Human-in-the-Loop Approach Based on MRI and ECG for Healthcare Diagnosis

Pavlo Radiuk[a], Oleksii Kovalchuk[a], Vitalii Slobodzian[a], Eduard Manziuk[a], Oleksander Barmak[a], Iurii Krak[b,c]

[a] Khmelnytskyi National University, 11, Institutes str., Khmelnytskyi, 29016, Ukraine
[b] Taras Shevchenko National University of Kyiv, 64/13, Volodymyrska str., Kyiv, 01601, Ukraine
[c] Glushkov Cybernetics Institute, 40, Glushkov ave., Kyiv, 03187, Ukraine

Abstract

The presented study investigates a human-centric approach to implementing human-in-the-loop models for healthcare diagnostics. The following tasks were considered and addressed in this work: a) identify the features necessary for future healthcare diagnosis based on electrocardiogram signals in the human-in-the-loop model: *P*, *T*-peaks, *QRS*-complex, *PQ* and *ST* segments, and b) detect inflammatory processes in the heart muscle (myocardium) based on cardiac magnetic resonance imaging. As a result of our investigation, a novel approach was proposed for embedding (integrating) clinical knowledge about the nature of these phenomena into the electrocardiogram signal and magnetic resonance imaging. Domain knowledge about the sample's nature is encoded similarly to the input information. Moreover, the convolution operation within our approach serves as an embedding mechanism. The results presented in the article are a starting point for using the models obtained by the proposed approach (human-in-the-loop models) for classification problems using deep learning and convolutional neural networks. Also, visual analysis shows the proposed approaches' ability to solve practical clinical problems. It also ensures transparent interpretation of the obtained results as the human-in-the-loop model, which, in turn, is built according to the human-centric approach. Overall, our contribution allows the implementation of a scheme for obtaining artificial intelligence solutions based on the principles of trust in them.

Keywords [1]

Human-centric approach, human-in-the-loop, trustworthiness in artificial intelligence, healthcare diagnosis, electrocardiogram, magnetic resonance imaging, autoencoder

1. Introduction

The acceleration of the development of information systems is accompanied by expanding the spheres of practical use. Information systems are taking on a new form in connection with integrating intelligent systems, which have the appearance of relatively simple algorithmic decision-making systems and artificial intelligence (AI) systems. Such systems are used in various subject domains, such as industry, education, transport, health care, and so forth [1]. Intelligent systems have considerably changed the life of society, processing a vast amount of data that is constantly growing. Intelligent systems demonstrate their effectiveness in applied tasks but, at the same time, become more complicated. The complexity of intelligent systems leads to their opacity in decision-making, which is an essential parameter of their introduction, especially in areas of critical use. Decisions made by AI

IDDM-2022: 5th International Conference on Informatics & Data-Driven Medicine, November 18–20, 2022, Lyon, France
EMAIL: radiukpavlo@gmail.com (P. Radiuk); losha.kovalchyk1998@gmail.com (O. Kovalchuk); vitalii.slobodzian@gmail.com (V. Slobodzian); eduard.em.km@gmail.com (E. Manziuk); alexander.barmak@gmail.com (O. Barmak); yuri.krak@gmail.com (I. Krak)
ORCID: 0000-0003-3609-112X (P. Radiuk); 0000-0001-9828-0941 (O. Kovalchuk); 0000-0001-8897-0869 (V. Slobodzian); 0000-0002-7310-2126 (E. Manziuk); 0000-0003-0739-9678 (O. Barmak); 0000-0002-8043-0785 (I. Krak)

systems depend on many parameters and are difficult to interpret. AI systems take the form of a black box in which the decision-making mechanisms are opaque, incomprehensible, possibly incorrect, and potentially dangerous. Today, there are known cases of incorrect and dangerous decisions made by artificial intelligence [2]. For example, accidents caused by autopilot cars hitting pedestrians are biased towards a particular category of people in hiring and others. Such examples indicate the need to develop AI systems that meet specific requirements for building socially responsible intelligent systems. This suggests that the simple application of artificial intelligence based only on technical performance characteristics in classification or clustering tasks is currently insufficient. It is necessary to expand the range of AI systems and specify them according to the specifics of practical applications.

Such manifestations of intelligent systems in tasks of practical importance necessitated the development of normative documents on the regulation of requirements and limitations of using AI to ensure safety, security, prevention of harm, etc. Guidelines for the development and use of AI are proposed. The Alliance for Artificial Intelligence of the European Union proposed ethical principles and frameworks for the management, development, and use of AI [3]. The General Data Protection Regulation (GDPR) [4] was adopted, within which the user's right to receive an explanation regarding decisions obtained thanks to AI systems or generated by such systems autonomously is recognised. Several requirements have been formed that AI must meet, including fairness, compliance with legislation, transparency in decision-making, interpretability, confidentiality, accountability, and several others. The combination of these requirements allows the application of AI systems that are more secure and dependable. Today, significant attention is paid to AI, whose decisions are transparent, explanatory, and interpretable. The practical application of such systems must be controlled; people must clearly understand what solutions the system can generate, what impact they have, what possible consequences are generated from the generated solutions and their limitations.

AI healthcare systems belong to the field of practical use, concerning which all the necessary safety, reliability, and criticality requirements are applied. AI in the healthcare field is undoubtedly necessary and vital [5] and can significantly affect human health, improve work processes, and improve the quality and efficiency of medical care. However, the application of AI is not limited to improving efficiency within specific tasks. The healthcare domain is an area of critical decision-making responsibility. In addition, AI must be able to work with data that has inaccuracy, is incomplete, has data gaps, is incorrect, erroneous, limited, and insufficient. It is not always possible to use AI systems, which by their characteristics, correspond to the black box, although they give the best results in terms of quality indicators. AI systems provide reliable solutions and can be practically applied [6]. Although AI systems are optimistic about changes in the healthcare field, there are significant caveats regarding their use in the healthcare field in responsible decisions. These issues follow from the following circumstances:

- AI systems at today's level of development generate a certain number of incorrect decisions, with a general, high-quality level of the received decisions.
- The developed systems do not provide an opportunity to determine which decision was wrong in each case but give a general assessment of the quality of a set of decisions.
- AI systems that meet the characteristics of a black box do not make it possible to determine based on which features they reached such a specific decision.

