



OPEN

Segmentation of gastroesophageal reflux events using a semi-U-Net architecture with 1D/2D CNNs

Azra Rasouli Kenari & Hossein Rabbani

U-Net has gained traction in biomedical signal processing, particularly for segmenting 1D waveforms. Building on this success, we propose a U-Net-inspired architecture that integrates both 2D and 1D CNNs to effectively learn and segment gastroesophageal reflux (GER) events from Multichannel Intraluminal Impedance (MII) signals—specifically, a 6-channel 1D impedance signal. Current methods for GER detection are limited by the absence of efficient software, leading to time-consuming manual interpretation that is prone to errors. As a key contribution, we are also releasing the dataset of MII signals and GER annotations publicly to facilitate further research and algorithm development. In our architecture, the 2D CNN serves as the first encoder in a semi-U-Net structure to capture features across all channels. Subsequently, all other encoders and decoders utilize 1D CNNs to preserve the 1D nature of the signal while minimizing the number of parameters. After network training, the model segments GER areas in the 6th channel, utilizing a post-processing unit that accurately segments GER areas across all six channels. This unit ensures that selected GER events align with clinically defined criteria. The proposed architecture is compact and efficiently utilizes parameters, demonstrating strong generalizability across diverse GER events, with average durations of 17.52 ± 6.39 s. Outperforming existing methods, our approach achieves a sensitivity of 95.24% and a positive predictive value of 100%, indicating superior segmentation quality. We evaluated the model's robustness using data from 202 episodes containing 208 GER events collected from 26 patients who underwent 24-h MII pH monitoring. This semi-U-Net architecture, with its low parameter count, offers robust generalizability and adaptability to varying input durations. By improving GER event segmentation, our approach enhances the utility of 24-h MII-pH monitoring, enabling clinicians to make better-informed decisions for patient selection in invasive surgical procedures.

Keywords 1D biomedical signal, Artificial intelligence, Elementwise classification, Gastroesophageal reflux disease, Semi-U-Net

Gastroesophageal Reflux Disease (GERD) is a prevalent gastrointestinal disorder characterized by the backflow of stomach acid into the esophagus, leading to symptoms such as heartburn, regurgitation, and chest pain. It affects a significant portion of the global population and can have a considerable impact on an individual's quality of life if left untreated¹.

The initial step in managing GERD typically involves initiating Proton Pump Inhibitor (PPI) therapies. For those patients with atypical symptoms who do not respond adequately to antisecretory treatments, Classical diagnostic methods such as endoscopy and pH monitoring are used². Both methods have been mainstays in the field for many years, providing valuable assistance in therapeutic and diagnostic guidance. However, there are still many patients who have normal endoscopic findings and pH monitoring results despite ongoing GERD symptoms and unresponsiveness to PPI therapy^{3,4}.

In the realm of GERD diagnosis and management, Multichannel Intraluminal Impedance pH (MII-pH) monitoring has emerged as a valuable tool for assessing reflux events and their association with symptoms. The integration of MII with pH monitoring, significantly enhances the detection of refluxate presence, proximal extent, and clearance. By combining MII with pH analysis, it becomes possible to accurately differentiate between acid and non-acid GER events³.

MII monitoring measures electrical resistance within the esophagus, offering detailed insights into the movement and properties of refluxate⁵. It allows for the differentiation between liquid, gas, and mixed reflux events while identifying the direction of bolus movement, differentiating swallows from refluxate. This robust

Medical Image and Signal Processing Research Center, School of Advanced Technologies in Medicine, Isfahan University of Medical Science, Isfahan, Iran. email: Rabbani.h@ieee.org

diagnostic tool enables clinicians to correlate symptoms with reflux events accurately. Currently, MII-pH monitoring is the gold standard for diagnosing GERD, offering superior diagnostic capabilities compared to conventional methods⁶.

While MII-pH monitoring represents a significant advancement in GERD diagnostics, the accurate interpretation of impedance data is paramount to its effectiveness in clinical practice⁷. Proper analysis of impedance measurements is essential for maximizing the utility of MII-pH monitoring and ensuring optimal patient outcomes in the management of GERD⁵.

Several clinical studies have highlighted the importance of thorough analysis and interpretation of impedance data^{5,7–9}. However, the automatic detection of GER events using MII data has been explored in only a limited number of studies^{10–13}.

In¹² a cascade Multivariate Long Short-Term Memory with Fully Convolutional Network (MLSTM-FCN) system was introduced for GER detection and diagnosis using MII-pH signals. In¹², the authors classified 84 acid reflux events versus 141 swallows or artifacts. Their detected events are characterized as either 6, 11, or 21 s long. This study primarily relied on pH sensor data, limiting its scope to identifying acid GER events. Their method relies on an initial preprocessing step based on pH threshold comparison to identify potential acid reflux events, followed by MLSTM-FCN classification to distinguish acid reflux from artifacts or swallows.

A limitation of their approach is that it only provides an approximate duration of a reflux event, categorized as 6, 11, or 21 s, without localizing the precise onset and offset times across the six channels. Another limitation is that they consider only acid reflux events, whereas GERs include both acid and non-acid refluxes. Non-acid refluxes are prevalent in clinical practice¹⁴ and should not be excluded.

Additionally, artifacts such as pH readings outside the 0–14 range are omitted in their method, which could lead to missing genuine reflux events, because many reflux events (acid and non-acid) can still be characterized via MII data. This ability to characterize diverse reflux events is precisely why MII-pH monitoring remains the gold standard for diagnosing GERD¹⁵.

In contrast¹³, introduced S4, a state-space sequence model that serves as a versatile building block for modeling signal data. The S4 model is treated as a “black box” with limited details provided. Initially developed for multidimensional signals such as images and videos, the S4 model was adapted into a multiscale architecture capable of handling extremely long audio sequences, including MII data. This innovative approach provided a universal framework for modeling diverse, multidimensional signals. However, the method achieved only 68.7% sensitivity and 80.8% specificity in identifying GER events within a dataset of 45 patients.

A limitation of this study is that in this study, MII data are segmented into 60-second clips labeled as containing reflux or not, but the exact start time of reflux within each segment is not specified. Moreover¹³, does not specify how they distinguish GER events from swallows or artifacts. Since swallows and peristaltic waves are highly prevalent, it is essential to address the need for their accurate differentiation.

Notably, both the MLSTM-FCN and S4 approaches were limited to event detection and did not address the segmentation of GER events across all six channels.

The method described in¹¹ was an initial exploration aimed at understanding the structure of unfamiliar MII-pH data. In¹¹, a time-domain analysis of MII tracings was conducted to characterize liquid events, deliberately excluding the common occurrence of mixed GERs. A limitation of this approach is that it focuses solely on liquid reflux events, neglecting the more common mixed refluxes.

Additionally¹⁰, successfully automated the characterization of both mixed and liquid GER events through sparse representation. This study posited that MII represents a sparse signal comprised of GER and swallow events and extended periods of inactivity, corresponding to isoelectric intervals, as well as noise and interferences. It demonstrated the effectiveness of sparse representations for modeling MII, given that GER or swallow events are infrequent in comparison to isoelectric intervals. However, we believe that the success of deep learning methods in¹⁰ has been hampered by the limited availability of a large dataset of MII tracings.

