ISSN 2603-4697 (Online)
Volume 10 / Issue 1 / March 2026

Complex Control Systems

Bulgarian Academy of Sciences

Institute of Robotics

COMPLEX CONTROL SYSTEMS

Volume 10 / Issue 1 / March 2026

Istitute of Robotics - Bulgarian Academy of Sciences
Sofia, 2026

COMPLEX CONTROL SYSTEMS

Volume 10 / Issue 1 / March 2026

Institute of Robotics Bulgarian Academy of Sciences

ISSN: 2603-4697 (Online)

Volume 10, Issue 1, March 2026, Sofia

Table Of Contents

Complex Control Systems ISSN 2603-4697 (Online)

A Critical Review of Hybrid AI Architectures for Analysis and Forecasting of Cardio Time Series

Dobromir Slavov(1), Ekaterina Popovska (2), and (3) Galya Georgieva-Tsaneva

(1)(2)(3) Institute of Robotics, Bulgarian Academy of Science, 1113 Sofia, Bulgaria

***Abstract*— The analysis of cardio time series, including signals derived from the electrocardiogram, photoplethysmogram and heart rate variability, play a central role in the assessment of cardiac function and autonomic regulation. These signals are characterized by strong nonlinearity, nonstationarity and susceptibility to noise and artifacts. Such properties impose fundamental constraints on both signal analysis and forecasting and challenge the applicability of modern artificial intelligence methods.**

Recent advances in artificial intelligence have led to the widespread use of neural network architectures for cardio time series analysis. Recurrent neural networks, convolutional neural networks and attention-based models have demonstrated promising performance in various tasks, yet their effectiveness is often limited by data scarcity, inter-subject variability, noise sensitivity and limited physiological interpretability. Consequently, reported improvements in predictive accuracy frequently lack robustness and generalizability across datasets and acquisition conditions.

This paper presents a comprehensive and critical review of contemporary artificial intelligence architectures for short-term cardio time series analysis. Classical statistical, fractal and entropy-based methods are examined alongside modern neural approaches, emphasizing their complementary roles. Furthermore, a structured taxonomy of hybrid artificial intelligence models is introduced, combining data-driven learning with established signal analysis techniques to enhance robustness and interpretability. Rather than focusing solely on predictive performance, this review adopts a dynamics-oriented perspective, explicitly addressing the limits of short-term forecasting and outlining future research directions for hybrid and physiology-informed cardio data modeling.

***Keywords*— Artificial intelligence, cardio—signals, electrocardiogram, heart rate variability, hybrid models, photoplethysmogram, short-term forecasting, time series analysis.**

I. Introduction

Cardio time series derived from the electrocardiogram (ECG), photoplethysmogram (PPG) and heart rate variability (HRV) are widely used for the assessment of cardiac electrophysiology, hemodynamic activity and autonomic nervous system regulation. These signals play a key role in both clinical diagnostics and continuous monitoring applications, including wearable and ambulatory systems. However, cardio signals are inherently nonlinear, nonstationary and strongly affected by noise and artifacts, such as motion interference, baseline wander, ectopic beats and sensor-dependent distortions, which significantly complicates their analysis and interpretation [1], [2].

In many practical scenarios, cardio signals are available only over short observation windows, often ranging from a few seconds to one minute. Such short-term recordings impose fundamental limitations on both signal analysis and forecasting, as physiological variability and stochastic influences restrict the degree of temporal predictability [3]. Traditional approaches to cardio time series analysis have relied on statistical models, spectral methods and nonlinear dynamics descriptors. Autoregressive and state-space models have been used to capture short-term temporal dependencies, while entropy-based and fractal measures have provided valuable insights into signal complexity and autonomic regulation [4], [5]. Although these methods offer physiological interpretability and robustness in low-data regimes, they are limited in their ability to model complex temporal interactions and morphological patterns present in raw ECG and PPG signals.

Recent advances in artificial intelligence have led to the widespread adoption of deep learning architectures for time series analysis. Recurrent neural networks, including long short-term memory and gated recurrent unit models, have demonstrated strong capabilities in modeling temporal dependencies [17]–[19]. Convolutional neural networks have been successfully applied to morphological analysis of cardio signals and large-scale arrhythmia detection [21], [22], while attention-based and Transformer architectures have been proposed to capture global temporal context [23]–[25]. Despite promising results, the application of such models to short-term cardio signals remains challenging due to limited data availability, inter-subject variability, noise sensitivity and reduced physiological interpretability. Moreover, reported improvements in predictive accuracy often lack robustness and generalizability across datasets and acquisition conditions [22], [35].

To address these challenges, recent research has increasingly focused on hybrid modeling approaches that combine deep learning architectures with classical statistical, fractal and entropy-based methods. These hybrid models aim to leverage the representation learning capabilities of neural networks while preserving robustness, interpretability and physiological grounding. However, the existing literature lacks a unified and critical overview that relates modern artificial intelligence architectures and hybrid designs to the intrinsic dynamics of cardio time series, particularly in the context of short-term analysis and forecasting.

This paper presents a comprehensive and critical review of contemporary artificial intelligence architectures for cardio time series analysis, with a focus on short-term ECG, PPG and HRV signals. Rather than emphasizing predictive accuracy alone, the review adopts a dynamics-oriented perspective, highlighting the limits of short-term forecasting and the complementary roles of classical and data-driven methods. A structured taxonomy of hybrid artificial intelligence models is introduced and open challenges and future research directions are discussed.

II. CHARACTERISTICS OF CARDIO TIME SERIES

Cardio time series exhibit a set of intrinsic properties that fundamentally distinguish them from many other classes of temporal data. These properties arise from the underlying physiological mechanisms governing cardiac activity and autonomic regulation, as well as from the conditions under which signals are acquired. Understanding these characteristics is essential for selecting appropriate analysis and modeling approaches, particularly in short-term observation scenarios.

A. Signal Types and Physiological Origin

Electrocardiogram signals represent the electrical activity of the heart and are characterized by well-defined morphological components associated with cardiac depolarization and repolarization. Photoplethysmogram signals are optical measurements reflecting blood volume changes in peripheral tissue and are therefore indirectly related to cardiac activity, making them more sensitive to motion artifacts and sensor placement. Heart rate variability time series are derived signals, typically obtained from beat-to-beat intervals and primarily reflect autonomic nervous system modulation rather than direct cardiac morphology [1], [2].

The heterogeneous physiological origin of these signals leads to significant differences in sampling rates, noise sensitivity and information content. While electrocardiogram signals preserve detailed morphological structure, heart rate variability reflects temporal dynamics and regulatory mechanisms. Photoplethysmographic signals contain morphological information, but with increased sensitivity to external disturbances [3].

B. Nonlinearity, Nonstationarity and Multiscale Dynamics

A defining feature of cardio time series is their pronounced nonlinearity and nonstationarity. Cardiac dynamics are governed by complex feedback mechanisms involving the heart, autonomic nervous system and respiratory processes, resulting in temporal variations that cannot be adequately described by linear or stationary models [4]. Moreover, cardio signals exhibit multiscale behavior, with fast beat-to-beat fluctuations superimposed on slower regulatory trends related to autonomic modulation and circadian rhythms.

In short observation windows, typically below one minute, these multiscale characteristics become particularly challenging to capture. The limited number of samples restricts reliable estimation of long-range correlations and complicates the separation of deterministic structure from stochastic variability. As a result, many conventional assumptions underlying time series modeling, such as stationarity or ergodicity, are violated in practical cardio monitoring scenarios [5].

C. Noise, Artifacts and Data Limitations

Cardio time series are strongly affected by noise and artifacts arising from both physiological and technical sources. Motion artifacts, baseline wander, electrode displacement and sensor-dependent distortions are common in ambulatory and wearable recordings, particularly for photoplethysmogram signals. Errors in peak detection and ectopic beats can significantly distort derived time series and bias subsequent HRV analysis [6], [7].

Another critical limitation is the frequent availability of only short-duration recordings. Short time series constrain the applicability of data-intensive modeling approaches and reduce the statistical reliability of many nonlinear descriptors. These limitations highlight the importance of robustness, interpretability and data efficiency in cardio time series analysis and motivate the use of hybrid modeling strategies that combine classical signal analysis with modern artificial intelligence techniques [8].

III. CLASSICAL STATISTICAL, FRACTAL AND ENTROPY-BASED METHODS

Classical methods for cardio time series analysis have been extensively studied and remain fundamental for understanding cardiac dynamics, particularly in short-term and low-data scenarios. These approaches are grounded in statistical signal processing and nonlinear dynamics theory and often provide physiologically interpretable descriptors that complement modern data-driven models.

A. Statistical and Linear Modeling Approaches

Statistical models have traditionally been employed to characterize temporal dependencies in cardio signals. Autoregressive and autoregressive moving average models have been applied to electrocardiogram-derived and heart rate variability time series to capture short-term linear correlations and rhythmic patterns [9]. State-space formulations further extend this framework by allowing time-varying parameters and explicit modeling of hidden physiological states [10].

