trained model to evaluate its prediction performance. The model output was then analysed to determine ulcer risk and assess classification accuracy. This process enabled validation of the trained system using newly acquired real-time data.

ix) Data Preprocessing

Raw biomedical datasets often include inconsistencies, noise, and variability that may negatively influence machine learning performance. Hence, thorough data preprocessing was carried out before model training to improve data quality and reliability. Numerical attributes, such as salivary pH and age, were normalized using min-max scaling to map all values into a consistent range between 0 and 1, as expressed in equation (1). This normalization limits the influence of features with larger numeric scales and supports faster and more stable model convergence during training. Categorical attributes including gender, dietary habits, smoking status, alcohol consumption, and NSAID usage were transformed into numerical formats using label encoding or one-hot encoding methods, based on the characteristics of each feature.

This transformation ensures compatibility with machine learning and deep learning algorithms. After preprocessing, the dataset was randomly shuffled and divided into three non-overlapping subsets to support reliable model development and evaluation. Specifically, 70% of the data was used for training, 15% for validation during hyperparameter tuning and performance optimization, and the remaining 15% for testing to provide an unbiased estimate of model generalization. This structured data partitioning enhances robust model evaluation and reduces the risk of overfitting. The pre-processed dataset thus forms a dependable foundation for AI-driven gastric ulcer risk prediction.

x) Artificial Intelligence Models

Artificial intelligence techniques were employed to analyse the preprocessed salivary pH dataset and predict gastric ulcer risk categories. Both machine learning and deep learning approaches were adopted to utilize their complementary strengths in interpretability and modelling complex non-linear relationships. The proposed framework incorporates a Random Forest classifier and a feedforward deep neural network to enable robust and accurate gastric ulcer risk prediction.

xi) Random Forest Classifier

Random Forest is an ensemble-based supervised learning algorithm that combines multiple decision trees to improve predictive accuracy and generalization performance. Each decision tree in the ensemble is trained on a randomly selected subset of the training dataset using bootstrap sampling, while a random subset of features is considered at each split node. This stochastic learning process reduces correlation among individual trees, thereby minimizing overfitting and enhancing model robustness.

In the proposed system, the Random Forest classifier processes salivary pH values together with demographic and lifestyle-related attributes to identify discriminative patterns associated with gastric ulcer risk. The final classification outcome is obtained by aggregating the predictions of all individual decision trees using majority voting or probability averaging. The probability of assigning a class label y given an input feature vector X is computed by averaging the outputs of all trees in the ensemble, as expressed in equation (1):

$$P(y | X) = \frac{1}{N} \sum_{i=1}^N T_i(X) \quad (1)$$

where N denotes the total number of decision trees in the forest and $T_i(X)$ represents the predicted class output or probability generated by the i^{th} decision tree. This aggregation strategy enhances classification stability by reducing variance and mitigating overfitting effects. Additionally, the Random Forest model provides feature importance measures, enabling identification of influential predictors such as salivary pH and NSAID usage. This interpretability is particularly valuable in biomedical applications, where transparent and explainable decision-making is essential.

Fig 1.4 Software Integration

III. Hardware Implementation

The hardware system is designed to enable real-time, non-invasive measurement of salivary pH for gastric ulcer risk assessment. An analog pH sensor (ELC1057) is used to measure hydrogen ion concentration in saliva, generating a voltage proportional to pH levels. The sensor output is interfaced with an ESP32 microcontroller, where the signal is digitized using its 12-bit analog-to-digital converter (ADC).

The ESP32 is selected for its low power consumption, compact size, and integrated Wi-Fi capability, enabling real-time wireless data transmission. Sensor calibration is performed using standard buffer solutions (pH 4.0, 7.0, and 10.0) to ensure measurement accuracy. Proper electrode maintenance using potassium chloride solution and distilled water is carried out to maintain stability and reliability. The overall system is portable, cost-effective, and suitable for continuous, non-invasive health monitoring.



Fig 1.5 Hardware Setup

xii) Experimental Results

The proposed system was evaluated using a pre-processed dataset divided into training, validation, and testing subsets. Random Forest classifier was trained using identical data partitions to ensure a fair performance comparison. Real-time salivary pH values were analysed both before and after food intake to capture physiological variations related to gastric acid secretion. In addition, detailed dietary and lifestyle factors, including meal timing, food type, and consumption patterns, were systematically examined and incorporated into the dataset. These combined features were provided as inputs to the Random Forest model, enabling effective learning of complex non-linear relationships between salivary pH variations and dietary behaviours.

IV. Results and Discussion

Although the proposed system shows promising results, certain limitations need to be considered. One major limitation is the use of a relatively small dataset for training and validation, which may affect the model's ability to perform consistently across different age groups, populations, and varying health conditions. Even though proper preprocessing and validation techniques were applied, increasing the dataset size would improve reliability and reduce potential bias.