Artificial Intelligence has recently expanded its applications to biomedical signal processing^{16–25}. Attempts have been made to leverage U-Net²⁶, for waveform segmentation tasks in various biomedical signals, including ECG^{27,28}, plethysmography²⁹, and heart sound signals³⁰. In this study, we aim to investigate the feasibility of employing a semi-U-Net deep learning architecture to MII data. As detailed in¹⁰, GER is associated with several variables of clinical importance. Our method accounts for both acid and non-acid, as well as liquid and mixed physical states. It accurately delineates the onset and offset times of each GER event, enabling detailed segmentation rather than simple classification. Furthermore, our approach effectively differentiates GER events from artifacts and swallows through model training, and employs a dedicated post-processing step to separate GERs from swallows. We leverage onset times across multiple channels and exploit the characteristic retrograde propagation pattern of GERs, with specific criteria discussed in³¹.

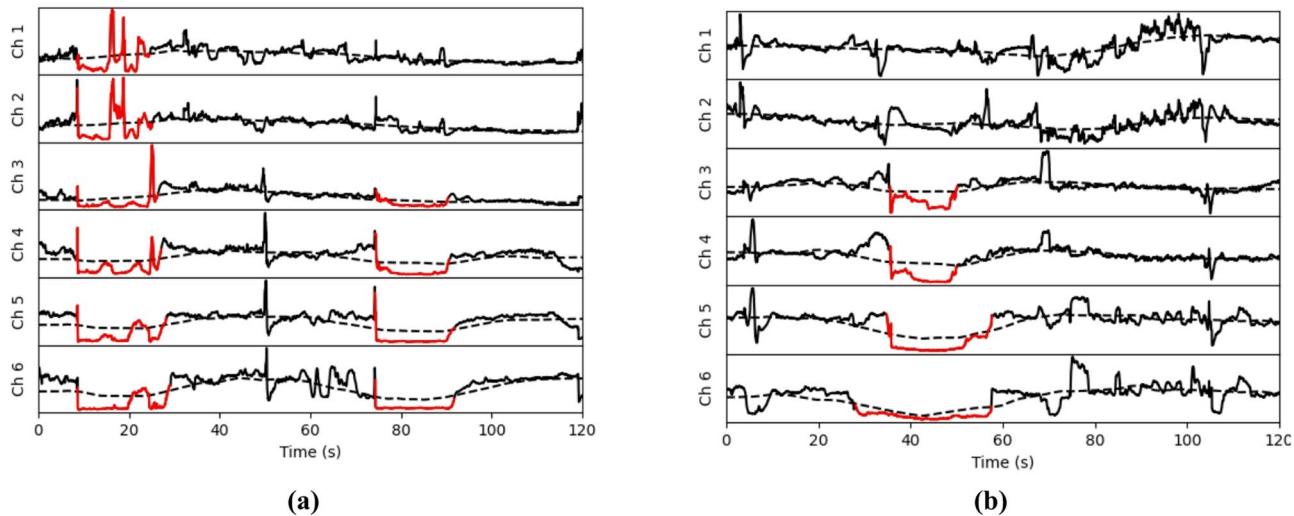
To our knowledge, most existing studies only identify the peak or approximate start of events in 1-D signals, often with significant error margins. In contrast, our method uses deep learning to simultaneously localize both the onset and offset of GERs with high precision. We introduce a semi-U-Net architecture combining 2D and 1D CNNs that exploit the relationships between channels for effective segmentation of GERs from MII signals. This compact design efficiently utilizes parameters, enhancing its ability to generalize across diverse GER events.

In the following sections, we will first review the characteristics of our dataset. Next, we will provide an overview of our neural network architecture, inspired by the U-Net. We will then present the statistical analyses and results obtained from our study. Finally, we will conclude with a summary of our findings.

Methods

In our study, we focused on learning GER patterns without directly addressing the extended duration of MII data, which covers 24 h. This challenge can be managed through a preprocessing approach that analyzes the MII signal in non-overlapping 2-minute segments. Candidate intervals can be efficiently identified by calculating the

MII dataset	# Episodes		# GERs		GER duration (s)	
	Initial ¹⁰	Extended	Initial ¹⁰	Extended	Initial ¹⁰	Extended
Train (train-validation)	141 (140-1)	182 (161-21)	142	187	2466	3345
Test	33	20	33	21	532	384
Total	174	202	175	208	2998	3729

Table 1. Dataset description.**Fig. 1.** Sample MII data, all non-zero labelled signal points are emphasized as red, dashed lines are representative of Baseline Impedance (BI) which is extracted using a moving average filter with a 30 s window size. **a** An episode including two GER events, **b** another episode that includes one GER event.

entropy of the 24-hour signal³². Segments with entropy values exceeding a predefined threshold are then selected for further analysis using our proposed models. However, in our dataset, these 2-minute episodes were manually chosen to ensure a diverse representation of GER patterns.

MII dataset

Our dataset is comprised of 202 episodes, including a total of 208 GER events, which is an extended version of the dataset that was initially presented in¹⁰. All setups for acquiring this extended dataset mirrors the procedures outlined in the previous study¹⁰. The dataset can be accessed at³³.

A typical 24-h MII-pH study recording often includes prolonged durations of isoelectric intervals, which may not be optimal for effective learning⁵. To address this issue, we extracted 2-min episodes from the MII data that are both informative and well-balanced. Each of these episodes spans a duration of 2 min and includes at least one GER event, with the potential inclusion of swallows, as detailed in¹⁰.

A thorough review process was conducted with three expert gastroenterologists to ensure accurate and consistent labeling of GER events across six channels. Their insights helped define event characteristics, leading to a standardized labeling system that improved data reliability and coherence.

The dataset overview in Table 1 provides a structural summary. Each episode spans a 2-min interval sampled at 50 Hz across six channels, with GER events labeling approximately 30.77% of the total data.

Figure 1 depicts plots of two samples from the dataset, showcasing two distinct episodes of MII data across the six channels plotted against the time axis, along with their corresponding labels.

Signal normalization

Normalization of the amplitude of MII data was done before inputting it into the semi-U-Net structure. Studies have highlighted the importance of considering impedance levels in the interpretation of MII data, with variations in impedance values observed between proximal and distal channels within an empty lumen³⁴. Normalization of the amplitude helps mitigate the impact of these differences, particularly given the differential presence of air and potential variations in cross-sectional areas between electrodes. Moreover, research has shown that esophageal impedance levels can vary among individuals, underscoring the need for standardized normalization procedures to reduce bias in the learning process³⁴.

Amplitude normalization is intended to adjust signal amplitudes to a consistent range. This process helps reduce the impact of differing magnitudes on the convergence speed of the model and promotes the stability of numerical calculations³⁵. The normalization step was implemented using (1):

$$x_n = \frac{x}{\|x\|_2} = \frac{(x_1, x_2, \dots, x_n)}{\sqrt{x^T x}} \quad (1)$$

In (1), x represents the impedance data of a specific channel. The normalization process was carried out individually for each channel.

Proposed approach

In this work we propose a two-stage scheme for the GER segmentation task which includes (1) elementwise classification at the 6th channel and (2) event segmentation across all of the channels. Our approach leverages the concept of making an initial estimation of potential GER events using MII data from all channels. Subsequently, we refine this initial estimation to determine the authenticity of the GER event, as well as to characterize its onset and offset times across all channels. Our proposed scheme, illustrated in Fig. 2 comprises two main components: the element-wise classification, detailed in section C1, and the GER event segmentation, described in section C2.

The neural network architecture A semi-U-Net architecture is introduced that integrates both 2D and 1D CNNs for effective segmentation of GER events from MII signals. The network takes as input normalized impedance data from all six channels, along with the ground-truth label of the 6th channel. A number of 182 episodes were used for training of the network with holdout cross validation method.