While such models are computationally efficient and interpretable, their applicability is limited by assumptions of linearity and stationarity, which are frequently violated in cardio signals. In short observation windows, parameter estimation becomes unreliable and model performance is highly sensitive to noise and artifacts. As a result, purely statistical models often serve as baselines or components within more complex hybrid frameworks rather than standalone solutions [11].

B. Fractal and Scaling Analysis

Fractal analysis has played a central role in the study of cardio dynamics, particularly in the analysis of heart rate variability. Methods such as detrended fluctuation analysis and scaling exponent estimation have been used to quantify long-range correlations and self-similar behavior in cardiac time series [12]. These approaches provide insight into the underlying regulatory mechanisms and have demonstrated sensitivity to pathological conditions and aging.

However, the reliability of fractal measures depends strongly on signal length and quality. In short-term recordings, the estimation of scaling exponents is often unstable and the distinction between deterministic structure and stochastic variability becomes blurred. Consequently, fractal analysis is most informative when used as a descriptive or comparative tool rather than as a basis for direct forecasting [13].

C. Entropy-Based Complexity Measures

Entropy-based methods are widely used to quantify the irregularity and complexity of cardio time series. Approximate entropy, sample entropy, permutation entropy and multiscale entropy have been applied to electrocardiogram and heart rate variability signals to assess autonomic regulation and system complexity [14], [15]. These measures are particularly attractive in short-term analysis due to their relative robustness to noise and minimal assumptions about signal structure.

Despite their advantages, entropy measures are sensitive to parameter selection and may yield biased estimates in the presence of artifacts or missing data. Moreover, entropy-based descriptors capture complexity rather than explicit temporal dynamics and therefore provide limited predictive capability when used in isolation. In recent studies, entropy measures have increasingly been incorporated as features within hybrid modeling frameworks, where they enhance interpretability and robustness while deep learning components handle representation learning and temporal modeling [16].

IV. AI ARCHITECTURES FOR TIME SERIES ANALYSIS

Recent advances in artificial intelligence have significantly expanded the set of tools available for cardio time series analysis. Deep learning architectures enable automatic representation learning from raw signals and have demonstrated strong performance across a variety of tasks. Nevertheless, their applicability to short-term cardio ~~data~~ is constrained by data availability, noise sensitivity and physiological interpretability. This section reviews the main classes of neural architectures used in cardio time series analysis, emphasizing their strengths and limitations.

A. Recurrent Neural Networks, Long Short-Term Memory and Gated Recurrent Units

Recurrent neural networks (RNNs) were among the first deep learning architectures applied to sequential biomedical data due to their ability to model temporal dependencies. Standard RNNs, however, suffer from vanishing and exploding gradient problems, which limit their capacity to capture long-range dependencies [17]. To address these issues, gated architectures such as long short-term memory (LSTM) networks and gated recurrent units (GRUs) introduce memory cells and gating mechanisms that regulate information flow over time [18], [19].

In cardio applications, LSTM and GRU models have been widely used for heart rate variability modeling, arrhythmia detection and short-term signal prediction. These architectures are particularly effective in capturing local temporal patterns and rhythm-related dynamics. However, their performance is strongly influenced by signal quality and window length. In short observation windows, recurrent models may overfit noise or device-specific artifacts and their hidden states often lack clear physiological interpretation. Moreover, training recurrent models can be computationally demanding and sensitive to hyperparameter selection [20].

B. Convolutional Neural Network-Based Models

Convolutional neural networks (CNNs) have become a dominant architecture for cardio signal analysis, particularly for electrocardiogram and photoplethysmogram data. One-dimensional convolutional layers are well suited for capturing local temporal patterns and morphological features, such as waveform shape and amplitude variations [21]. Multi-scale and dilated convolutional designs further enhance the ability of CNNs to model patterns across different temporal resolutions.

CNN-based models have demonstrated strong performance in tasks such as arrhythmia classification and signal quality assessment. Their feedforward structure allows for efficient training and parallelization, making them attractive for real-time and embedded applications. Nevertheless, CNNs inherently focus on local receptive fields and lack explicit mechanisms for modeling long-range temporal dependencies. As a result, CNNs alone may be insufficient for capturing global dynamics in cardio time series, particularly in tasks involving temporal evolution rather than static morphology [22].

C. Transformer and Attention-Based Architectures

Attention mechanisms and Transformer architectures have recently been introduced to time series analysis to address the limitations of recurrent models in capturing long-range dependencies. By computing pairwise attention weights across time steps, Transformers can model global temporal relationships without relying on sequential processing [23]. This property has motivated their application to cardio signals, including electrocardiogram classification and multivariate physiological monitoring [24].

Despite their theoretical advantages, Transformer-based models present several challenges in cardio applications. These architectures are typically data-intensive and may require large training datasets to generalize effectively. In short-term cardio recordings, the available temporal context is often insufficient to fully exploit global attention mechanisms. Furthermore, attention weights may capture spurious correlations related to noise or subject-specific patterns rather than meaningful physiological dynamics. Consequently, Transformers are often more effective as representation learners or components within hybrid architectures than as standalone models for short-term cardio forecasting [25].

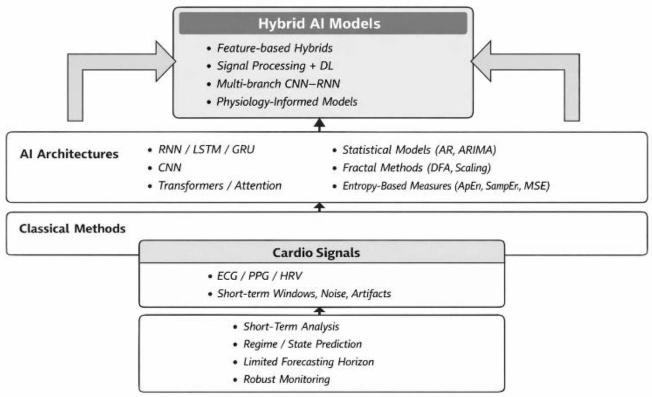


Fig. 1. *Conceptual taxonomy of classical, artificial intelligence, and hybrid modelling approaches for short-term analysis and forecasting of cardio time series.*

Figure 1 summarizes the relationships between classical methods, modern AI architectures and hybrid modeling approaches for short-term cardio time series analysis.

V. HYBRID AI MODELS FOR DYNAMIC ANALYSIS AND FORECASTING OF CARDIO SIGNALS

As illustrated in Fig. 1, hybrid AI models integrate classical signal analysis with neural architectures to address the limitations of short-term cardio data. Hybrid artificial intelligence models have emerged as a promising direction for cardio time series analysis, particularly in short-term and low-data regimes. These approaches aim to combine the representation learning capabilities of deep neural networks with the robustness, interpretability and physiological grounding of classical statistical, fractal and entropy-based methods.

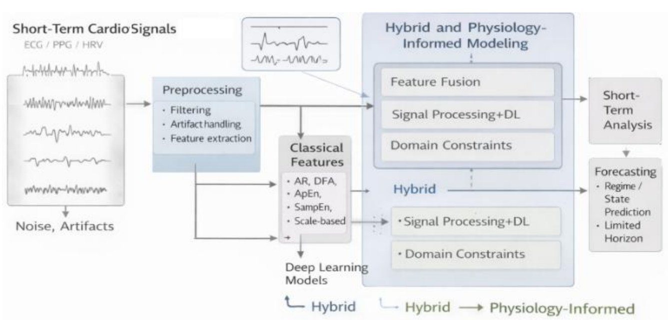


Fig. 2. *Conceptual pipeline of hybrid artificial intelligence models for short-term cardio signal analysis*

To clarify the integration of classical signal analysis and modern artificial intelligence techniques, Fig. 2 presents an end-to-end conceptual pipeline for short-term cardio time series analysis and forecasting. The diagram illustrates how preprocessing, handcrafted feature extraction, deep learning architectures and physiology-informed constraints are combined within hybrid modeling frameworks to address the limitations of short-term data and enhance robustness and interpretability. By integrating complementary modeling paradigms, hybrid architectures address key limitations of standalone data-driven or analytical approaches.

A. *Taxonomy and Design Principles*

Hybrid models for cardio signals can be broadly categorized according to the nature of integration between classical methods and neural architectures. A first category combines signal processing techniques, such as filtering, wavelet decomposition, or empirical mode decomposition, with deep learning models. In this design, classical preprocessing enhances signal quality and stabilizes subsequent neural representation learning, particularly for noisy electrocardiogram and photoplethysmogram data [26].