The presented study proposes the use of AI in diagnosing clinical diseases, considering the human-centric approach and the human-in-the-loop model. This application of AI intelligence allows the transformation of the information field of its practical use. Integrated transformation bridges the gap between theoretical AI research and its practical application in the healthcare field with the development of medical AI. The objective circumstances of practical application necessitate the creation of AI systems that consider ethical aspects, comply with legal regulations, and build trust.

Consequently, the contribution of this work is presented in the following aspects.

- Implementation of human-centric approach and human-in-the-loop models for healthcare diagnostics for analysis tasks: a) electrocardiogram (ECG) signals to intend to identify features necessary for further diagnosis: *P*, *T*-peaks, *QRS*-complex, *PQ* segments and *ST*; and b) MRI images of the heart for detected inflammatory processes in the heart muscle (myocardium).
- An approach for embedding (integration) human knowledge about the nature of these phenomena into the ECG signal and MRI image; as an embedding mechanism, it is proposed to use the convolution operation.

- Visual analysis of the ability of the proposed approaches to solve the tasks.

The structure of the article is as follows: in section 2, an overview of sources that consider the set of requirements for trust in AI systems is given; in section 3, the approaches proposed by the authors to the use of AI in the tasks of healthcare diagnostics are given, considering the human-in-the-loop model and human-centric approach; Chapter 4 presents the results of research on the integration of doctors' knowledge into the process of obtaining AI solutions.

2. Related work

The problem of trust in AI systems became relevant due to the acceleration of practical implementation and revealed new aspects that need to be paid attention to, which in some places become the main reasons for the impossibility of using AI. New aspects of the use of AI not only go beyond the technical difficulties of building AI but also create new directions and varieties of AI. In healthcare, trust has two important directions [7]: social and technical. Socially, trust is an essential aspect of patient-doctor interaction. The patient comes to the doctor in a state different from ordinary life functionality and is vulnerable. In this state, the patient cannot help himself and is forced to seek external help.

On the other hand, patients will use the services and follow the instructions according to the doctor's recommendation if there is a trusting relationship with the doctor based on the level of maintenance. Trust in the doctor is not the last factor in the success of the treatment that has a therapeutic effect. However, the specified aspect of trust relates to the medical side of the patient-doctor interaction. The use of AI systems introduces a new aspect of patient-doctor interaction that can undermine general trust.

AI systems can yield decent results, but their level of trust is low, so they cannot be considered dependable. In those circumstances, patients will be forced to rely on AI systems to obtain final decisions of medical importance, which may lead to a decrease in trust in the clinical practice of patient-doctor interaction [8].

Several studies have been devoted to creating AI systems that meet the requirements of trust [9]-[11]. Studies [12] and [13] are dedicated to studying the concept of trust based on the definition of a set of ethical principles that AI can be considered trustworthy. Proposed metrics for assessing trust in AI using the explainability of the expert-in-the-loop system [14]. The metric defines the difference between the explanations provided by the AI system and those obtained thanks to experts based on their reasoning and experience. The metric can be applied to diverse groups of experts to determine their confidence level in their recommendations. It allows for reducing the concept of confidence to a quantitative number by determining the distance between AI explanations and expert explanations. In this case, trust is reduced to explanation and equates to these two concepts. According to this approach, trust is entirely determined by the explainability of the decision.

In recent years, there has been growing concern in the scientific community about the potential dangers of black-box algorithms used in various fields of human activity, including healthcare diagnostics. This observation narrows the practical application in those medical aspects when the trust and transparency of the obtained decisions are not essential and critical [15]. Since AI systems still give high results, their use is justified, but this is not enough for the normative service of doctors. The potential applications of AI become limited due to the low level of trust. As stated in work [16], one should accept AI as a black box, but it is necessary to gain experience in the interaction of doctors with bioinformatics to acquire the necessary skills and expertise. This will improve the quality of medical image analysis. Another way to address the black box issue is transforming the neural network's complex structure into an understandable linear polynomial form [17], which allows reliable interpretation of the result of healthcare classification. However, this way of implementing medical AI requires highly qualified doctors to acquire specific competencies in bioinformatics. So, such approaches may limit and slow the spread of AI and might be considered too costly.

An essential aspect of implementing medical AI is the availability of quality data for establishing decision-making models. In many cases, quality is the determining factor for developing effective and explainable AI, like cardiac MRI measurements and interpretation [18]. However, the availability of such data can be limited for valid reasons, and significant difficulties can accompany the collection of quality data. A significant amount of data in the healthcare field for training neural networks is focused on images. Synthesis algorithms are used to expand such data to obtain training data [19]. Examples of such data are clinical imaging data [20], electrocardiogram signals [21], electronic medical records [22], and so forth. In this aspect, trust in AI is recognised as the data generation necessary for distribution with compliance with restrictions and rules. The lack of data and its limitations is another area of development of reliable medical AI [15]. In particular, neural networks for their training require a large amount of clinical data to obtain high results.

The prospects for the use of medical AI are promising, and today AI is used to predict caries on images and the development of reliable AI, which allows the explanation of the reasons according to which the prediction was made [23], [24]. However, in these studies, the capabilities of AI are limited by the need to trust AI and explain the reasons for the prediction. To improve the explainability of AI, systems for evaluating the results of prediction on images are proposed, involving a person as an expert in the prediction process [25]. In order to achieve the required results of trust and reliability of AI, three research areas have been identified that must be combined to obtain the required result. According to the authors, the combination of neural networks and their predictions, graphical causal models and methods of verification and explanation is the path that plays a transformative role in bridging the gap between theoretical research and the practical application of AI in clinical medicine [26].

Many studies reveal the need to develop medical AI with a set of characteristics that can be considered dependable. According to the conducted analysis, it can be considered that the urgent need is not so much the result of forecasting, but the feature set according to which the AI generated the forecast. The form of obtaining the required features can be presented in different forms. AI can independently indicate those features that are decisive in the obtained forecasts. Furthermore, the doctor can provide AI with a feature set that, according to clinical recommendations, play a decisive role in healthcare diagnosis. In this case, the AI should be able to focus computational algorithms for obtaining decisions on the given feature set. At the same time, other available features are also considered by AI with the necessary weighting of the influence and the difference in values.

3. Methods and materials

To implement the principles of trust in the results of healthcare diagnostics obtained thanks to AI and within the framework of the human-in-the-loop model and human-centric approach, we propose to integrate the knowledge of doctors about these data into the input data of medical research (ECG signal and MRI image). The proposed approach may allow modelling and classifying features that are understandable for doctors and enable them to interpret the result obtained by the AI systems.