Processing impedance data was based on the utilization of a U-Net like structure and CNN layers to extract feature maps. We proposed to keep the number of network parameters low, in order to be able to effectively learn from the limited input dataset³⁶. The design of the neural network architecture is based on the following principles.

Convolutional layers play a crucial role in extracting meaningful features from raw MII data. By utilizing stacked convolutional layers and multiple feature maps, the model can effectively extract rich and valuable features³⁷. Studies have shown a correlation between input length and segmentation accuracy in models³⁸. Consequently, our proposed architecture featured an initial two-dimensional convolutional layer. This two-dimensional convolution enables the network to monitor changes in impedance throughout the esophagus effectively.

The pooling layer reduced the input length by half and captured relevant MII features. The temporal location of a GER in an episode was not considered a crucial feature to preserve. Pooling layers helped create representations that were almost invariant to translation³⁹.

A sigmoid function was employed to scale output values within the range of 0 to 1⁴⁰. We evaluated with the DICE score, a well-established metric for segmentation accuracy evaluation in the field of computer vision⁴¹. The proposed algorithm employs an encoding-decoding architecture, with detailed specifications provided in Table 2 and a corresponding block diagram illustrated in Fig. 3.

The input of the network was a matrix of size $6 \times l$, where l represented the length of the two-minute MII episode, each episode contains $l = 2 \times 60 \times f_s$ samples for each channel. Despite other frameworks that suggest independent channel analysis⁴², we proposed simultaneous analysis of all 6 channels. In section III we will compare result of our proposed framework to the case in which each channel is analyzed individually.

The initial convolutional layer is two-dimensional with a kernel size of (211×6) to expand the receptive view of the network. During the annotation process, gastroenterologist experts considered all impedance channels together ensuring a thorough understanding of the dynamics and propagation of GER events within the esophageal lumen. This idea was inspired in choosing a 2D convolution at the first layer to gain a comprehensive understanding of GER event dynamics.

Reflux events typically initiate and terminate at the most distal site, making the 6th channel crucial for reflux duration detection. Based on this information, the network was trained to classify each sample point at the distal site as GER or Non-GER. The output was a vector of size $1 \times l$, compared to the target vector representing the label of the 6th channel. The output of the network is an intermediate result that presents elementwise classification of the 6th channel. To localize a GER event, additional process steps are needed. Post processing generates final segmentation result in terms of GER event onsets and offsets across all 6 impedance sites.

Post-Processing To finalize GER segmentation the post processing step was conducted. The output of the previous step is the primary flag of the 6th channel. As described earlier, our final goal is to segment a GER event across all channels, meaning that we should find a final flag, the same size as input MII data that provides element wise classification for each sample at each channel. The post processing step can be considered as three blocks that are described below.

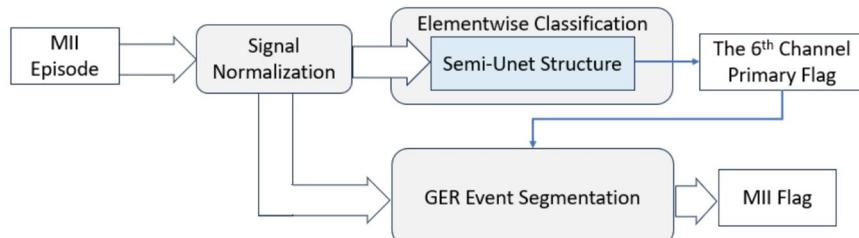


Fig. 2. Block diagram of the proposed method.

Layer	Layer type	#Channels	Unit type	Kernel size	Stride	Padding	Output
<i>Encoder</i>							
Input*		1					(1, 6000, 6)
C1**	Conv 2-d	8	Relu	(211, 6)		(105, 0)	(8, 6000)
P	Pooling		Max		(2, 1)		(8, 3000)
C2	Conv 1-d	16	Relu	21		10	(16, 3000)
P	Pooling		Max		2		(16, 1500)
<i>Decoder</i>							
D1	Transpose Conv 1-d	8	Relu	2	2		(8, 3000)
D2	Transpose Conv 1-d	1		2	2		(1, 6000)
Output	Softmax		Sigmoid				(1, 6000)

Table 2. The proposed NN architecture for elementwise classification of input MII data. *, **: Attribute functions unsqueeze(1) and squeeze(3) were used before input and after C1 layers, respectively.

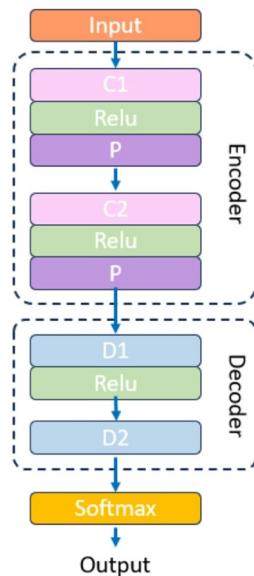


Fig. 3. The architecture of the network. The first layer, C1, is special as it uses a 2D CNN to capture features across all channels of MII data, the rest of the encoders and decoders use 1D CNNs to preserve the 1D nature of the signal.

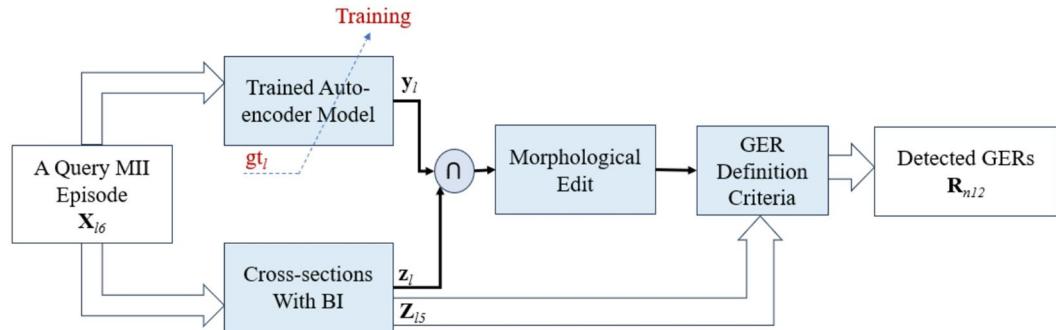
Cross sections with BI According to clinical literature a GER event typically initiates right after the impedance data reaches 50% of its baseline value¹¹. In order to investigate this criterion, Baseline Impedance (BI) is determined by applying a 30-second moving average to each channel individually⁴³. Next, BI is used to identify cross sections, where the impedance value reaches 50% of the BI. Subsequently the output is divided into a vector z_l , specific to channel 6 and a matrix Z_{l5} for the remaining channels. The intersection between z_l and output of the semi-U-Net model, y_l is computed to ensure the selection of GER events that align with the defined criteria, potentially enhancing the specificity of the detected GERs.

In addition to this, the definition of a GER event has other criteria, which are further investigated in the third block of our post-processing step.

Morphological editing In the second block, a morphological editing process is applied to the outcomes of the preceding step. This stage consists of four morphological sub-blocks that function similarly to their two-dimensional counterparts, known as opening and closing operations⁴⁴. Each sub-block involves the application of a moving average filter followed by thresholding. The specifics of kernel size and threshold values (θ) for each sub-block are summarized in Table 3. All kernels are rectangular in shape and normalized to have an L2-norm of 1. The result of this block is the determination of the onset and offset points for each GER event at the 6th channel.

GER definition criteria The last block of the post-processing procedure is dedicated to channels 5 through 1. This block takes as input Z_{l5} together with the predicted labels from the 6th channel. Its objective is to determine onset and offset points for each assigned GER across channels 5 to 1.