A second category consists of feature-based hybrid models, where handcrafted statistical, fractal and entropy-derived features are used as inputs to neural networks. This approach preserves physiological interpretability and reduces data requirements, while neural models capture nonlinear interactions among features [27]. Such designs are particularly effective for heart rate variability analysis, where derived features carry well-established clinical meaning.

Multi-branch hybrid architectures represent a third category, explicitly separating morphological and dynamical modeling. In these systems, convolutional neural networks are typically employed to extract morphological features from raw signals, while recurrent or attention-based modules capture temporal dependencies and regulatory dynamics [28]. This separation mirrors the physiological distinction between waveform morphology and autonomic modulation and has shown improved robustness in heterogeneous datasets.

Finally, physiology-informed hybrid models incorporate explicit constraints, priors, or regularization terms derived from cardio physiology. These constraints may encode heart rate limits, rhythm regularity, or known relationships between signal components. By embedding domain knowledge into the learning process, such models improve generalization and reduce the risk of learning spurious correlations [29].

B. *Comparison with Classical Approaches*

Compared to purely classical methods, hybrid AI models offer enhanced expressive power and adaptability to complex signal patterns. While statistical, fractal and entropy-based techniques provide stable and interpretable descriptors, they lack the capacity to automatically learn hierarchical representations from raw data. Hybrid approaches bridge this gap by augmenting classical descriptors with learned representations, resulting in improved robustness under noise and inter-subject variability [30].

At the same time, hybrid models mitigate several weaknesses of deep learning architectures. By incorporating classical features or constraints, they reduce data requirements and enhance interpretability, addressing key concerns in clinical and wearable applications. Empirical studies have shown that hybrid designs often outperform standalone neural models in short-term cardio analysis, particularly when training data are limited or signal quality varies across acquisition conditions [31].

Importantly, hybrid models do not aim to replace classical methods but rather to integrate them within a unified modeling framework. This integration enables complementary strengths: classical approaches provide physiologically meaningful structure, while neural components offer flexibility and nonlinear modeling capacity. As a result, hybrid architectures represent a pragmatic and theoretically grounded solution for short-term cardio time series analysis and forecasting.

VI. SHORT-TERM FORECASTING: LIMITS AND OPPORTUNITIES

Short-term forecasting of cardio time series represents one of the most debated and methodologically sensitive topics in cardiac signal analysis. While advances in artificial intelligence have enabled increasingly accurate modeling of complex temporal patterns, the intrinsic physiological variability of cardio systems imposes fundamental limits on predictability. A clear distinction between meaningful and ill-posed forecasting tasks is therefore essential.

A. What is Predictable in Short-Term Cardio Signals

In short observation windows, forecasting is generally feasible only over very limited horizons. Predictable components typically include local temporal continuity, short-term trends and regime transitions rather than exact point-wise signal values. For electrocardiogram and photoplethysmogram signals, short-horizon continuation of waveform morphology may be achievable under stable conditions, while for heart rate variability signals, limited predictability may exist for beat-to-beat intervals over one to a few future steps [32].

An alternative and often more meaningful formulation of forecasting focuses on state or regime prediction rather than precise value estimation. Examples include the anticipation of rhythm changes, detection of transitions between stable and unstable cardiac states, or prediction of increased variability regimes. Such formulations align more closely with physiological processes and clinical relevance and are less sensitive to noise and inter-subject variability [33].

B. Fundamental Limits of Predictability

Cardio dynamics arise from the interaction of multiple regulatory mechanisms operating across different time scales, including autonomic control, respiration and circadian influences. These interactions introduce stochastic components and nonlinear feedback that limit long-horizon predictability, even in the absence of measurement noise [34]. As a result, attempts to perform long-term point-wise forecasting of cardio signals are often ill-posed and may lead to misleading conclusions.

Moreover, short-term recordings exacerbate these limitations by restricting the available temporal context for model training and inference. In such settings, deep learning models may inadvertently learn subject-specific or device-specific signatures rather than underlying physiological dynamics. This phenomenon contributes to the lack of generalizability observed in many reported forecasting results [35].

C. Evaluation Protocols and Practical Considerations

Robust evaluation of short-term forecasting models requires careful selection of performance metrics and validation protocols. Traditional error-based metrics, such as mean absolute error or root mean square error, should be complemented by robustness analyses under noise, artifact contamination and cross-subject testing. Uncertainty-aware evaluation, including confidence intervals or probabilistic forecasts, provides additional insight into model reliability [36].

From a practical perspective, forecasting approaches must be aligned with application-specific objectives. In clinical and wearable monitoring systems, interpretability, robustness and reliability often outweigh marginal gains in numerical accuracy. Consequently, hybrid and physiology-informed models that emphasize stability and meaningful prediction horizons are better suited for real-world deployment than purely data-driven approaches optimized solely for short-term accuracy [37].

VII. OPEN CHALLENGES AND FUTURE DIRECTIONS

Despite significant progress in cardio —time series analysis, several open challenges remain that limit the widespread adoption and reliability of artificial intelligence-based approaches, particularly in short-term and real-world monitoring scenarios. Addressing these challenges is essential for advancing both methodological rigor and practical applicability.

One major challenge is the analysis of ultra-short time windows, often below thirty seconds, which are increasingly common in wearable and opportunistic monitoring. In such regimes, both classical and deep learning models suffer from insufficient temporal context, leading to unstable estimates and reduced generalizability. Developing data-efficient and uncertainty-aware models that can operate reliably under these constraints remains an open research problem [38].

Robustness to noise, artifacts and sensor heterogeneity represents another critical challenge. Cardio signals acquired from different devices, body locations and operating conditions exhibit substantial variability. Models trained on clean, curated datasets often fail to generalize to real-world data, where motion artifacts and signal quality fluctuations are prevalent. Future research should emphasize robustness-oriented training strategies, including noise-aware modeling, domain adaptation and explicit signal quality integration [39].

Physiology-informed and constrained learning constitutes a promising direction for improving interpretability and generalization. By embedding domain knowledge, such as heart rate limits, rhythm regularity, or known autonomic relationships, models can be guided toward physiologically plausible solutions. This approach reduces the risk of learning spurious correlations and enhances trustworthiness in clinical applications [40].

Finally, the integration of hybrid artificial intelligence models within digital twin frameworks for cardio monitoring represents an emerging research frontier. Digital twins combine data-driven models with mechanistic understanding to enable personalized analysis, simulation and decision support. Such frameworks have the potential to transform cardio monitoring from passive signal analysis to proactive, individualized health management [41].

The comparative analyses presented in Table 1 indicate that classical statistical and fractal methods maintain high interpretability but demonstrate limited robustness to noise in short observation windows. Deep neural networks, particularly CNN and LSTM architectures, achieve greater capacity for modelling complex temporal dependencies but suffer from low interpretability and increased sensitivity to input data quality. Hybrid AI approaches combine the strengths of both paradigms, providing improved noise robustness and a balanced level of interpretability, which makes them well-suited for real-time HRV analysis in practical monitoring applications.

Table 1. *Performance comparison of classical, deep learning and hybrid methods for short-term HRV analysis*

Method	Input type	Window length (s)	Robustness to noise	Interpretability
Statistical (AR)	RR intervals	30	Medium	High
Fractal (DFA)	HRV series	30	Medium	High
Entropy-based	HRV series	30	High	Medium
CNN	Raw ECG	30	Low	Low
LSTM	HRV series	30	Medium	Low
Hybrid (CNN + traditional)	**ECG + features**	**30**	**High**	**Medium–High**

Statistical and entropy-based methods, as summarized in Table 2, provide strong physiological interpretability, which supports their use in clinical analysis and expert-driven evaluation.

Table 2. *Interpretability and physiological relevance of different model classes*

Model type	Physiological interpretability	Clinical relevance	Suitability for digital twin
Statistical	High	Medium	Medium
Fractal / Entropy	High	High	Medium
Deep learning	Low	Medium	Low

Deep learning models exhibit limited inherent interpretability, posing barriers to their direct adoption in medical decision workflows despite their predictive capacity. Hybrid neural-analytical designs offer the most effective trade-off, jointly preserving physiological relevance, clinical applicability, and alignment with digital-twin paradigms for personalized cardio modelling.

VIII. CONCLUSION

This paper presented a comprehensive and critical review of contemporary artificial intelligence architectures for cardio time series analysis, with a focus on short-term electrocardiogram, photoplethysmogram and heart rate variability signals. The intrinsic characteristics of cardio signals, including nonlinearity, nonstationarity, noise sensitivity and limited observation windows, were discussed as fundamental factors shaping model performance and predictability.

Classical statistical, fractal and entropy-based methods were reviewed alongside modern deep learning architectures, highlighting their complementary strengths and limitations. While deep learning models offer powerful representation learning capabilities, their effectiveness is often constrained by data scarcity, noise sensitivity and limited physiological interpretability. Hybrid artificial intelligence models were identified as a pragmatic and theoretically grounded solution, integrating classical signal analysis with data-driven learning to enhance robustness, interpretability and generalization.