As of today, the most prominent results in medical image processing have been obtained using deep learning methods and means, in particular, convolutional neural networks (CNNs) [27]. The convolution operation, when applied to two functions, *f* and *g*, returns a third function that corresponds to the cross-correlation functions $f(x)$ and $g(-x)$. The operation of convolution can be interpreted as the “similarity” of one function to a mirrored and shifted copy of another [28]. The concept of convolution is generalised for functions defined on arbitrary measurement spaces and can be considered a special integral transformation. In the discrete case, the convolution corresponds to the sum *f* of values with coefficients corresponding to the shifted values g and defines as

$$(f * g)(x) = f(1)g(x-1) + f(2)g(x-2) + f(3)g(x-3) + \ldots \quad (1)$$

The critical point in (1) is that a square (rectangular) impulse convolution (rectangular function, rectangular impulse, rectangular window) is a triangular (trapezoidal) impulse (function) [29]. That is, placed synchronously, the input and rectangular signal, because of convolution, give a signal with more pronounced peaks (known as features) with the input signal. We suggest using the given convolution property as a mechanism for integrating knowledge about the nature of the signal (image).

It will look like integrating the knowledge about the ECG signal and the MRI image shown below.

3.1. Integration of knowledge into the ECG signal

Let us consider what knowledge of the subject area (domain) can be for the ECG signal (Fig. 1).

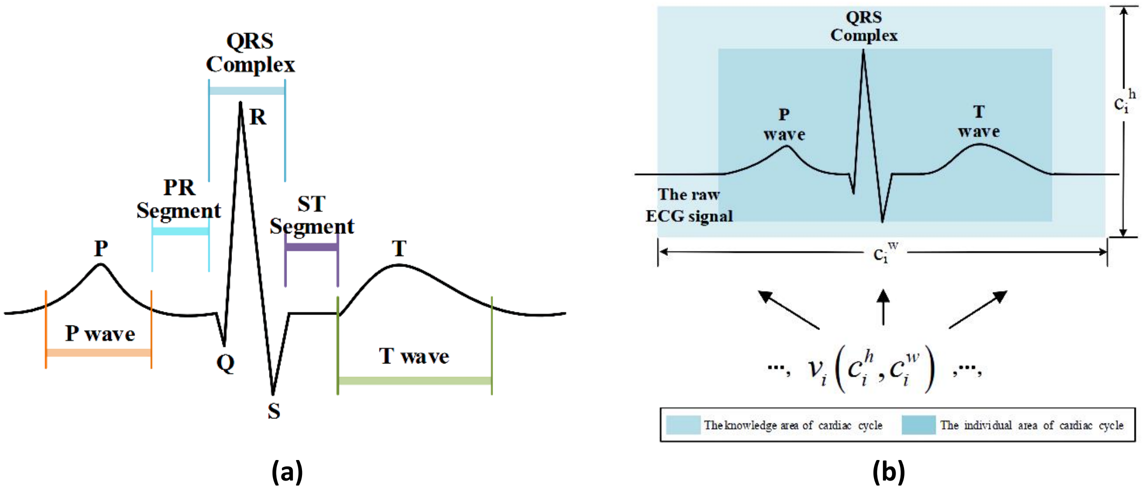
QRS
Complex
R
PR
Segment
ST
Segment
P
T
P wave
Q
T wave
S
QRS
Complex
P
wave
T
wave
The raw
ECG signal
c_i^h
c_i^w
$\dots, v_i(c_i^h, c_i^w), \dots,$
The knowledge area of cardiac cycle
The individual area of cardiac cycle
(a)
(b)

Figure 1: An illustrative sample of an ECG signal: (a) of a regular cardiac cycle; (b) of a cardiac cycle with individual feature knowledge implemented for an ECG signal [30]

Identifying feature points for ECG signals usually involves identifying the onset and offset of the *P* wave, the *QRS* complex, and the *T* wave. The higher amplitude of the *QRS* complex is frequently easily identified. Distinguishing *P* and *T* waves is tricky because their amplitudes are lower and sometimes accompanied by noise. Delineation of feature points (reference points) allows for more information, such as intervals and amplitudes, and provides essential information for further ECG analysis.

Domain knowledge is encoded in the form of the input signal. Since ECG signals $(s_1, s_2, \dots, s_n)$ are one-dimensional time series data, domain knowledge concerning possible pathologies in a medical image is encoded similarly.

Alternatively, knowledge of the ECG domain can be presented in Fig. 1b). For example, knowledge about the *P* wave is encoded as follows

$$\left(\dots,0,\underbrace{h_i,\dots,h_i}_{w_i},0,\dots\right). \tag{2}$$

The knowledge encoded as (2) can be represented as a rectangular pulse. Similarly, knowledge about the *QRS* complex and the *T* wave is encoded.

3.2. Integrating knowledge into MRI imaging

This subsection presents a trust AI model applicable to interpretive cardiovascular segmentation using multimodal MRI data. Healthcare professionals rarely use a multimodal approach in practice because labelling these many images is highly laborious and time-consuming. Meanwhile, annotations for images with larger slice thicknesses are more common and readily available, while images with thinner slice thicknesses are not. Thus, in this work, we propose a thickness-free multimodal image segmentation model that can be applied to both thick-slice and thin-slice images but only needs to annotate the thick-slice images during the training procedure.

Let us denote a set of images with thick slices as $I_C = \{(x_c, y_c) | x_c \in \mathbb{R}^{H \times W \times 3}, y_c \in \mathbb{R}^{H \times W}\}$ and with thin slices as $I_P = \{x_p | x_p \in \mathbb{R}^{H \times W \times 3}\}$. The proposed model uses unlabeled thin-slice images I_P to minimise the gap in model performance between thick and thin-slice images. In other words, this approach applies domain knowledge from one modality to image segmentation from another modality, resulting in trustful AI.

Here, a CNN architecture of the encoder-decoder type was used to segment medical images. We used the vanilla U-Net architecture and replaced the original coder with a pre-trained ResNet-50 [29]. ResNet-50 was utilised as it better represents the features of the input images. The proposed decoder uses subpixel convolution to construct segmentation results. Subpixel convolution is defined as

$$ConvSub^L = SP(W_L * F^{L-1} + b_L), \tag{3}$$

where operator $SP(\cdot)$ transforms a matrix of $H \times W \times D \times r^2$ into a matrix of $rH \times W \times D$, r is a scale factor for H, F^{L-1} and F^L stand for the input and output feature maps, W_L and b_L represent parameters of the sub-pixel convolution operators for layer L.