Sub-block name	Kernel size	Thresholding	
		Type	θ
Opening	80	Relu	0.95
Closing	200	Anti-Relu	0.01
Opening	100	Relu	0.85
Closing	200	logical	0.01

Table 3. Details of morphological editing block.**Fig. 4.** Block diagram of the proposed post-processing procedure.

The ground truth labels of channels 5 to 1 from the training dataset were used to regularize parameters of this block with the objective to best adapt with the definition criteria of GER events^{11,31}. Then the optimized block was employed to predict labels for a query MII episode. The output of this block, known as R_{n12} is the final result of the post-processing procedure which is a matrix with n rows and 12 columns. With n representing the number of detected GER events, and the number of columns corresponds to 6 couple of onset and offset points belonging to the channels. In reporting values of R_{n12} , Any empty values were denoted by a “-” mark.

Figure 4 shows details of the post-processing step. Line and bold arrows symbolize the flow of single (vector) and multi-channel (matrix) data among different blocks, respectively. Each matrix is displayed in capitalized bold format, while each vector is in non-capital bold format. The subscript of each matrix or vector denotes its size. The vector gt_l represents the ground truth annotations of the 6th channel which was used to train the semi-U-Net network. The output of the network is denoted by y_l which is the predicted label flag of GER at semantic segmentation level.

Figure 5 further elaborates on Fig. 4 and illustrates the roles of each component in the post-processing pipeline and how they collaboratively contribute to the final GER detection for a query MII episode.

In summary, Fig. 6 provides an overview of our proposed model’s entire architecture, illustrating the process from MII episode input to GER event detection output. This figure integrates Figs. 2 and 4, highlighting the input, processing steps, and final detection results of our approach.

Results

The proposed semi-U-Net neural network was trained and experiments were conducted using the extended MII dataset. To assess the performance of the method the dataset was randomly divided into train and test sets, comprising 182 and 20 episodes, respectively, as described in Table 1. Within the training episodes, 21 were set aside for validation during the neural network training and later combined for the final fine-tuning step.

To provide more insight on the regularization of the proposed method, Fig. 7 shows the accuracy and loss plots against iteration for both the train and test folds. Based on the analysis of accuracy and loss plots, our proposed method demonstrates effective regularization, indicating a good balance between model complexity and generalization ability. The consistent performance across both training and testing folds suggests that the model is not overfitting to the training data and can reliably segment GER events in unseen data.

The Adam optimization method was employed to minimize loss by adjusting the network weight. During the initial training, the learning rate was set to 0.001, the mini-batch size was 32, and the training was limited to 150 epochs. To prevent overfitting, the maximum epoch was set to 55 and model parameters at epoch 55 were saved for subsequent fine-tuning. The best network weights, determined by the maximum F1-score, were saved at this stage.

To further enhance the network’s performance and capitalize on the knowledge gained during the initial training, a fine-tuning stage was also implemented. During fine-tuning, the training and validation data were combined to create a larger training dataset. The network’s weights with the best F1-score were loaded, and training with the expanded dataset took place for the specified number of epochs for fine-tuning (which was set to 10).

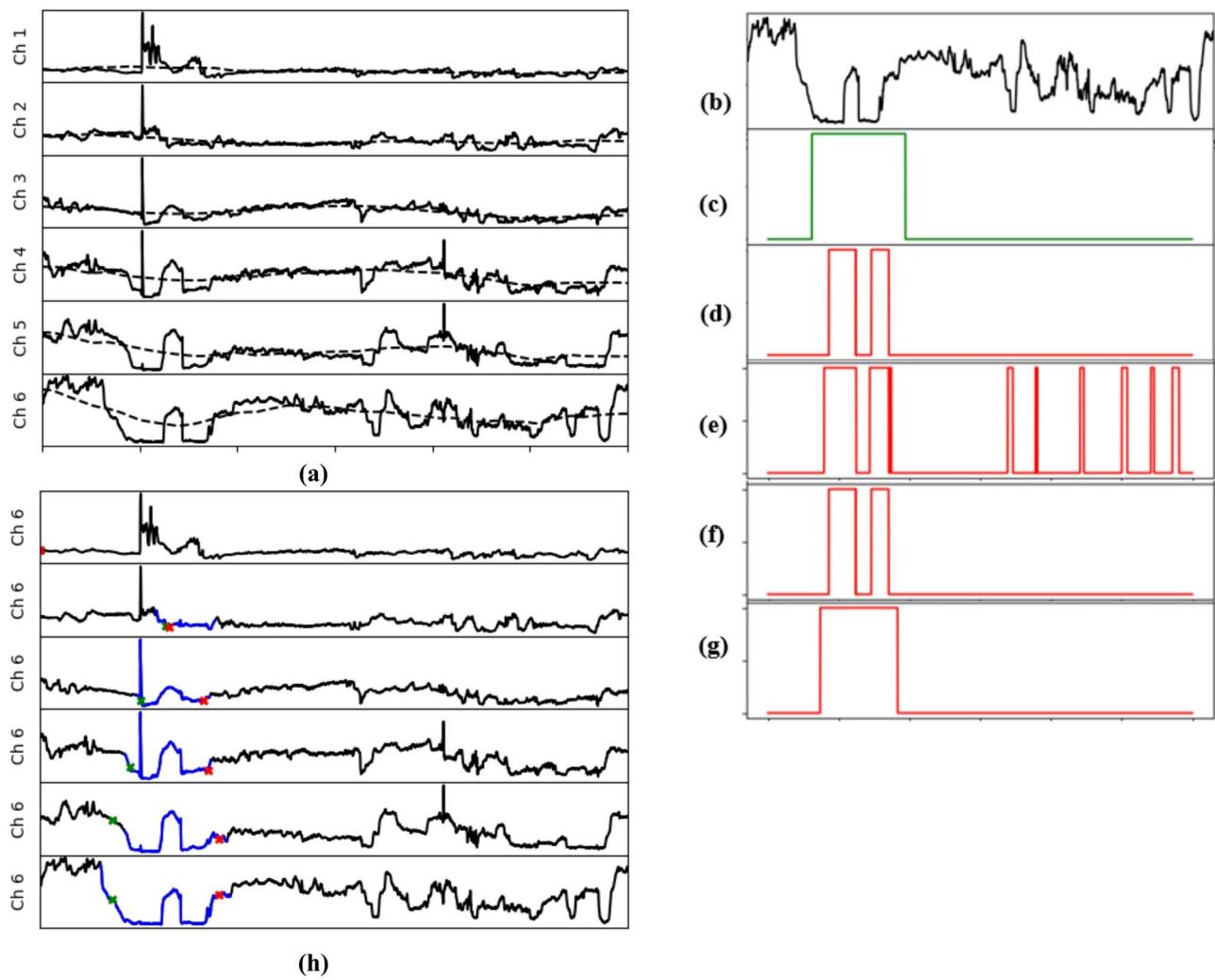


Fig. 5. **a** The 133th MII episode, X_{133} , along with its baseline impedance, **b** the 6th channel Impedance data. **c** The ground truth annotation for the 6th channel, g_{t_p} . **d** The output of the model, y_p . **e** the output of the cross sections with BI block, z_p . **f** the output of the dot product of **d** and **e**. **g** output of the morphological Edit block. **h** the same episode from **a** with ground truth annotation highlighted in blue, and detected GER event, R_{n12} , marked with green crosses indicating the onset points and red crosses indicating the offset points. The horizontal axis in all panels indicates time.