A dynamics-oriented perspective on short-term forecasting was emphasized, clarifying the limits of predictability in cardio signals and distinguishing meaningful forecasting tasks from ill-posed formulations. Finally, open challenges and future research directions were outlined, including ultra-short window analysis, robustness-aware modeling, physiology-informed learning and digital twin-based cardio monitoring.

Overall, this review underscores the importance of aligning artificial intelligence methodologies with the intrinsic dynamics and physiological constraints of cardio systems. Future progress in this field will depend not on increasingly complex models alone, but on the thoughtful integration of domain knowledge, robust modeling principles and hybrid architectures tailored to real-world cardio data.

ACKNOWLEDGMENT

This research was funded by the National Science Fund of Bulgaria (scientific project “Research on AI-based methods for modeling, analyzing and forecasting biomedical time series by applying a digital twin concept”), Grant Number КП-06-Н97/12, 10 December 2025.

REFERENCES

[1] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, “Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use”, Circulation, 1996.

[2] F. Shaffer and J. P. Ginsberg, “An overview of heart rate variability metrics and norms,” Frontiers in Public Health, vol. 5, pp. 1–17, 2017.

[3] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," Physiological Measurement, vol. 28, no. 3, pp. R1–R39, 2007.
[4] A. L. Goldberger, L. A. N. Amaral, J. M. Hausdorff, P. C. Ivanov, C.-K. Peng and H. E. Stanley, "Fractal dynamics in physiology: Alterations with disease and aging," Proc. Natl. Acad. Sci. USA, vol. 99, suppl. 1, pp. 2466–2472, 2002.
[5] G. D. Clifford, F. Azuaje and P. McSharry, Advanced Methods and Tools for ECG Data Analysis. London, U.K.: Artech House, 2006.
[6] D. E. Lake and J. R. Moorman, "Accurate estimation of entropy in very short physiological time series," Am. J. Physiol. Heart Circ. Physiol., vol. 300, no. 1, pp. H319–H329, 2011.
[7] G. D. Clifford, J. Behar, Q. Li and I. Rezek, "Signal quality indices and data fusion for determining clinical acceptability of electrocardiogram signals," Physiological Measurement, vol. 33, no. 9, pp. 1419–1433, 2012.
[8] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih and U. R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," Comput. Methods Programs Biomed., vol. 161, pp. 1–13, 2018.
[9] G. E. P. Box, G. M. Jenkins, G. C. Reinsel and G. M. Ljung, Time Series Analysis: Forecasting and Control, 5th ed. Hoboken, NJ, USA: Wiley, 2015.
[10] J. Durbin and S. J. Koopman, Time Series Analysis by State Space Methods, 2nd ed. Oxford, U.K.: Oxford Univ. Press, 2012.
[11] J. D. Hamilton, Time Series Analysis. Princeton, NJ, USA: Princeton Univ. Press, 1994.
[12] C.-K. Peng, S. Havlin, H. E. Stanley and A. L. Goldberger, "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series," Chaos, vol. 5, no. 1, pp. 82–87, 1995.
[13] J. W. Kantelhardt, E. Koscielny-Bunde, H. H. A. Rego, S. Havlin, A. Bunde, "Multifractal Detrended Fluctuation Analysis of Nonstationary Time Series", Physica A: Statistical Mechanics and its Applications, 2002.
[14] S. M. Pincus, "Approximate entropy as a measure of system complexity," Proc. Natl. Acad. Sci. USA, vol. 88, no. 6, pp. 2297–2301, 1991.
[15] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," Am. J. Physiol. Heart Circ. Physiol., vol. 278, no. 6, pp. H2039–H2049, 2000.
[16] M. Costa, A. L. Goldberger and C.-K. Peng, "Multiscale entropy analysis of biological signals," Phys. Rev. E, vol. 71, no. 2, pp. 1–18, 2005.
[17] Y. Bengio, P. Simard and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," IEEE Trans. Neural Netw., vol. 5, no. 2, pp. 157–166, 1994.
[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[19] K. Cho et al., "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 2014, pp. 1724–1734.
[20] Z. C. Lipton, J. Berkowitz and C. Elkan, "A critical review of recurrent neural networks for sequence learning," arXiv:1506.00019, 2015.
[21] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, M. Adam, "Automated Detection of Arrhythmias Using Different Intervals of Tachycardia ECG Segments with Convolutional Neural Network", Information Sciences, vol. 405, pp. 81–90, 2018. DOI: 2018.
[22] A. Y. Hannun et al., "Cardiologist-level arrhythmia detection with convolutional neural networks," Nat. Med., vol. 25, pp. 65–69, 2019.
[23] A. Vaswani et al., "Attention is all you need," in Proc. Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 2017, pp. 5998–6008.
[24] S. Li, X. Li, S. Wang, H. Zhang, "A Lightweight Transformer Model for ECG Classification", in Proc. IEEE Int. Conf. on Engineering in Medicine and Biology Society (EMBC), Berlin, 2019, pp. 1–4. DOI: 2019.
[25] T. Zhou, H. Qi, J. Sun, B. Xu, H. Jin, Q. Wu, "Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting", AAAI Conference on Artificial Intelligence, 2021.
[26] M. Elgendi, I. Norton, M. Brearley, D. Abbott and D. Schuurmans, "Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions," PLoS One, vol. 8, no. 10, pp. 1–11, 2013.
[27] A. Choi, M. T. Bahadori and J. Sun, "Doctor AI: Predicting clinical events via recurrent neural networks," in Proc. Machine Learning for Healthcare Conf., Los Angeles, CA, USA, 2016, pp. 301–318.
[28] O. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," Comput. Biol. Med., vol. 96, pp. 189–202, 2018.
[29] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, A. Y. Ng, "Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks", arXiv:1707.01836, 2017.
[30] O. Faust, U. R. Acharya, Y. Hagiwara, T. J. Hong and O. S. Lih, "Hybrid deep learning for ECG-based diagnosis," Biomed. Signal Process. Control, vol. 42, pp. 32–41, 2018.
[31] N. Y. Chen, Y. H. Liu, X. Q. Zhang, "An Automated HRV Classification Framework Using 1D-CNN and LSTM for Cardiac Abnormality Detection", IEEE Access, vol. 7, pp. 1–10, 2019. DOI: 2019.
[32] L. Glass, "Synchronization and rhythmic processes in physiology," Nature, vol. 410, no. 6825, pp. 277–284, 2001.
[33] A. Porta, S. Guzzetti, N. Montano, R. Furlan, M. Pagani, A. Malliani and S. Cerutti, "Entropy, entropy rate and pattern classification as tools to typify complexity in short heart period variability," IEEE Trans. Biomed. Eng., vol. 48, no. 11, pp. 1282–1291, 2001.
[34] J. W. Kantelhardt et al., "Multifractal detrended fluctuation analysis of nonstationary time series," Physica A, vol. 316, nos. 1–4, pp. 87–114, 2002.
[35] A. H. Ribeiro et al., "Automatic diagnosis of the 12-lead ECG using a deep neural network," Nat. Commun., vol. 11, pp. 1–9, 2020.
[36] T. Gneiting and M. Katzfuss, "Probabilistic forecasting," Annu. Rev. Stat. Appl., vol. 1, pp. 125–151, 2014.
[37] S. Nemati, M. M. Ghassemi and G. D. Clifford, "Optimal medication dosing from suboptimal clinical data: A deep reinforcement learning approach," in Proc. AAAI Conf. Artificial Intelligence, Phoenix, AZ, USA, 2016, pp. 297–305.
[38] D. E. Lake, J. S. Richman, M. P. Griffin and J. R. Moorman, "Sample entropy analysis of neonatal heart rate variability," Am. J. Physiol. Regul. Integr. Comp. Physiol., vol. 283, no. 3, pp. R789–R797, 2002.
[39] C. Liu, P. Li, X. Di, G. D. Clifford, "Robust Heart Rate Estimation From Photoplethysmography Signals Using Adaptive Filtering and an Error-Correction Strategy", IEEE Transactions on Biomedical Engineering, vol. 65, no. 4, pp. 968–977, 2018. DOI: 2018.
[40] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nat. Med., vol. 25, pp. 44–56, 2019.
[41] J. Corral-Acero et al., "The 'digital twin' to enable the vision of precision cardiology," Eur. Heart J., vol. 41, no. 48, pp. 4556–4564, 2020.