A multimodal procedure was used to train the CNN with (3), which provides joint optimisation for both types of images. The objective function of the proposed multimodal training is defined as

$$\mathcal{L}(x_c, x_p) = \mathcal{L}_c(q_c, y_c) + \ell\mathcal{L}_p(q_p), \tag{4}$$

where ℓ represents a hyperparameter for weighting the impact of $\mathcal{L}_c$ and $\mathcal{L}_p$, q_c and q_p stand for predictions of the segmentation probability maps of $rH \times W \times D$ for images with thick and thin slices, respectively. For (4), the cross-entropy loss is determined as

$$\mathcal{L}_c(q_c, y_c) = -\frac{1}{HWD}\sum_{n=1}^{HW}\sum_{d=1}^{D} y_c^{n,d} \ln q_c^{n,d}.$$

In the case of images with thin slices, $\mathcal{L}_p$ pushes the features away from the decision boundary of the feature distribution of thick-slice images, obtaining a flattening of the distribution.

3.3. Evaluation of the quality of the obtained results

Applying only a qualitative assessment of the obtained results at this research stage is possible. The purpose of these evaluations is to prove the capability of the proposed approaches for use in the given tasks. It is also proposed to visually evaluate changes in the signal with integrated knowledge concerning the input signal. Analyse how these changes affected the feature points.

For the task of MRI analysis, the segmentation quality of a network trained with multimodal datasets is evaluated through the Dice coefficient.

$$Dice = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \tag{5}$$

where, for the segmentation task, TP stands for true positive, TN – true negative, FP – false positive, and FN – false negative cases.

Considering the relationship between CNN performance and input samples remains unclear, a multi-layer perceptron was used to pick decomposed samples and their corresponding Dice scores for whole-space estimation. Such an approach can provide insight into Dice scores for individual regions of interest in latent space where no data are available. Consequently, it becomes possible to obtain information about the relationships between samples and their predictive ability by analysing the characteristics of samples in the hidden space. As a result, we can achieve an elevated level of trustfulness in the CNN model.

4. Results and discussion

4.1. Analysis of the ECG signal with integrated knowledge

Several experiments were conducted to evaluate the proposed mechanisms for integrating knowledge about a signal into a signal. The results of the experiments showed the ability of the proposed approach to solve the following tasks:

1. Clearer selection of signal features (*R*, *P*, *T*-peaks, *PR*-segment, *ST*-segment, *P*, *T*-waves).
2. Cleaning of «noise» in the signal for a more straightforward interpretation of the behaviour of *P*, *T*-waves.

The results of the conducted research can be visually evaluated in Fig. 2.

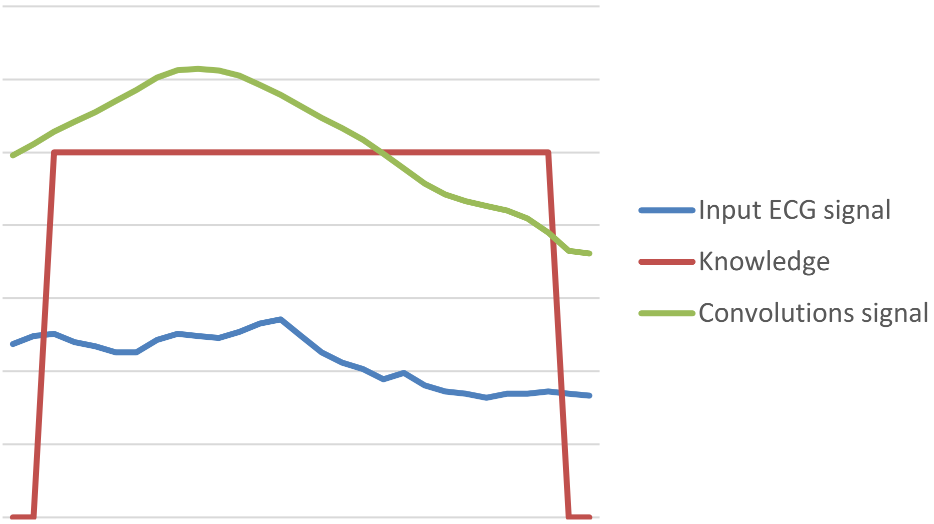

Figure 2: Input ECG signal (P wave fragment), signal knowledge, systolic signal

The picture shows a *P*-wave fragment. Applying the convolution operation to the given fragment resulted in a more apparent expression of the *P*-peak and wave behaviour.

The following steps were taken to incorporate signal knowledge into the input signal.

To incorporate knowledge, we need to match that knowledge with the corresponding ECG signals; for each cardiac cycle of the ECG, we use the position of the peak of the *R* wave as a reference point for matching knowledge; to find the peaks, we used the approach based on Shannon's entropy [31], as the one that gave the best results. The results of such incorporation are shown in Fig. 3.