The performance of the proposed method was evaluated with accuracy (Acc), sensitivity (Sen), positive predictive value (PPV), F1 score and Temporal Error (TE) determined as follows⁴⁵:

$$Acc \ (\%) = 100 \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$F1_Score \ (\%) = 100 \frac{2 \times TP}{2 \times TP + FP + FN} \quad (3)$$

$$Sen \ (\%) = 100 \frac{TP}{TP + FN} \quad (4)$$

$$PPV \ (\%) = 100 \frac{TP}{TP + FP} \quad (5)$$

$$TE = |actual \ GER \ duration - detected \ GER \ duration| \quad (6)$$

The TP, TN, FP and FN are calculated using output of the network and label annotations of the 6th channel and is described by the confusion matrix of Table 4.

Building upon the evaluation metrics outlined above, Table 5 reports performance of our proposed semi-U-Net architecture at elementwise classification level for the 20 test episodes, that contain 21 GER events. Notably, prior research in¹², indicated that using a single classifier yields unsatisfactory results across GER events of

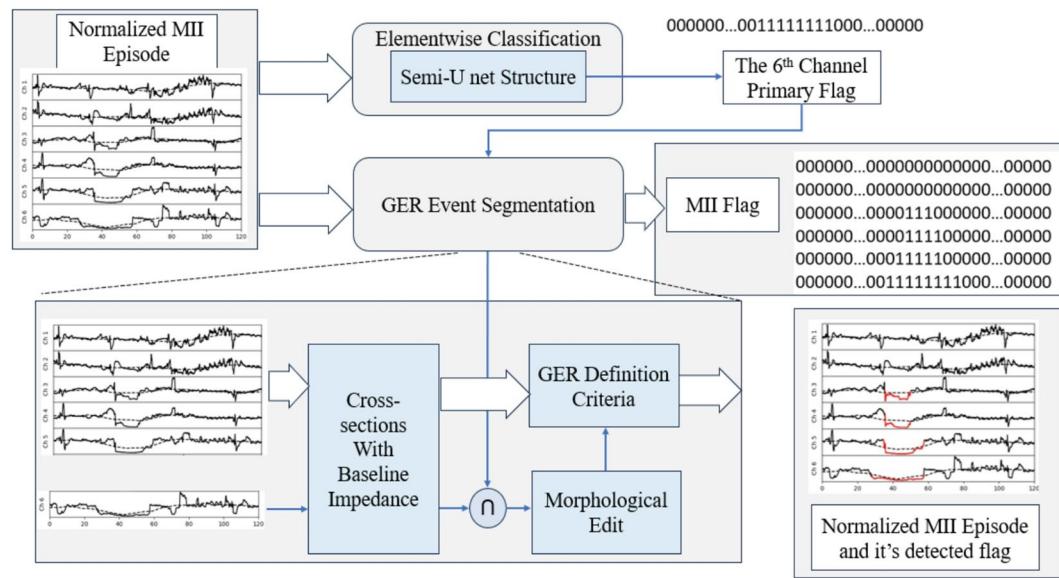


Fig. 6. The proposed two-stage scheme for the GER segmentation task which includes *elementwise classification* at the 6th channel and *event segmentation* across all of the channels. It makes an initial estimation of potential GER events then refines it to determine the authenticity of the GER event, as well as to characterize its onset and offset times across all channels.

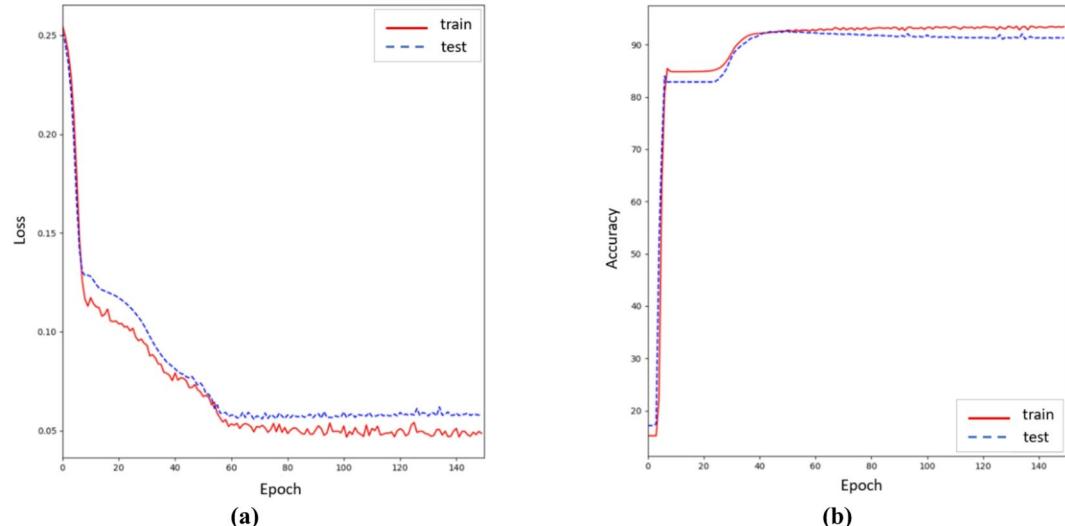


Fig. 7. Performance of the network versus epochs, for the train and test phases: **a** MSE loss and **b** accuracy.

Output of network	0	1
Ground truth annotations		
0	TN	FP
1	FN	TP

Table 4. Confusion matrix for elementwise classification.

Episode #	Semi-U-Net model	Acc (%)	Sen (%)	PPV (%)	F1	Duration* (s)	TE* (s)
197	2D-1D	90.8167	91.4538	66.7383	0.7717	20.36	1.82
	1D	89.7000	93.6149	63.2802	0.7552		1.84
80	2D-1D	93.7667	85.5225	74.5547	0.7966	13.90	0.24
	1D	93.0750	86.5733	71.1612	0.7811		1.44
4	2D-1D	91.7778	64.8620	78.3456	0.7097	20.36	10.50
	1D	91.5278	66.8698	75.6285	0.7098		10.36
93	2D-1D	92.4125	64.6980	82.8826	0.7267	19.06	2.02
	1D	92.2417	68.3057	79.0842	0.7330		0.70
133	2D-1D	92.7083	60.0715	86.3599	0.7086	26.46	4.58
	1D	92.3806	63.7801	80.5324	0.7118		3.70
172	2D-1D	93.4762	61.9015	87.8173	0.7262	11.14	2.70
	1D	93.2691	65.7863	82.5000	0.7320		3.32
52	2D-1D	93.8625	63.9344	89.5717	0.7461	18.04	0.58
	1D	93.7562	67.8334	84.8670	0.7540		1.08
142	2D-1D	94.2019	65.9539	90.9887	0.7647	18.90	0.08
	1D	94.1500	69.6734	86.7797	0.7729		0.40
102	2D-1D	94.4750	66.2464	91.6293	0.7690	12.24	0.40
	1D	94.4183	70.0648	87.2067	0.7770		0.58
191	2D-1D	94.6606	67.5999	91.1111	0.7761	14.18	1.50
	1D	94.6545	71.5946	87.0677	0.7858		1.80
97	2D-1D	94.5944	67.7558	92.0766	0.7807	23.70	3.74
	1D	94.6972	72.0505	88.4578	0.7942		3.70
24	2D-1D	94.1141	69.2521	88.4040	0.7766	26.08	0.04
	1D	94.1923	73.4166	85.2337	0.7889		0.52
157	2D-1D	94.1893	71.7924	87.0909	0.7871	20.76	3.92
	1D	94.3643	75.6129	85.0492	0.8005		4.24
120	2D-1D	94.3956	71.7088	87.5431	0.7884	10.78	1.58
	1D	94.5944	75.6163	85.5761	0.8029		2.38
148	2D-1D	93.9896	72.1398	82.6488	0.7704	6.28	4.30
	1D	94.1573	76.1422	80.9252	0.7846		4.74
2	2D-1D	93.8951	70.9266	83.4339	0.7667	20.24	10.14
	1D	94.1265	75.1750	81.8271	0.7836		7.52
32	2D-1D	93.5500	70.6131	82.3407	0.7603	24.28	7
	1D	93.7472	74.9409	80.5372	0.7764		10.88, 4
108	2D-1D	93.6561	72.2506	83.1230	0.7731	28.12	2.72
	1D	93.8605	76.3329	81.4495	0.7881		2.52
181	2D-1D	93.4933	71.6150	83.5766	0.7713	10.42, 7	0.02, –
	1D	93.7675	76.3513	81.7716	0.7897		0.36, –
144	2D-1D	93.4683	71.8502	82.9500	0.7700	15.72	1.96
	1D	93.5063	76.2255	80.1371	0.7813		3.00
Total	2D-1D	93.5743	69.8250	84.6066	0.7603	17.52 ± 6.39	2.99 ± 3.09
	1D	93.5027	73.5129	81.3523	0.7683		3.25 ± 3.09