Robotics in healthcare and some of its applications

Ani Boneva
Institute of Information and Communication Technologies, Bulgarian Academy of Sciences
Sofia, Bulgaria
ani.boneva@iict.bas.bg

Veronika Ivanova
Institute of Robotics, Bulgaria Academy of Sciences
Sofia, Bulgaria
iwanowa.w@abv.bg

***Abstract*—In recent decades, medicine has undergone a profound transformation brought about by the advent of robotic systems. Robots have gradually been integrated and are changing the way the human body is diagnosed, treated and restored. Robotic technologies are becoming an essential part of modern healthcare – from the operating room to home rehabilitation and telemedicine. The paper suggests some applications associated with robotic technology and its response to medical challenges and leads to a new generation of robots that can meet the patient's needs. In particular, the paper presents a robotic modular platform that is suitable for medical training. Directions for future work are mentioned such as a device to monitor and adjust the level of CO2 in abdominal cavity.**

***Keywords*—Robotics, Technology, Device, Medicine, Healthcare**

I. Introduction

The development of science opens up new opportunities for the application of robotics. This development is accompanied by the use of new methods and approaches in the design and implementation of robotic systems, which use modern achievements in areas such as electronics, mechanics, information systems and artificial intelligence. With its achievements, robotics finds its place in medicine. Medical robotics includes systems and devices, which use automation, sensors, AI and precision mechanics in order to assist medical staff. The idea is not to replace doctors, they simply don't have a chance at this stage, but to give them super capabilities such as more accuracy, less fatigue, faster patient recovery [1, 2, 3].

Robotics application in healthcare includes:

1. Surgical Robots and surgical robot training One of them is Da Vinci [4].

Advantages microincisions, less bleeding, fast recovery.

These robots find their application in urology, gynecology, cardiac surgery, orthopedics

Modern trends in the field of minimally invasive surgical robotics are Single port access - access through one hole. : The method is gentle, although it is not suitable for emergency cases due to the time delay

2.-Exoskeletons and robotic walkways (Lokomat). They assist patients with stroke, spinal cord injuries, neuromuscular diseases

3-Bionics and Smart prostheses-AI- controlled prosthetic limbs and sensor-enhanced prostheses.

4 Assistive robots for Daly-living support-feeding , lifting ets.

5. Social and therapeutics robots, robots nurses and companionships:

- PARO (the seal robot) for dementia [5].
- Robots for children with autism — teach empathy, social skills materials.

6. Logistics robots – for delivering medications and other support materials. They deliver medications, samples, materials. They save time and reduce errors — plus they reduce crowding around staff.

7 Cleaning robots • perform autonomous robots that use UV-C to destroy pathogens in a hospital room. With the development of COVID-19, this class of work is given a serious voice.

8 . Telepresence and telemedicine robots.

Mobile platforms with cameras and sensors that allow a doctor to remotely examine a patient. Part of the new trends — they are increasingly combined with IoT [6] and wearable devices.

Pill robots -Pill robots are motorized pill robots for human body observation. The obvious benefits of this technology are immediate diagnosis anywhere on the planet with minimum invasiveness, but the company believes that one day in the not-too-distant future, an army of rice-grain-sized surgeons may operate on you while you go about your normal daily routine. The obvious advantages of this technology are immediate diagnosis anywhere with minimal invasiveness, but a trend in the near future is that an army of surgeons the size of a grain of rice can operate while you go about your normal daily life [7].

9. Diagnostic and imaging robots. Precision biopsy Robots, MRI/CT navigation, robotic ultrasound. They perform precisely activities that are achieved difficultly by a human hand.

Robotics applications in healthcare is shown in Figure 1.

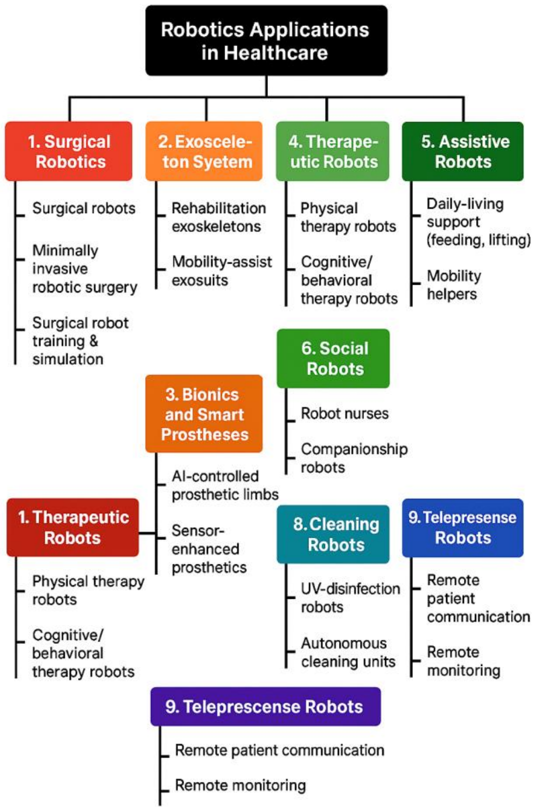


Fig.1. Robotics applications in healthcare

Technological enabling healthcare robots:

- Remote, mimics or gesture control;
- Voice control ;
- Machine visions ;
- Tactile sensors .

Main Benefits of robots in healthcare –save costs, reduce waste and improve patient care. Robots are revolutionizing medicine, although they don't come with much fanfare, they are changing everyday life and treatment outcomes. Fewer errors, faster surgeries, more accurate diagnoses, a safer environment. Combined with artificial intelligence, the entire sector is literally being "renovated." [2, 3].

II. ROBOTIC MODULAR PLATFORM AND ITS ELEMENTS

This chapter discusses a robotic modular platform and its elements. The first of them is the System for Analysis and Control of Mechanical Properties of Biological Tissues. Its application is for diagnostics of anomalous formations of a studied object. [8]

System for analysis and control of the mechanical properties of biological tissues or diagnostics device

The system uses controlled micromovements of the mounted manipulator to detect areas with structural or mechanical deviations and to determine their precise position in real time.

It is designed as an intelligent mechatronic device, the advantage of which is the ability to determine the parameter "relaxation time" τ, characterizing the internal structure of the tissue (elasticity and viscosity). This parameter is used as the main characteristic of the studied object and its change outside the specified limits signals the presence of anomalous areas in the object. The diagnostic approach applies both macro- and micro-stimulation, in which an instrument equipped with force sensors estimates the mechanical structure and response of the biological tissue.

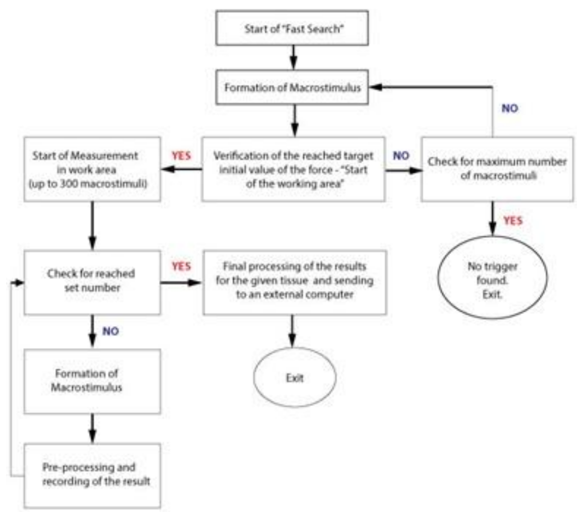


Fig 2. Block diagram of the operation of the diagnostics device

The 21st century has witnessed new and profound knowledge about cancer and significant progress in its treatment. However, problems related to the increasing incidence and successful outcome of its treatment remain. Radiation therapy (RT) is used for all types of cancer and is part of the procedure for its treatment. One of the disadvantages is excessive scattering and irradiation of healthy tissues. Advances in robotics and mechatronics can be used to optimize tumor treatment by personalizing the therapeutic dose based on clinical parameters and anatomical information.

A solution for a surgical robot device in local tumor therapy is proposed, consisting of a diagnostic device, upgraded with a generator that generates a programmed frequency, forms the necessary radio signal and feeds it to an emitter for performing radiotherapy. An electromagnetic field with a pre-programmed frequency and intensity is formed in the patient's body. The advantage is a gentle local therapy, in which the tumor is under control, while protecting the organs that are close to the diseased tissues. Another significant advantage of the solution is that the device simultaneously studies the characteristics of biomechanical tissues and applies local tumor therapy, unlike other solutions. Robotic Device for local tumor therapy is shown in Figure 3.