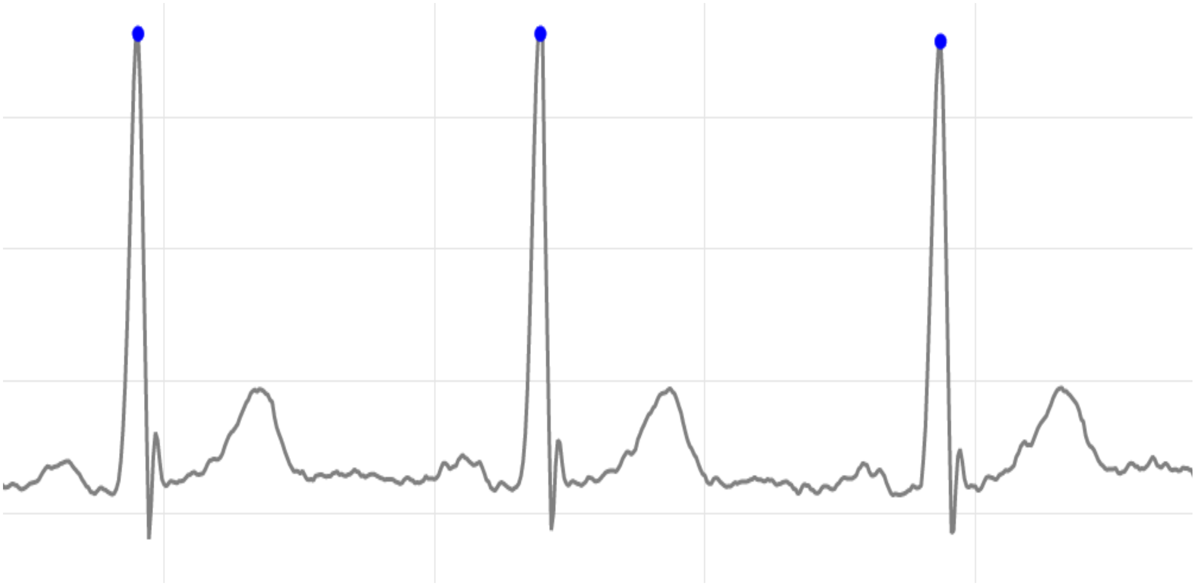

Figure 3: Detection of R-peaks in the ECG signal using Shannon entropy

Since ECG signals $(s_1, s_2, \dots, s_n)$ are one-dimensional time series data, we encode knowledge in the same form; additional three data channels (knowledge of *P*, *R* and *T* waves) are added to the primary input ECG signal (Fig. 4).

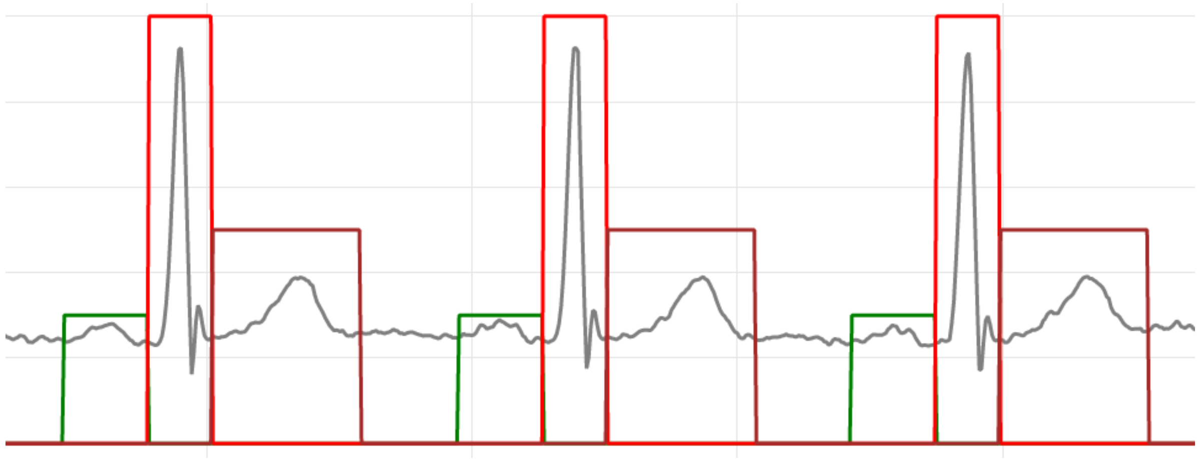

Figure 4: Channels of ECG signal and knowledge about ECG signal

The process of levelling knowledge includes three stages:

1. Alignment of the central point of the rectangular wave *R* with the identified reference point of the *R* peak.
2. Displacement of the control point of the *R* peak to the left by a fixed length (from the central point of the rectangular *P*-wave).
3. A shift of the control point of the *R* peak to the right by a fixed length (from the central point of the rectangular *T*-wave).

For domain knowledge, these three types of knowledge are encoded in the ECG signal data (*channel0*) as follows:

$$\begin{array}{llllllllllllllll}
channel0 & \cdots & s_k & s_{k+1} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & s_{k+m-1} & s_{k+m} \\
channel1 & \cdots & 0 & h_i^P & \cdots & h_i^P & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
channel2 & \cdots & 0 & 0 & 0 & 0 & 0 & h_i^R & \cdots & h_i^R & 0 & 0 & 0 & 0 & 0 & \cdots \\
channel3 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & h_i^T & \cdots & h_i^T & 0 & \cdots
\end{array}$$

After the encoding and knowledge matching is complete, we include this data as input to a neural network model of a hidden convolutional layer encoder-decoder [30]. The results obtained by an encoder are presented in Fig. 5.

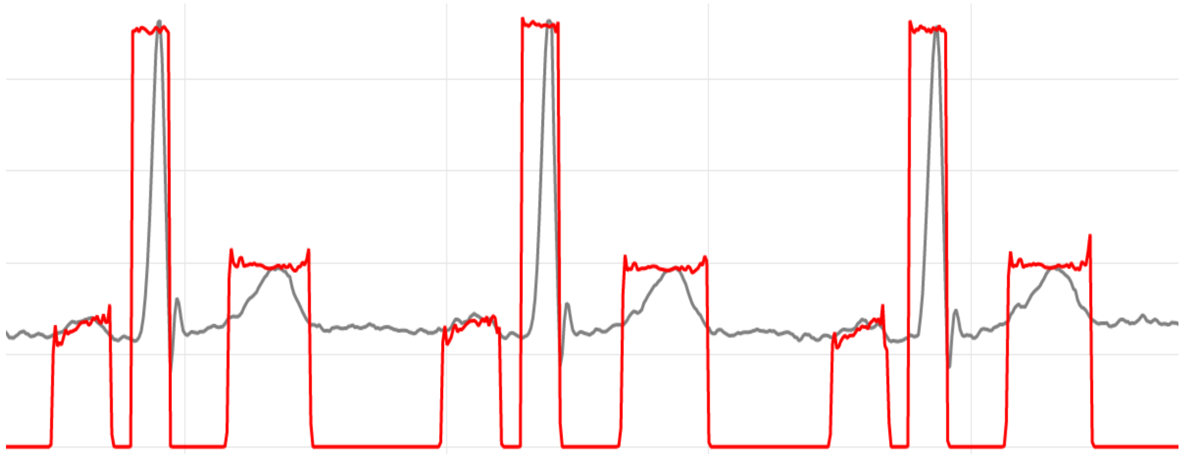

Figure 5: The output layer of the encoder-decoder neural network with a hidden convolutional layer; red lines are forecasting

According to Fig. 6, as a result of the operation of the encoder-decoder neural network, the *PQ* and *ST* segments and the width of the *QRS* complex are quite successfully selected.

The selection of *P*, *R*, and *T* peaks requires certain postprocessing, illustrated in Fig. 6.