Table 5. Performance of the proposed GER detection method for the 6th channel at elementwise classification level. *Amounts are reported as mean \pm std. For the total case (averaged results across all 20 test episodes), the bold values highlight the highest scores achieved when comparing the two methods, 2D-1D and 1D.

varying durations. Our analysis, as illustrated in Table 5, reveals that the mean duration of GER events is 17.52 s, accompanied by a standard deviation of 6.39 s. This significant variation underscores the challenges associated with analyzing MII data, highlighting the diverse nature of GER events.

In our investigation, we also evaluated a method that analyzes each channel of the MII data independently, called as 1D semi-U-Net model. The results obtained from this approach are comparable to those of our proposed method. However, the semi-U-Net model, which integrates data from all channels, demonstrates slightly superior performance. Specifically, Table 5 indicates that the total TE for the proposed method is 2.99 ± 3.09 , compared to 3.25 ± 3.09 for 1D model, which underscores the effectiveness of our approach in accounting for the complete set of MII channels.

The results presented in the Table 5 correspond to the output of the trained network before applying the dedicated post-processing steps. In the following, we aim to systematically assess the impact of each post-

Configurations		Acc (%)	Sen (%)	PPV (%)	F1
Name	Description				
C1	Without GER definition criteria	93.12 \pm 3.82	88.60 \pm 4.85	73.62 \pm 9.33	0.79 \pm 0.06
C2	Without morphological edit	93.48 \pm 1.18	66.41 \pm 6.38	87.08 \pm 6.90	0.74 \pm 0.03
C3	Without cross sections with BI	92.37 \pm 3.88	89.99 \pm 4.78	70.15 \pm 8.50	0.78 \pm 0.05
C4	No post-process	93.57 \pm 0.92	69.82 \pm 7.08	84.60 \pm 5.91	0.76 \pm 0.02
C5	Proposed method	96.72 \pm 0.88	84.90 \pm 6.16	92.05 \pm 1.71	0.88 \pm 0.03

Table 6. Quantitative evaluation of component contributions in the proposed framework via ablation study. The bold values in the Proposed Method row denote the best performance scores for the metrics presented.

Method	# GERs		TP	FP	FN	F1-score (%)
	Train	Test				
Sparse representation	141	33	32	2	1	95.52
1D Conv semi-U-Net	187	21	20	1	1	95.24
2D-1D Conv semi-U-net	187	21	20	0	1	97.56

Table 7. Comparative analysis of GER event localization between proposed and other Methods. The bold value corresponds to the highest F1-score, achieved by the 2D-1D Conv Semi-U-Net method.

processing component by analyzing different configurations of the proposed method. Table 6 summarizes the quantitative effects of each component within the framework. To evaluate the contribution of individual blocks, we tested various configurations, each omitting a specific processing step. Excluding the “GER Definition Criteria,” “Morphological Edit,” or “Cross Sections with BI” blocks, or bypassing post-processing entirely—using only the raw network output—allows us to understand how each component influences overall detection performance and robustness. Table 6 reports the results of each configuration on the test set at the element-wise classification level.

As shown in Table 6, the elimination of the two blocks ‘GER Definition Criteria’ and ‘Cross sections with BI’ does not have a significant effect on the final results. This can be justified by the fact that these two blocks are dedicated to other objectives more than just to enhance the performance metrics. The ‘GER Definition Criteria’ and ‘Cross sections with BI’ blocks were added to provide GER segmentation across all channels and to ensure that each detected GER fulfills clinical literature criteria. They might be essential in a real-world setting for ensuring clinical validity but could be optional for purely performance-focused tasks.

The C4 and C2 configurations results in degraded performance, as indicated in Table 6. In C2, the removal of the ‘morphological edit’ block slightly degrades overall performance, indicating its role in improving the quality of results. Interestingly, metrics such as accuracy, sensitivity, and F1 score are slightly higher in C4 (raw network output) than in C2. This may be because, in C4, the raw output isn’t truncated via intersection operations, whereas in C2, some GER areas might be truncated during those intersection procedures with the BI block. Meaning that if we use the output of the model solely we get a higher sensitivity but the PPV in C2 -after some operations of the post processing step, except for the morphological edit-, is higher meaning that each candidate GER has a higher probability of being a true positive, possibly reducing false positives.

The proposed method, C5, outperforms all ablation configurations, confirming that each component contributes positively to the system’s robustness and accuracy. The superior performance of C5 over C1 may be attributed to its ability to effectively distinguish true GER events from swallows and artifacts, thereby reducing false positives and ensuring more precise detections.

At event segmentation level, Table 7 presents the performance metrics for GER localization, comparing the results of our proposed method with those obtained from our previous approaches that utilized sparse coding and discriminative dictionary learning methods¹⁰. Additionally, we include a comparison with the 1D convolutional semi-U-Net model. Our algorithm achieved an F1-score of 97.56% for localizing GER events, outperforming the other methods.

Examples of GER event segmentation using our proposed method are illustrated in Figs. 8 and 9. Figure 8 specifically presents the segmentation results for the 197th episode, which was excluded from the database in¹⁰ due to excessive noise. As evident in the figure, this episode exhibits significant noise. Remarkably, our proposed method effectively handles both mixed GER types and challenging noisy episodes. As indicated in Table 5, the proposed method segmented this GER event with a TE of just 1.82 s. This error is relatively minimal compared to the overall duration of the GER, which is 20.36 s, making it negligible for practical purposes.

Figure 9 shows segmentation results for the 4th episode which has a considerable amount of TE. We believe that the insufficiency of the segmented area is initiated from the post-processing which require refinement to optimize the output results effectively.

Processing of the algorithm takes 26 ms per 2 min of the MII signal on a system equipped with an Intel(R) Xeon(R) E-2176 M CPU, 32 GB of RAM, and an Intel(R) UHD Graphics P630 GPU with 16 GB of VRAM, which makes it suitable for online implementations as well.