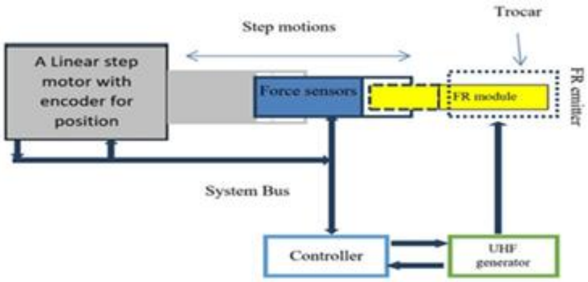


Fig 3 Robotic Device for local tumor therapy

The experiment involves searching for a deviation in the specified area with a specified power - the red graph. The blue graph shows the frequency of the generated RF signal used for irradiation. When the deviation is detected, the formation of microsteps is terminated and the generator starts operating in accordance with the specified program. Upon reaching the specified frequency, in the case of 434 MHz, the irradiation is maintained at the specified frequency and intensity for the time specified by the therapeutic program - in this case 10 seconds. Then the generator is turned off and the frequency drops to a minimum. For the red graph, the number of microsteps is located on the X axis. The search for the specified power value is shown on the Y axis. For the blue graph, the microsteps of the device are shown on the X axis, and the frequency of the irradiating signal in MHz is shown on the Y axis. Results obtained in simulation are shown in Figure 4.

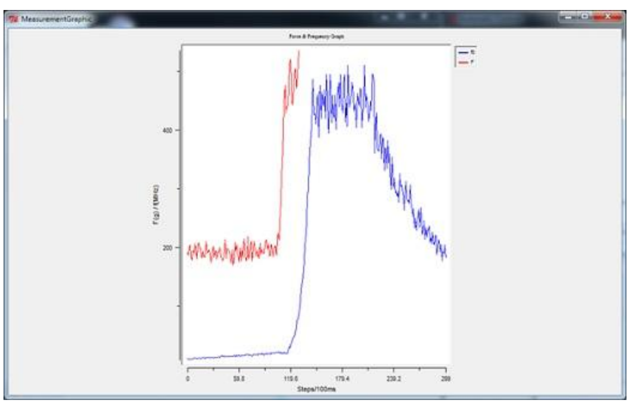


Fig. 4. Local therapy results obtained in simulation

System for collecting, processing and visualizing data with application in medicine – MOS. In order for a control device to function adequately, it is necessary to have appropriate software, an appropriate system for A Multifunctional Operator Station has been developed - the architecture of which can be seen in the figure. In addition to the other elements, the interest here is the training and simulation block and the Expert System block.

A part of software programmed using artificial intelligence techniques.

Such systems use data with expert knowledge to offer advice or make basic decisions in the field of medical diagnosis.

The components required for this block are:

- Knowledge database with facts, rules and constraints;
- Interpreter-Component for signal processing - using logical functions.

The Operating Station serves as a fundamental interface for the supervision and control of technological processes. To meet the requirements of real-time control systems, Operating Station (OS) software must possess capabilities for real-time processing, multitasking execution, and graphical visualization. Several software platforms are available for constructing Operating Stations, each offering specific advantages in terms of system architecture, functionality, and efficiency. Among the most commonly used platforms are UNIX, Linux, and Windows, which function as operating systems equipped with built-in features essential for a wide range of applications.

An alternative approach involves the use of open-source software implemented through scripting languages such as Tcl/Tk , Perl, and Python. These languages enable the seamless integration of executable code with resource requests directed to the operating system, embedded within procedural constructs. The adoption of scripting languages provides developers with direct access to system resources, significantly reducing development time while enhancing flexibility and scalability.

This chapter focuses on the design and implementation of an Operating Station for a Robotic System for Diagnostic and Therapeutic Tasks in Laparoscopic Surgery, leveraging the Tcl/Tk [9] scripting language under the Windows operating system. The development of such a system presents numerous challenges, as modern Operating Stations (OS) are inherently complex, integrating multiple interdependent subsystems that must communicate efficiently with each other as well as with external data sources. To simplify development and enhance system reliability, this work proposes an implementation based on widely adopted and well-documented technologies. In particular, Tcl/Tk is utilized due to its rich set of built-in libraries, multi-threading support, and cross-platform capabilities, making it highly suitable for real-time control applications. Furthermore, game engines written in C-based languages—both compiled and highly computationally efficient—are integrated to provide virtual reality (VR) functionalities for enhanced visualization and simulation in surgical applications.

The expert system block interacts with a graphical user interface that visualizes in a meaningful way all possible solutions, with the possibility of selection (pop-up windows, radio buttons, etc.). Multifuntional Operationg station is shown in Figure 5 [10].

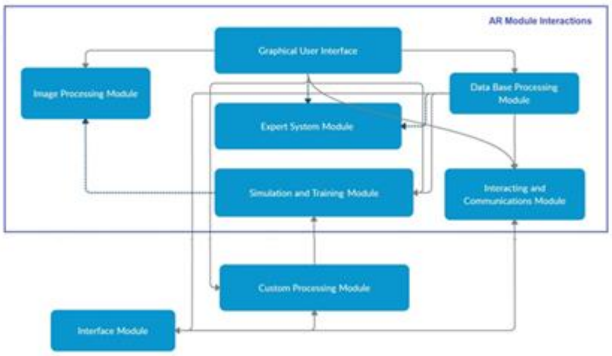


Fig 5. Multifuntional Operationg station

A mobile application for virtual and augmented reality has also been developed, as an upgrade to an operator station. An additional database has been developed for storing data from a lap instrument connected to the MOS, storing characteristics of various instruments and tissues.

When receiving contact data with a surface and depending on the data on its characteristics, the mobile application starts a training video and provides data on the measured values of the studied object (organ).

The application is suitable for training medical students and novice doctors.

Unity [11] is a software platphorme, that provides implementation of AR-VR

Two modes of operation are possible here.

AR mode provides the core functionality of MOSAR [10] , with live video enhanced with instrument and control displays, and AI-based video analysis. VR Surgical Training Simulator Mode allows surgeons to acquire, develop, and maintain skills in a variety of realistic environments with controls for safe operations. After surgery or virtual training, surgeons can review their performance. Architecture of Mobile AR-VR application for MOS is shown in Figure 6.

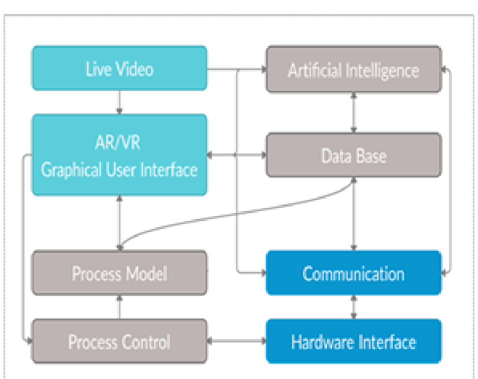


Fig 6. Architecture of Mobile AR-VR application for MOS

The ES block is replaced by an artificial intelligence block, and the graphical user interface is based on Virtual and Augmented Reality.

Robotics combined with simulation has been successfully used in the field of surgical training, providing a realistic environment for surgeons to practice complex procedures, allowing them to refine their acquired skills before performing real operations. A conceptual model of a training platform is proposed, where the monitoring of training tasks includes a virtual environment, simulation of the surgical workspace and surgical skills, robot simulation, force and position feedback, and a graphical user interface. A Model of Robot Platphorme for surgical training is shown in Figure 7.

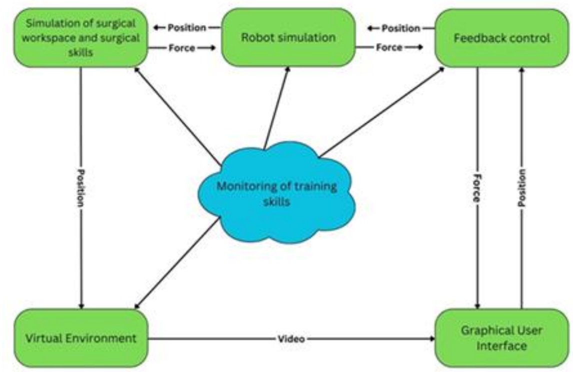


Fig 7. Conceptual model of platform for training skills in MIS and telesurgical robotics

Wireless mobile ECG device - The developed wireless mobile ECG (DECG) measures the electrical potential between two points of the human body, forms an ECG-gram and sends it on request via the wireless network to a specialized controller. The device consists of two modules - Measuring, developed based on MAX30003 [13]; Control, developed based on JN5168-001-M00 [14]. The control module controls the operation of the measuring device and provides communication of the device via a wireless interface - uMac or wired - USB. uMAC is based on the IEEE 802.15.4 stack. The measuring module generates the patient's ECG via two serviced electrodes and sends the information via an internal SPI interface to the control module. The MAX30003 (Analog Devices) specialized integrated circuit is a single-lead ECG front-end chip that has a built-in heart rate detection (R-R detection) function. Suitable for portable applications, it detects the ECG signals of the heart. The MAX30003, which performs all the analog signal processing from the electrodes and supports an interface via SPI [12, 13].