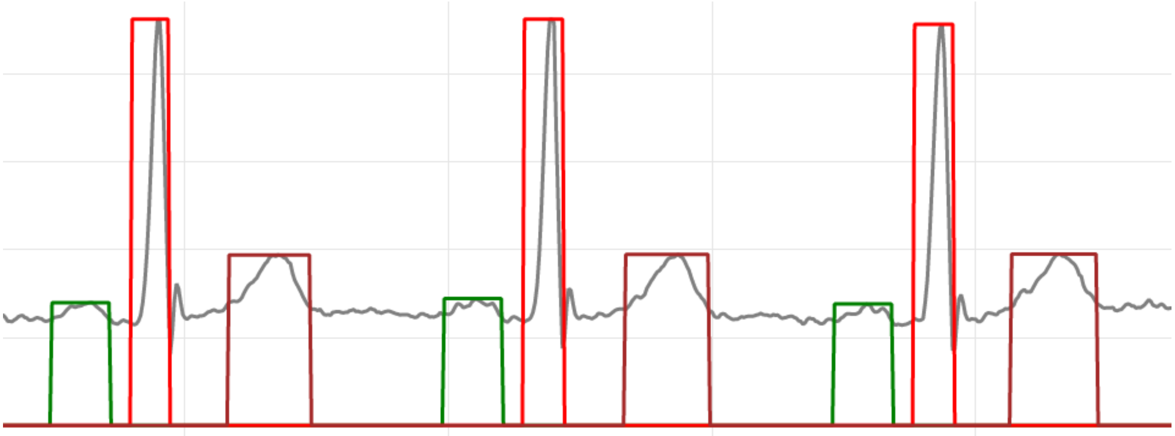

Figure 6: Postprocessing of the output layer of the encoder-decoder with the implemented knowledge

The *P*, *R*, and *T* peaks selected from the input image are shown in Fig. 7.

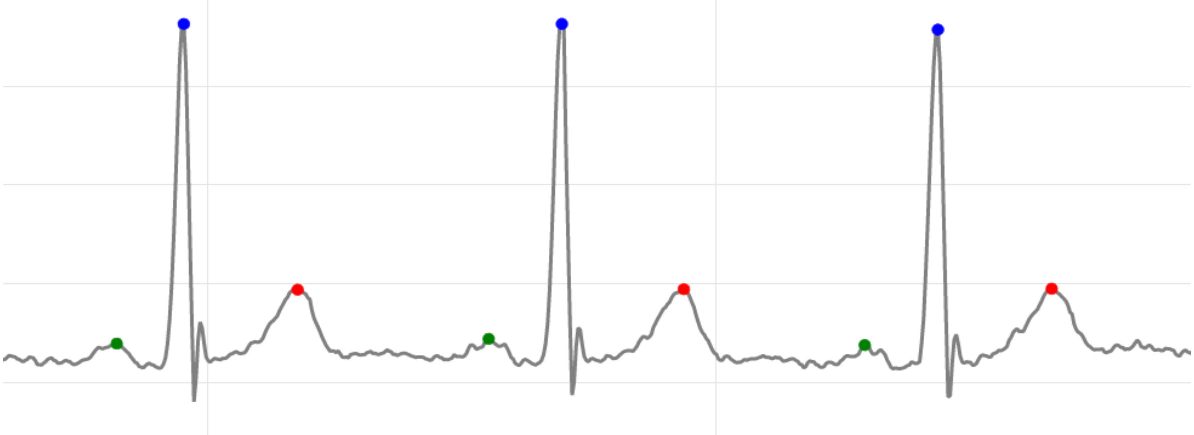

Figure 7: The result of the determination of *P*, *R*, and *T* peaks with the implemented knowledge

As can be seen from Fig. 5-7, the signal convoluted by the encoder-decoder neural network allows extracting the necessary information from the input ECG signal reliably: *P* and *T* peaks, *QRS*-complex, *PQ* and *ST* segments. The approach will not work only on signal sections where the *R*-peak is not detected. It is not critical because the specified areas are areas with artefacts, and they are removed from the analysis as they do not contain the necessary information.

4.2. Cardiac MRI studies with integrated knowledge

The dataset used for the experiments contained 1890 cardiac MRI samples excluded from 136 patients. The result of short-axis stack segmentation during the cardiac cycle with implemented domain knowledge is presented in Fig. 8.

As a result of computational experiments, it was found that the total percentage of unsuccessful segmentations obtained by the AI system reached 1.5% (that is, 28 unsuccessfully segmented images out of 1890). According to the domain knowledge, almost all failures were caused either by congenital heart diseases, such as a ventricular septal defect (Fig. 8a) or by visual artefacts and technical problems that affected image quality. At once, in 43 samples out of 1,890 (2.3%), segmentation errors were caused by a poor image of the apex of the heart (Fig. 8b).

The analysis of Dice score (5) demonstrated a decent correspondence between the CNN and manual LV and RV contours in both the internal and external test cohorts. The final values of the Dice coefficient for the internal cohort were obtained at 83,4-85,1% in the LV, while in the RV – 82,7-84,9%. Meanwhile, for the external cohort, the CNN model achieved 82,8-83% in the LV and 80,4-83,1% in the RV.