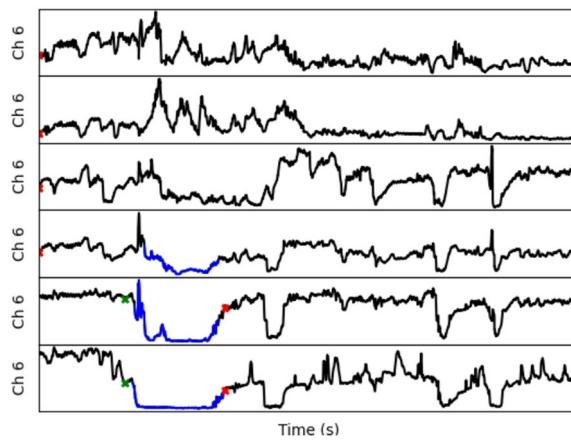


Fig. 8. Example of GER event localization demonstrating a low TE. The blue highlighted areas represent the ground truth, while the green cross marks indicate the onset and red cross marks denote the offset of the localized GER event.

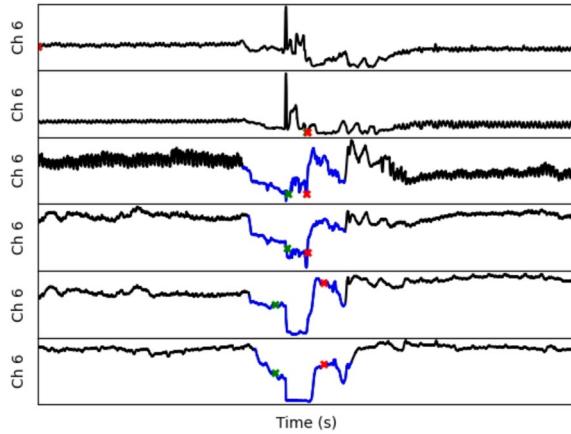


Fig. 9. An example of GER event localization with a high TE. The blue highlighted areas represent the ground truth, while the green cross marks indicate the onset and red cross marks denote the offset of the localized GER event.

Discussion

In this work, we presented a novel semi-U-Net architecture that combines 2D and 1D convolutional neural networks to effectively learn and identify GER events. The model was evaluated using a robust dataset of 202 episodes containing 208 GER events, collected from 26 patients undergoing 24-h MII pH monitoring. After training of the network, a post-processing unit was implemented that accurately segmented GER areas across all 6 channels, ensuring alignment with clinically defined criteria.

Our proposed architecture is characterized by its compact size, low weight, and efficient utilization of a small number of parameters, enhancing its generalizability. Our method demonstrates adaptability to varying durations of input data, making it versatile for application in scenarios with diverse data collection settings.

Our method achieved a sensitivity of 95.24% and a positive predictive value of 100%, demonstrating superior segmentation quality compared to existing methods. Notably, our method outperforms existing approaches in terms of segmentation quality, boasting an F1-score of 97.56%. Its robust generalizability is underscored by its ability to effectively handle challenging episodes that posed difficulties for previous methods.

The potential application of our approach in segmenting GER events holds promise for enhancing the utility of 24-hour MII-pH monitoring devices. The algorithm processes MII signals at a speed of 26 ms per 2 min, making it suitable for online implementations in clinical settings. Healthcare providers can leverage the segmentation results generated by our network to potentially improve decision-making processes, particularly in selecting patients for invasive surgical procedures.

While our method has yielded improvements in GER characterization, several limitations warrant further investigation. Currently, our pipeline relies on a post-processing step to refine initial detections, verify the authenticity of GER events, and determine their onset and offset times across all channels. This adds complexity

to the workflow and may introduce additional sources of error. Additionally, there is a need for models that more accurately extract the proximal extents and capture the dynamics of GER across all channels.

To address these challenges, future research could focus on developing end-to-end models that inherently capture the dynamics and spatial correlations of GER events, eliminating the need for post-processing. Recent advances, such as Dense Associative Networks⁴⁶ or Dense Attention Mechanisms^{47,48}, offer promising solutions for this purpose.

Additionally, our analysis revealed that the statistical distributions of GER and non-GER regions resemble Laplace and normal distributions, respectively. Leveraging this insight, future directions could include the development of novel knowledge distillation techniques⁴⁹, to transfer knowledge from complex, high-capacity models to smaller, more efficient models suitable for clinical deployment.

Data availability

The datasets generated and analyzed is available at <https://github.com/azrarasouli/24Hr-Multichannel-Intraluminal-Impedance-Dataset>. The dataset is also provided in³³.