The MAX30003 has been carefully selected to take care of almost all analog operations, including acquisition, filtering, amplification, and digital conversion of the microvoltages produced by the heart. ECG device and uMAC conection with robots are shown in Figure 8. The Diagnostic Tool (DECG) feeds the measured information via the μMAC protocol to Block A of the Dedicated Controller, which via RS 232 C protocol feeds the gen4-

IoD-28T information (Block of the RMLI Dedicated Controller). Detailed information on the operation of the gen4-IoD-28T (Figure 9) is given in the description of [15, 16]. The measured information is accumulated in a local database on a SD card and is submitted to the RMLI Operator's Station upon request

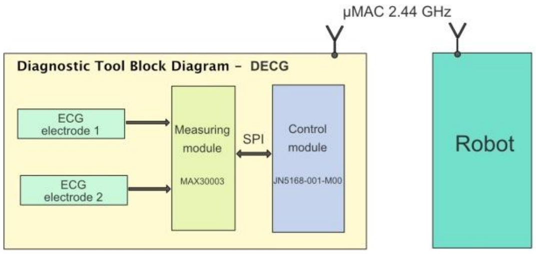


Fig. 8. ECG device and uMAC conection with robots

Fig. 9. Device gen4-IoD-28T [16].

More detailed information about the device can be found at Ivanova et. al. [15].

III. CONCLUSIONS AND INTENTIONS FOR FUTURE WORK.

The paper outlines key applications of robotic technologies in addressing the challenges of minimally invasive surgery (MIS), highlighting their role in shaping a new generation of surgical systems capable of better meeting patient-specific needs. Building on these advancements, several directions for future research are identified. One important avenue involves the development of a dedicated device for monitoring and regulating carbon dioxide levels within the abdominal cavity. Another promising direction is the creation of simulation software based on augmented and virtual reality (AR/VR) to support advanced training of emergency medical personnel in the operation of ECG devices. Additional work includes the design of pulse-monitoring systems and realistic models of laparoscopic instruments and human organs to improve surgical training and procedural planning. Further research should also explore the integration of STEMM (Science, Technology, Engineering, Mathematics, and Medicine) principles in the context of MIS, as well as the broader potential of the Internet of Medical Things (IoMT) for interconnected, data-driven surgical environments.

REFERENCES

[1] Ivanova, V., Boneva, A., Metodiev., V., Integrating STEMM Principles in Laparoscopic Surgery Training, In 34th International Scientific and Technical Conference Automation of Discrete Production Engineering, 29.06– 2.07 2025, Bulgaria, 2025, pp. 123-137, DOI: https://doi.org/10.53656/adpe-2025.11

[2] Arun Agrawal, Rishi Soni, Deepak Gupta, Gaurav Dubey, The role of robotics in medical science: Advancements, applications, and future directions, Journal of Autonomous Intelligence, 7(3), 2024, pp. 1-8, DOI: 10.32629/jai.v7i3.1008

[3] Standard Robots, Robotics in healthcare, October 2025, https://standardbots.com/blog/10-ways-robotics-are-being-used-in-healthcare-today

[4] Intuitive, https://www.intuitive.com/en-us/products-and-services/da-vinci

[5] Sense Medical Limited, 2025, https://www.paroseal.co.uk/training

[6] Ivanova, V., Boneva, A., The Internet of Medical Things Application - Challenges and Future Directions, j. „Complex Control Systems“ - Proceedings of the International Scientific Conference "Robotics & Mechatronics 2024", 7, Art. No. 5, 2024, pp. 1-6, http://ir.bas.bg/ccs/2024/07/5.pdf

[7] National Institute of Biomedical Imaging and Bioengineering, Robotic pill can orally deliver large doses of biologic drugs, March 2, 2022, https://www.nibib.nih.gov/news-events/newsroom/robotic-pill-can-orally-deliver-large-doses-biologic-drugs

[8] Ivanova, V., Bachvarov, D., Boneva, A., Andreev, R., Dobrinkova, N., System for analysis and control of mechanical properties of biological tissues, Utility Model., Registration №: 4347/BG/U/2019/4347-[10]/04.06.2025 г, Utility model number: 3323/31.10.2019, Bulletin on Patent Office in the Republic of Bulgaria, 12.1, 2019-2029), pp. 6622-6631, https://bulletin.bpo.bg/2019/binder-2019-12-16.pdf (in Bulgarian)

[9] Tcl Developer Xchange, https://www.tcl-lang.org/

[10] Ivanova V., Vasilev, P., Stoianov, I., Andreev, R., Bonev, A., Design of Multifunctional Operating Station based on Augmented Reality (MOSAR). Journal Cybernetics and Information Technologies, 21 (1), 2021, pp. 119-136, DOI: 10.2478/cait-2021-0009

[11] Unity Technologies, Scripting API, https://docs.unity3d.com/ScriptReference/Physics.html

[12] Maxim Integrated, MAX30003: Ultra-Low Power, Single-Channel Integrated Biopotential (ECG, R-to-R Detection) AFE, Data Sheet, 2021, https://datasheets.maximintegrated.com/en/ds/MAX30003.pdf

[13] ADI Engineer Zone, Are there instructions to connect the ECG leads for the MAX30003 or MAX30001?, 2024, https://ez.analog.com/temperature_sensors/w/documents/21199/are-there-instructions-to-connect-the-ecg-leads-for-the-max30003-or-max30001

[14] NXP Laboratories UK, Data Sheet: JN5168-001-Myy, JenNet-IP, ZigBee PRO and IEEE802.15.4 Module, Data Sheet, 2013, https://www.nxp.com/docs/en/data-sheet/JN5168-001-MXX.pdf

[15] Ivanova, V., Boneva, A., Ivanov, S., Vasilev, P., A Wireless Device to Modular Robotized Instrument for Health Information, Computer Science & Engineering: An International Journal (CSEIJ), 13(2), 2023, pp. 21- 33, DOI:10.5121/cseij.2023.13203

[16] 4DSystems, Gen4-Internet of Displays Series, Data Sheet, 2021, pp. 1-23, https://eu.mouser.com/datasheet/2/451/gen4_iod_datasheet_r_1_5-1627290.pdf

Unsupervised Physiological State Differentiation and Anomaly Detection

Galya Georgieva-Tsaneva

Institute of Robotics, Bulgarian Academy of Science, 1113 Sofia, Bulgaria

Abstract— **The article presents an analytical extension of our previously published study on HRV-based physiological state recognition using unsupervised learning. Building upon the original PCA + K-Means + DBSCAN framework, the current work focuses on a deeper interpretation of selected outcomes, particularly the interplay of fractal metrics (DFA α1/α2, Hurst exponent) and nonlinear entropy features in differentiating rest, stress, and physical load conditions. Additional analysis highlights the physiological relevance of borderline and anomaly cases identified by DBSCAN, suggesting early deviations in autonomic regulation. A comparison of the findings with recent HRV clustering studies is also made to highlight methodological advantages and limitations. The results reinforce the potential of label-free clustering for real-time monitoring and decision support in wearable health systems.**

***Keywords*—DBSCAN, HRV, K-Means, PCA.**

I. INTRODUCTION

Unsupervised machine learning techniques have gained increasing attention in recent years for their ability to extract structure and detect physiological states from Heart Rate Variability (HRV) data without requiring predefined labels. Prior studies have applied K-Means clustering to time- and frequency-domain HRV metrics to separate daily-life conditions such as rest, stress, and physical load, extending the approach to risk prediction using Gradient Boosting. In rehabilitation contexts, cluster analysis combined with Principal Component Analysis (PCA) has shown potential in identifying heterogeneous responders among patients with Parkinson's disease undergoing aerobic exercise. Wearable validation studies further demonstrated that HRV-derived parameters are sensitive to physiological differences related to fitness and body composition.

The present article serves as an analytical extension of our previously published work [1], in which a hybrid unsupervised framework—integrating PCA, K-Means, hierarchical clustering (Ward), and DBSCAN (Density-Based Spatial Clustering of Applications with Noise)—was used to distinguish physiological states and detect anomalies from combined linear, spectral, nonlinear, and fractal HRV features. The contribution of this extended analysis lies not in introducing a new model, but in re-examining key outcomes to better understand the latent structure of the data, the behavior of fractal and entropy metrics, and the physiological interpretation of borderline or atypical cases identified through density-based clustering. By contextualizing these findings within related literature, this study provides deeper insights into the utility and limitations of label-free HRV clustering for real-time monitoring and decision support in wearable health applications.

Related Works

Several recent studies have applied unsupervised learning approaches to HRV data to identify physiological states. In [2], K-Means clustering was applied to HRV features in the time and frequency domains to distinguish physiological states such as "rest", "stress" and "exertion" during daily activities, and Gradient Boosting was subsequently used to predict fall risk.

Assessing HRV in Parkinson's disease rehabilitation, a study [3] on 110 Parkinson's disease patients performed cluster analysis and PCA, identifying four different types of responders to aerobic exercise (vigorous, moderate, mixed and low responders).

Validation of the Polar H7 wearable device during physical activity [4] showed that the K-Means algorithm effectively classified participants according to fitness level and body composition. The results show that HRV parameters measured by the device are sensitive to physiological differences related to aerobic capacity and body composition.