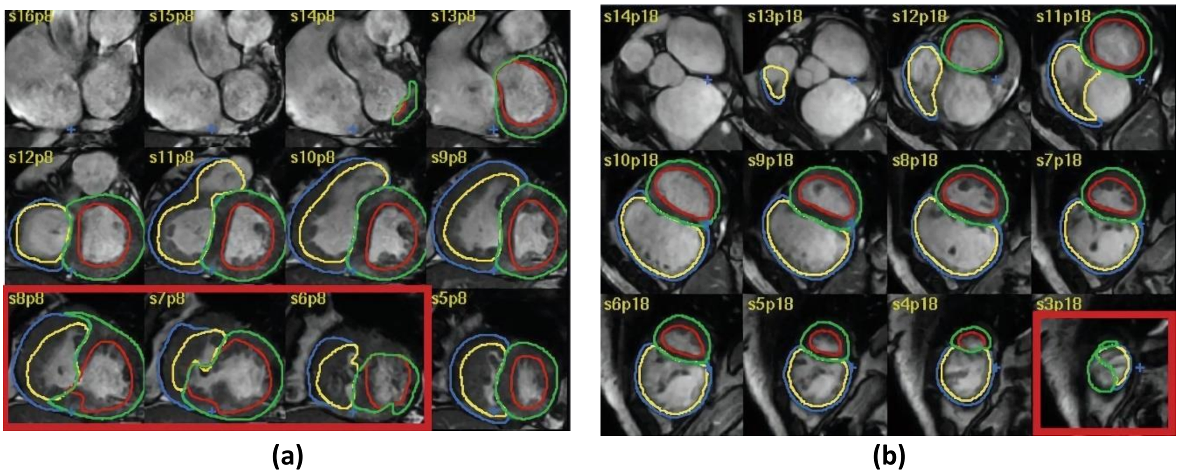

Figure 8: Instances of successful and unsuccessful segmentation by the AI system: (a) significant insufficiency due to congenital heart defect causing widening of the left ventricular (LV) contours in the right ventricular (RV) (red segment); (b) slight insufficiency at the apex, where the RV was incorrectly labelled as LV (red segment); red, green, blue and yellow ovals indicate the selection by the healthcare professionals of endocardial LV, epicardial LV, endocardial RV and epicardial RV lineaments

The analysis of the Dice coefficient demonstrated promising results for automatic segmentation conducted using AI with our human-in-the-loop approach. It is worth noting the constant differences in the automatic segmentation of the scan-re-scan cohort, for example, the exclusion of parts of the outflow tract of the RV (Fig. 8b). While this sequence maintained excellent repeatability, the Dice coefficient took smaller values (82,7-84,9% for the internal cohort and 80,4-83,1% for the external cohort).

Despite the promising results of the proposed human-in-the-loop approach, it has a limitation concerning the need for up-to-date databases for ECG signals and MRI, which contain more significant variability for a broader range of cardiac pathologies. Specifically for ECG, our approach based on the individual feature knowledge highly depends on R-wave peak detection when matching with ECG signals. As a result, automated delimitation may produce systematic errors in T-waves because the autoencoder predicts ups and downs as independent waves for very long T waves. For MRI, our approach fails to predict the junction region between the third and fourth ventricles because it is too small to be distinguished. In sum, the proposed human-in-the-loop approach is subjective to the domain knowledge and currently remains a proof of concept. The approach's performance might be improved by applying intelligent data techniques, such as further partial observation in large databases without annotations or through realistic simulations of ECG and MRI samples.

5. Conclusions and Future work

This study proposes a novel human-centric approach to healthcare diagnostics. Our contribution is based on resolving the following tasks: a) ECG signals to identify the features necessary for future diagnosis in the human-in-the-loop model: *P* and *T* peaks, *QRS*-complex, *PQ* and *ST* segments, and b) cardiac MRI for detected inflammatory processes in the heart muscle (myocardium). An approach is proposed for embedding human knowledge about the nature of these phenomena into a signal or an image. It is proposed to use the convolution operation as an embedding mechanism. For the problems under consideration, knowledge about the nature of the signal and image is encoded in the same form as the input information. The visual analysis revealed the ability of the proposed approaches to solve the problems under investigation. Moreover, experimental results on MRI demonstrated a decent correspondence between the CNN and manual LV and RV contours in both the internal and external test cohorts. Despite the promising results of the proposed human-in-the-loop approach, it has a limitation concerning the need for up-to-date databases for ECG signals and MRI, which contain more significant variability for a broader range of cardiac pathologies. In addition, our approach is subjective to the domain knowledge and remains proof of concept for now.

Further research will be directed to using models from the approach given in the article (human-in-the-loop models) for classification problems using convolutional neural networks and deep learning. A unique feature will be that such technology allows transparent interpretation of the obtained results in terms of the human-in-the-loop model, which, in turn, is built according to the human-centric approach. It might allow the implementation of a scheme for obtaining an AI solution based on the principles of trust.

6. References

[1] Z. Sun et al., A review of Earth artificial intelligence, Computers & Geosciences, vol. 159, p. 105034, Feb. 2022, doi:10.1016/j.cageo.2022.105034.

[2] Y. Duan, J. S. Edwards, and Y. K. Dwivedi, Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda, International Journal of Information Management, vol. 48, pp. 63–71, Oct. 2019, doi:10.1016/j.ijinfomgt.2019.01.021.

[3] N. A. Smuha, The EU approach to ethics guidelines for trustworthy artificial intelligence, Computer Law Review International, vol. 20, no. 4, pp. 97–106, Aug. 2019, doi:10.9785/cri-2019-200402.

[4] R. Chatila and J. C. Havens, The IEEE global initiative on ethics of autonomous and intelligent systems, in Robotics and Well-Being, vol. 95, M. I. Aldinhas Ferreira, J. Silva Sequeira, G. Singh Virk, M. O. Tokhi, and E. E. Kadar, Eds. Cham: Springer International Publishing, 2019, pp. 11–16. doi:10.1007/978-3-030-12524-0_2.

[5] S. Secinaro, D. Calandra, A. Secinaro, V. Muthurangu, and P. Biancone, The role of artificial intelligence in healthcare: A structured literature review, BMC Med Inform Decis Mak, vol. 21, no. 1, p. 125, Apr. 2021, doi:10.1186/s12911-021-01488-9.

[6] S. Keel, J. Wu, P. Y. Lee, J. Scheetz, and M. He, Visualising deep learning models for the detection of referable diabetic retinopathy and glaucoma, JAMA Ophthalmology, vol. 137, no. 3, pp. 288–292, Mar. 2019, doi:10.1001/jamaophthalmol.2018.6035.

[7] O. Asan and A. Choudhury, Research trends in artificial intelligence applications in human factors health care: Mapping review, JMIR Human Factors, vol. 8, no. 2, p. e28236, Jun. 2021, doi:10.2196/28236.

[8] J. J. Hatherley, Limits of trust in medical AI, Journal of Medical Ethics, vol. 46, no. 7, pp. 478–481, Jul. 2020, doi:10.1136/medethics-2019-105935.

[9] O. V. Barmak, Yu. V. Krak, and E. Manziuk, Characteristics for choice of models in the ansables classification, Problems in Programming, vol. 2–3, pp. 171–179, Jan. 2018, doi:10.15407/pp2018.02.171.

[10] E. Manziuk, Approach to creating an ensemble on a hierarchy of clusters using model decisions correlation, Electrotechnical Review, vol. 1, no. 9, pp. 110–115, Sep. 2020, doi:10.15199/48.2020.09.23.