Received: 23 February 2025; Accepted: 18 September 2025

Published online: 23 October 2025

References

- Altay, D., Ozkan, T. B., Ozgur, T. & Sahin, N. U. Multichannel intraluminal impedance and pH monitoring are complementary methods in the diagnosis of gastroesophageal reflux disease in children. *Eurasian J. Med.* **54** (1), 22–26. <https://doi.org/10.5152/eurasianjmed.2022.20265> (2022).
- Frazzoni, M. et al. Impedance-pH monitoring for diagnosis of reflux disease: new perspectives. *Dig. Dis. Sci.* **62**, 1881–1889 (2017).
- Gyawali, C. P. et al. Updates to the modern diagnosis of GERD: Lyon consensus 2.0. *Gut* **73**, 361–371. <https://doi.org/10.1136/gutjnl-2023-330616> (2024).
- Katz, P. O., Gerson, L. B. & Vela, M. F. Guidelines for the diagnosis and management of gastroesophageal reflux disease. *Off J. Am. Coll. Gastroenterol. ACG.* **108** (3), 308–328 (2013).
- Gyawali, C. P. et al. Modern diagnosis of GERD: the Lyon consensus. *Gut* **67** (7), 1351–1362 (2018).
- Krause, A. & Yadlapati, R. Reflux testing: wireless pH, impedance-pH, and mucosal impedance. *Gastrointest Endosc. Clin.*, **35**(3), 587–601 (2025).
- Sifrim, D., Castell, D., Dent, J. & Kahrilas, P. J. Gastro-oesophageal reflux monitoring: review and consensus report on detection and definitions of acid, non-acid, and gas reflux. *Gut* **53** (7), 1024–1031 (2004).
- Roman, S. et al. Nov., Ambulatory 24-h oesophageal impedance-pH recordings: reliability of automatic analysis for gastro-oesophageal reflux assessment. *Neurogastroenterol. Motil.* **18** (11), 978–986 <https://doi.org/10.1111/j.1365-2982.2006.00825.x> (2006).
- Gyawali, C. P. et al. Inter-reviewer variability in interpretation of pH-impedance studies: the wingate consensus. *Clin. Gastroenterol. Hepatol.* **19** (9), 1976–1978. <https://doi.org/10.1016/j.cgh.2020.09.002> (2021).
- Kenari, A. R. et al. A multichannel intraluminal impedance gastroesophageal reflux characterization algorithm based on sparse representation. *IEEE J. Biomed. Heal Inf.* **25** (9), 3576–3586 (2021).
- Rasouli, A. et al. Liquid gastroesophageal reflux characterization by investigating multichannel intraluminal impedance-pH monitoring data. In: *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 4636–4639 (2019).
- Fu, J. et al. A cascade deep learning model for diagnosing pharyngeal acid reflux episodes using hypopharyngeal multichannel intraluminal Impedance-pH signals. *Intell. Med.* **8**, 100131. <https://doi.org/10.1016/j.ibmed.2023.100131> (2023).
- Zhou, M. J. et al. Development and validation of a machine learning system to identify reflux events in esophageal 24-hour pH/impedance studies. *Clin. Transl. Gastroenterol.*, **14**(10), e00634 (2022).
- Felsenreich, D. M. et al. Update on esophageal function, acid and non-acid reflux after one-anastomosis gastric bypass (OAGB): high-resolution manometry, impedance-24-h pH-metry, and gastroscopy in a prospective mid-term study. *Surg. Endosc.* <https://doi.org/10.1007/s00464-025-11606-7> (2025).
- Geeratragool, T. et al. Association between laryngopharyngeal reflux clinical scores and esophageal multichannel intraluminal impedance pH monitoring interpretation according to Lyon consensus 2.0. *Dis. Esophagus.* **38** (1), doae098 (2025).
- Le Berre, C. et al. Application of artificial intelligence to gastroenterology and hepatology. *Gastroenterology* **158** (1), 76–94. <https://doi.org/10.1053/j.gastro.2019.08.058> (2020). e2.
- Yang, Y. J. & Bang, C. S. Application of artificial intelligence in gastroenterology. *World J. Gastroenterol.* **25** (14), 1666–1683. <https://doi.org/10.3748/wjg.v25.114.1666> (2019).
- Kröner, P. T. et al. Artificial intelligence in gastroenterology: A state-of-the-art review. *World J. Gastroenterol.* **27** (40), 6794–6824. <https://doi.org/10.3748/wjg.v27.i40.6794> (2021).
- Ruffle, J. K., Farmer, A. D. & Aziz, Q. Artificial intelligence-assisted gastroenterology— promises and pitfalls. *Off. J. Am. Coll. Gastroenterol. | ACG* **114** (3) (2019).
- Christou, C. D. & Tsoulfas, G. Challenges and opportunities in the application of artificial intelligence in gastroenterology and hepatology. *World J. Gastroenterol.* **27** (37), 6191–6223. <https://doi.org/10.3748/wjg.v27.i37.6191> (2021).
- Parinitha, M. S., Doddawad, V. G., Kalgeri, S. H., Gowda, S. S. & Patil, S. Impact of artificial intelligence in endodontics: precision, predictions, and prospects. *J. Med. Signals Sens.* **14** (9) (2024).
- Amiri, M., Ranjbar, M. & Mohammadi, G. F. Automatic diagnosis of COVID-19 pneumonia using artificial intelligence deep learning algorithm based on lung computed tomography images. *J. Med. Signals Sens.* **13** (2) (2023).
- Doğan, Y. & Bor, S. Computer-based intelligent solutions for the diagnosis of gastroesophageal reflux disease phenotypes and chicago classification 3.0. *Healthcare (Basel, Switzerland)* **11** (12) <https://doi.org/10.3390/healthcare11121790> (2023).
- Ullah, Z., Usman, M., Latif, S., Khan, A. & Gwak, J. SSMD-UNet: semi-supervised multi-task decoders network for diabetic retinopathy segmentation. *Sci. Rep.* **13** (1), 9087. <https://doi.org/10.1038/s41598-023-36311-0> (2023).
- Ullah, Z., Usman, M. & Gwak, J. Multi-task semi-supervised adversarial autoencoding for COVID-19 detection based on chest X-ray images. *Expert Syst. Appl.* **216**, 119475 (2023).
- Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention-MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, proceedings, part III* 18 234–241 (2015).
- Jang, J., Park, S., Kim, J. K., An, J. & Jung, S. Cnn-based two step r peak detection method: Combining segmentation and regression. In *44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* 1910–1914 (2022).
- Tapotee, M. I., Saha, P., Mahmud, S., Alqahtani, A. & Chowdhury, M. E. H. M2ECG: wearable mechanocardiograms to electrocardiogram Estimation using deep learning. *IEEE Access* **12**, 12963–12975 (2024).

29. Guo, Z., Ding, C., Hu, X. & Rudin, C. A supervised machine learning semantic segmentation approach for detecting artifacts in plethysmography signals from wearables. *Physiol. Meas.* **42** (12), 125003 (2021).
30. He, Y. et al. Research on segmentation and classification of heart sound signals based on deep learning. *Appl. Sci.* **11** (2), 651 (2021).
31. Rasouli, A. et al. Reflux definitions in esophageal multi-channel intraluminal impedance. *Gastroenterol. Hepatol. Bed Bench.* **16** (4), 408 (2023).
32. Gell-Mann, M. & Lloyd, S. Information measures, effective complexity, and total information. *Complexity* **2** (1), 44–52 (1996).
33. MII dataset. <https://misp.mui.ac.ir/en/mii-data-0>
34. Patel, A., Wang, P. A., Sainani, D. & Sayuk, N. GS, Gyawali CP Distal mean nocturnal baseline impedance on pH-impedance monitoring predicts reflux burden and symptomatic outcome in gastro-oesophageal reflux disease. *Aliment Pharmacol. Ther.* **44**, 890–898 (2016).
35. Zhu, B. et al. Review of phonocardiogram signal analysis: insights from the PhysioNet/CinC challenge 2016 database. *Electronics* **13** (16), 3222 (2024).
36. Shaikhina, T. & Khovanova, N. A. Handling limited datasets with neural networks in medical applications: a small-data approach. *Artif. Intell. Med.* **75**, 51–63 (2017).
37. Du, B. et al. Stacked convolutional denoising auto-encoders for feature representation. *IEEE Trans. Cybern.* **47** (4), 1017–1027 (2016).
38. Chen, Y., Sun, Y., Lv, J., Jia, B. & Huang, X. End-to-end heart sound segmentation using deep convolutional recurrent network. *Complex. Intell. Syst.* **7**, 2103–2117 (2021).
39. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016).
40. Grossberg, S. Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, **52**(3), 213–257 (1973).
41. Bertels, J. et al. Optimizing the dice score and jaccard index for medical image segmentation: theory and practice. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, Proceedings, Part II* 92–100 (2019).
42. Moskalenko, V., Zolotykh, N. & Osipov, G. Deep learning for ECG segmentation. In *Advances in Neural Computation, Machine Learning, and Cognitive Research III: Selected Papers from the XXI International Conference on Neuroinformatics, October 7–11 246–254 (Dolgoprudny, Moscow Region, 2019).*
43. Chiarella, C., He, X. Z. & Hommes, C. A dynamic analysis of moving average rules. *J. Econ. Dyn. Control.* **30**, 9–10 (2006).
44. Bhutada, S., Yashwanth, N., Dheeraj, P. & Shekar, K. Opening and closing in morphological image processing. *World J. Adv. Res. Rev.* **14** (3), 687–695 (2022).
45. Ning, X. & Selesnick, I. W. ECG enhancement and QRS detection based on sparse derivatives. *Biomed. Signal. Process. Control.* **8** (6), 713–723 (2013).
46. Ullah, Z. & Kim, J. DAM-Seg: Anatomically accurate cardiac segmentation using Dense Associative Networks. arXiv preprint [arXiv:2502.15128](https://arxiv.org/abs/2502.15128) (2025).
47. Ullah, Z., Usman, M., Latif, S., & Gwak, J. Densely attention mechanism based network for COVID-19 detection in chest Xrays. *Sci Rep.* **13**(1), 261 (2023).
48. Ullah, Z., Usman, M., Jeon, M., & Gwak, J. Cascade multiscale residual attention CNNs with adaptive ROI for automatic braintumor segmentation. *Information Sciences*, **608**, 1541–1556 (2022).
49. Ahmad, s., Ullah, Z., & Gwak, J. Multi-teacher cross-modal distillation with cooperative deep supervision fusion learning for unimodal segmentation. *Knowledge-Based Systems*, **297**, 111854 (2024).

Acknowledgements

The authors gratefully acknowledge the financial support provided by the National Institute for Medical Research Development (NIMAD) of Iran (Project 4002280).

Author contributions

A.R.K designed/implemented the final method and wrote the main manuscript. H.R. modified the main method and evaluated the final results. Both authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to H.R.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025