Fig 1.5 Output

V. CONCLUSION

In addition to prediction, the system is designed to generate a detailed PDF report containing the patient's input data, salivary pH values, predicted risk level, and recommendations. This report can be shared with healthcare professionals for further clinical evaluation and decision-making. Future work will focus on improving diagnostic accuracy by incorporating additional salivary biomarkers such as buffering capacity, salivary flow rate, enzyme activity, and inflammatory markers. Expanding the dataset and performing real-world clinical validation will further enhance the system's effectiveness and usability in practical healthcare settings.

VI. FUTURE WORK

Although the proposed system produced encouraging results, certain limitations should be considered. One key limitation is the relatively small dataset used for training and validation, which may limit the model's ability to generalize across different age groups, populations, and health conditions. While appropriate preprocessing and evaluation strategies were applied, expanding the dataset would improve reliability and reduce the possibility of bias.

Future research will focus on strengthening the diagnostic performance by incorporating additional salivary biomarkers such as buffering capacity, salivary flow rate, enzyme activity levels, and inflammatory indicators. Integrating multiple biomarkers is expected to enhance prediction accuracy and provide a more comprehensive evaluation of gastric health. Further clinical validation and real-world testing will also be necessary to refine the system and improve its practical applicability in healthcare settings.

REFERENCES

- [1] M. Prunella et al., "Automated Pathomic Analysis of Angiogenesis and Immune Profiles Unveils an Interpretable Prognostic Biomarker in Colon and Gastric Cancers," IEEE Journal of Biomedical and Health Informatics, 2025.
- [2] J.-H. Kim, J.-W. Chae, H. Chin Cho and H.-C. Cho, "Enhancing Robustness in Gastroscopy Analysis With Poisson-Gaussian Noise-Based Adversarial Augmentation," IEEE Access, vol. 13, pp. 192391–192402, 2025.
- [3] D. C. Bui, B. Song, K. Kim and J. T. Kwak, "Spatially-Constrained and -Unconstrained Bi-Graph Interaction Network for Multi-Organ Pathology Image Classification," IEEE Transactions on Medical Imaging, vol. 44, no. 1, pp. 194–206, Jan. 2025.
- [4] S. Varadhaganapathy, S. Nandha, P. Pramanik and D. Rajasekar, "Gastrointestinal Ulcer Detection Using Deep Learning," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1–7.
- [5] Z. Wang, X. Zheng, J. Zhang and M. Zhang, "Three-Branch BERT-Based Text Classification Network for Gastroscopy Diagnosis Text," International Journal of Crowd Science, vol. 8, no. 1, pp. 56–63, Feb. 2024.
- [6] T.-H. Tran et al., "GIFCOS-DT: One Stage Detection of Gastrointestinal Tract Lesions From Endoscopic Images With Distance Transform," IEEE Access, vol. 12, pp. 163698–163714, 2024.
- [7] J.-Y. Jhang et al., "Gastric Section Correlation Network for Gastric Precancerous Lesion Diagnosis," IEEE Open Journal of Engineering in Medicine and Biology, vol. 5, pp. 434–442, 2024.
- [8] D. Ravenscroft and L. G. Occhipinti, "A Predictive Point-of-Care Platform for Early Detection of Periodontal Disease," 2023 IEEE SENSORS, Vienna, Austria, 2023, pp. 1–4.

- [9] J.-W. Chae and H.-C. Cho, "Enhanced Classification of Gastric Lesions and Early Gastric Cancer Diagnosis in Gastroscopy Using Multi-Filter AutoAugment," IEEE Access, vol. 11, pp. 29391–29399, 2023.
- [10] S. Ahmad et al., "Automated Detection of Gastric Lesions in Endoscopic Images by Leveraging Attention-Based YOLOv7," IEEE Access, vol. 11, pp. 87166–87177, 2023.
- [11] M. S. Alam, M. M. Rahman and M. A. Hossain, "IoT-based smart healthcare monitoring system using wearable sensors," IEEE Access, vol. 8, pp. 174951–174960, 2020.
- [12] S. Zhang, J. Li and Y. Chen, "Machine learning-based clinical decision support systems for gastrointestinal disease diagnosis," IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 6, pp. 1729–1738, 2020.
- [13] P. Kaufmann, A. Smolle and M. Graninger, "Noninvasive biomarkers in gastrointestinal diseases: current status and future perspectives," World Journal of Gastroenterology, vol. 25, no. 27, pp. 3456–3472, 2019.
- [14] R. T. Humphrey and J. P. Wong, "Saliva as a diagnostic fluid," Journal of Oral Microbiology, vol. 10, no. 1, pp. 1–8, 2018.
- [15] I. Goodfellow, Y. Bengio and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.