[11] A. Singh, S. Sengupta, and V. Lakshminarayanan, Explainable deep learning models in medical image analysis, Journal of Imaging, vol. 6, no. 6, Art. no. 6, Jun. 2020, doi:10.3390/jimaging6060052.

[12] E. Manziuk, O. Barmak, I. Krak, O. Mazurets, and T. Skrypnyk, Formal model of trustworthy artificial intelligence based on standardisation, in Proceedings of the 2nd International Workshop on Intelligent Information Technologies & Systems of Information Security (IntelITSIS-2021), Khmelnytskyi, Ukraine, March 24–26, 2021, 2021, vol. 2853, pp. 190–197. [Online]. Available: http://ceur-ws.org/Vol-2853/short18.pdf

[13] E. Manziuk, I. Krak, O. Barmak, O. Mazurets, V. Kuznetsov, and O. Pylypiak, Structural alignment method of conceptual categories of ontology and formalised domain, in Proceedings of International Workshop of IT-professionals on Artificial Intelligence (ProfIT AI 2021), Kharkiv, Ukraine, September 20–21, 2021, Sep. 2021, vol. 3003, pp. 11–22. [Online]. Available: http://ceur-ws.org/Vol-3003/

[14] D. Kaur, S. Uslu, A. Durresi, S. Badve, and M. Dundar, Trustworthy explainability acceptance: A new metric to measure the trustworthiness of interpretable ai medical diagnostic systems, in Complex, Intelligent and Software Intensive Systems, Asan, Korea, July 1–3, 2021, 2021, vol. 278, pp. 35–46. doi:10.1007/978-3-030-79725-6_4.

[15] J. M. Durán and K. R. Jongsma, Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI, Journal of Medical Ethics, vol. 47, no. 5, pp. 329–335, May 2021, doi:10.1136/medethics-2020-106820.

[16] W. J. von Eschenbach, Transparency and the black box problem: Why we do not trust AI, Philos. Technol., vol. 34, no. 4, pp. 1607–1622, Dec. 2021, doi:10.1007/s13347-021-00477-0.

[17] I. Izonin, R. Tkachenko, N. Kryvinska, P. Tkachenko, and M. Greguš ml., Multiple linear regression based on coefficients identification using non-iterative sgtm neural-like structure, in Advances in Computational Intelligence, Gran Canaria, Spain, June 12-14, 2019, 2019, vol. 11506, pp. 467–479. doi:10.1007/978-3-030-20521-8_39.

[18] A. Janik, J. Dodd, G. Ifrim, K. Sankaran, and K. Curran, Interpretability of a deep learning model in the application of cardiac MRI segmentation with an ACDC challenge dataset, in Medical Imaging 2021: Image Processing, Feb. 2021, vol. 11596, pp. 861–872. doi:10.1117/12.2582227.

[19] G. Yang, Q. Ye, and J. Xia, Unbox the black-box for the medical explainable AI via multimodal and multi-centre data fusion: A mini-review, two showcases and beyond, Information Fusion, vol. 77, pp. 29–52, Jan. 2022, doi:10.1016/j.inffus.2021.07.016.

[20] D. B. Larson, D. C. Magnus, M. P. Lungren, N. H. Shah, and C. P. Langlotz, Ethics of using and sharing clinical imaging data for artificial intelligence: A proposed framework, Radiology, vol. 295, no. 3, pp. 675–682, Jun. 2020, doi:10.1148/radiol.2020192536.

[21] Y.-Y. Jo et al., Explainable artificial intelligence to detect atrial fibrillation using electrocardiogram, International Journal of Cardiology, vol. 328, pp. 104–110, Apr. 2021, doi:10.1016/j.ijcard.2020.11.053.

[22] J. Duell, X. Fan, B. Burnett, G. Aarts, and S.-M. Zhou, A comparison of explanations given by explainable artificial intelligence methods on analysing electronic health records, in 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Jul. 2021, vol. 2021, pp. 1–4. doi:10.1109/BHI50953.2021.9508618.

[23] N. Hasani et al., Trustworthy artificial intelligence in medical imaging, PET Clin, vol. 17, no. 1, pp. 1–12, Jan. 2022, doi:10.1016/j.cpet.2021.09.007.

[24] J. Ma et al., Towards trustworthy AI in dentistry, J Dent Res, vol. 101, no. 11, pp. 1263–1268, Oct. 2022, doi:10.1177/00220345221106086.

[25] D. Kaur, S. Uslu, and A. Durresi, Trustworthy AI explanations as an interface in medical diagnostic systems, in Advances in Network-Based Information Systems, Cham, 2022, vol. 526, pp. 119–130. doi:10.1007/978-3-031-14314-4_12.

[26] A. Holzinger et al., Information fusion as an integrative cross-cutting enabler to achieve robust, explainable, and trustworthy medical artificial intelligence, Information Fusion, vol. 79, pp. 263–278, Mar. 2022, doi:10.1016/j.inffus.2021.10.007.

[27] L. Wang et al., Trends in the application of deep learning networks in medical image analysis: Evolution between 2012 and 2020, European Journal of Radiology, vol. 146, p. 110069, Jan. 2022, doi:10.1016/j.ejrad.2021.110069.

[28] X. Liang et al., ECG_SegNet: An ECG delineation model based on the encoder-decoder structure, Computers in Biology and Medicine, vol. 145, p. 105445, Jun. 2022, doi:10.1016/j.compbiomed.2022.105445.

[29] I. Krak, O. Barmak, and P. Radiuk, Detection of early pneumonia on individual CT scans with dilated convolutions, in Proceedings of the 2nd International Workshop on Intelligent Information Technologies & Systems of Information Security (IntelITSIS-2021), Khmelnytskyi, Ukraine, March 24–26, 2021, 2021, vol. 2853, pp. 214–227. Accessed: May 09, 2021. [Online]. Available: http://ceur-ws.org/Vol-2853/

[30] J. Wang, R. Li, R. Li, and B. Fu, A knowledge-based deep learning method for ECG signal delineation, Future Generation Computer Systems, vol. 109, pp. 56–66, Aug. 2020, doi:10.1016/j.future.2020.02.068.

[31] S. Modak, L. Y. Taha, and E. Abdel-Raheem, A novel method of QRS detection using time and amplitude thresholds with statistical false peak elimination, IEEE Access, vol. 9, pp. 46079–46092, 2021, doi:10.1109/ACCESS.2021.3067179.