Research Questions:

Q1. Which hidden patterns and relationships within HRV, entropy, and fractal descriptors can be revealed when these metrics are projected into a reduced latent feature space?

Q2. Are unsupervised machine learning methods capable of reliably distinguishing between different physiological states when no prior class labels are available?

Q3. How effectively can AI-driven visual interpretation tools (e.g., PCA plots, heatmaps, cluster maps) assist in identifying atypical responses, natural subgroups, and borderline cases within HRV datasets?

II. Methods

This study utilizes the same dataset, protocol, and HRV feature extraction pipeline presented in our previous work [1] on unsupervised physiological state recognition. HRV signals were recorded in three controlled conditions—rest, psychological stress, and physical load—and preprocessed through digital filtering and hybrid R-peak detection to derive time-domain, spectral, nonlinear, and fractal descriptors. The feature set included SDNN, RMSSD, LF/HF ratio, sample entropy, Hurst exponent, fractal dimension, and DFA $\alpha1/\alpha2$.

Dimensionality reduction was performed using Principal Component Analysis to reveal latent axes of variation. K-Means clustering ($k = 3$), hierarchical clustering (Ward), and DBSCAN were applied to evaluate complementary grouping behavior and anomaly detection. Unlike the original study, which focused on methodological development, the present work concentrates on reexamining these results—comparing the contributions of fractal and entropy metrics, analyzing borderline classifications, and contextualizing the observed patterns with recent HRV clustering literature.

A total of 22 healthy male athletes participated in the study [1], with HRV signals recorded in three physiological conditions—rest, psychological stress, and physical load—resulting in three recordings per participant.

III. Rezults

How the techniques used contribute to answering the research questions of the study (Table 1):

Fractal metrics (DFA $\alpha1$, $\alpha2$, Hurst exponent, fractal dimension) provide a robust and theoretically sound framework for capturing the self-similarity, scaling, and complexity of physiological time series. These features go beyond the capabilities of linear HRV parameters and are sensitive to long-term correlations and the degree of autonomic flexibility. Their use allowed us to uncover latent regulatory dynamics that are not visible in classical time- or frequency-domain analysis – effectively supporting the discovery of hidden structures in HRV data (Research Question 1).

Unsupervised machine learning algorithms (K-Means, DBSCAN, hierarchical clustering) offer several key advantages:

• They operate on unlabeled data, which is consistent with real biomedical settings where ground-truth annotation is often not available.

• K-Means provided efficient baseline clustering for the hypothetical three states (rest, stress, exercise).

• DBSCAN handled noise and outliers robustly, revealing transient and anomalous cases (e.g., participants in borderline physiological states).

• External validation metrics (Purity = 0.933, ARI = 0.89, NMI = 0.92) [1] indicate that the quality of the clustering is not only internally consistent, but also compatible with ex-pert-based physiological classifications (Research Question 2).

The AI-based visualizations (PCA biplots, heatmaps, clustermaps, and DBSCAN scatter plots) were instrumental in:

• Providing intuitive interpretation of high-dimensional HRV data.

• Revealing natural groupings in a reduced space.

• Highlighting the contribution of features to the principal components, which helps to understand the physiological basis of each cluster.

• Visual identification of anomalies, especially with DBSCAN, where noise points marked with an "X" indicate participants with potentially atypical autonomic profiles (Research Question 3).

Table 1. Strengths of the applied methods in relation to the research questions.

Method	*Research Question Addressed*	*Key Strengths*
Fractal Metrics (e.g., DFA α1/α2, Hurst exponent, Fractal Dimension)	Q1.	Capture long-range correlations, self-similarity, and complexity not visible in linear HRV measures.
K-Means Clustering		Simple, interpretable clustering into 3 states (Rest, Stress, Load); good baseline structure detection.
Hierarchical Clustering (Ward linkage)	Q2.	Provides nested structure and interpretability; confirms cluster separation aligned with physiological states.
PCA Biplots	Q1 & Q3.	Reveals major variance directions; links physiological features to clustering axes.
Heatmaps and Clustermaps	Q3.	Enable visual assessment of inter-individual similarity and distinct group profiles; identify hidden subgroups.
DBSCAN	Q2 & Q3.	Handles noise and outliers robustly; detects borderline cases and non-standard physiological profiles.

PCA revealed a dominant axis (PC1 ≈ 66%) that aggregates vagal-related indices (RMSSD, SDNN, Mean RR), while entropy and fractal metrics (SampEn, Hurst, DFA) articulate secondary axes that separate psychological stress from physical load. This indicates that combining magnitude (variability) and structural (complexity/fractality) descriptors provides complementary information for physiological state recognition. DBSCAN's isolation of borderline cases (detected outliers) illustrates the approach's potential for personalized early-warning in wearable systems

The PCA cluster analysis identified outliers that revealed deviations from the dominant physiological patterns in each condition. One participant from the Rest cluster was positioned at PC1 = –0.45, significantly higher than the typical Rest range (PC1 = –2.3 to –2.0), placing it closer to the stress centroid. This suggests possible increased sympathetic activation despite a nominal resting state. Another participant from the Load cluster showed PC1 = –0.10 and PC2 = –0.60, deviating from the typical Load position (> 0.3 in PC1 and < –1.5 in PC2) and aligning more closely with the Rest profile.

A comparison in terms of methodology, data types and clustering techniques of several articles with the present study is presented in Table 2.

Table 2. Methodology, data types, and clustering techniques.

Study	*Methodology*	*Data*	*Clustering Technique*
Ref. [2]	HRV-based stress detection and fall risk monitoring during daily activities; supervised ML models	Wearable HRV recordings during daily activities	K-Means
Ref. [3]	HRV as a biomarker for autonomic function in Parkinson's rehabilitation; exercise-induced changes	HRV during rehabilitation exercises	hierarchical clustering
Ref. [5]	Unsupervised identification of mental stress in firefighters using deep learning	HRV from firefighters under stress tasks	deep clustering (autoencoder-based)
Ref. [6]	Stratifying autonomic nervous system regulation patterns in healthy men	HRV from healthy men at rest	K-Means
Current study	HRV-based classification across rest, stress, and load using multimodal analysis	Experimental HRV dataset (N=22) in 3 conditions	Combination of clustering methods: K-Means, hierarchical, and DBSCAN

In contrast to the referenced studies, which often adopt a single clustering approach (e.g., K-Means in Messaoud et al. [2], hierarchical clustering in Basri et al. [3], or a deep learning pipeline without comparative unsupervised baselines in Oskooei et al. [5]), the present work leverages a multi-algorithmic clustering framework combining K-Means, hierarchical clustering, and DBSCAN. This diversification mitigates the method-specific sensitivity and instability seen in single-technique designs, enabling cross-method validation of the derived clusters. Moreover, while several prior works limit their feature sets to conventional time-domain and frequency-domain HRV indices, our methodology extends the feature space with nonlinear and fractal descriptors (SampEn, DFA α1, DFA α2, Hurst exponent), thus capturing subtler autonomic regulation patterns that traditional metrics fail to detect. The resulting clusters are therefore not only algorithmically more robust but also physiologically richer.

Justification of the PCA + K-Means + DBSCAN Framework

The combined use of Principal Component Analysis (PCA), K-Means, and DBSCAN offers several methodological advantages for unsupervised exploration of HRV data:

Dimensionality reduction via PCA allows the projection of heterogeneous HRV features (linear, spectral, nonlinear, and fractal) into a compact orthogonal space. This mitigates multicollinearity, improves computational efficiency, and highlights latent axes of physiological variation.

K-Means clustering is well-suited for discovering spherical and balanced group structures in the principal component space, providing a fast and interpretable baseline for group separation based on central tendencies.

DBSCAN complements this by identifying arbitrary-shaped clusters and outliers, which K-Means cannot capture. It is particularly effective for detecting transitional or rare physiological states, including those not represented in the majority of the population.

When combined, PCA simplifies the input space, K-Means uncovers global groupings, and DBSCAN identifies local density anomalies. When combined, PCA simplifies the input space, K-Means uncovers global groupings, and DBSCAN identifies local density anomalies.

In this study, PCA and DBSCAN complemented each other particularly well. PCA reduced the dimensionality of the HRV feature set to the first two principal components, which explained over 70% of the variance, enabling clear visual separation between Rest, Stress, and Load states while preserving the key physiological differences. DBSCAN then operated directly on these reduced coordinates to form clusters based on point density and identify anomalies (Cluster = –1). While PCA provided an intuitive visualization of grouping tendencies, DBSCAN quantitatively defined membership and automatically isolated atypical cases—such as participants with mixed autonomic responses—labeling them as “noise.” This dual approach yielded consistent overall cluster structures, with DBSCAN excelling in the detection of boundary and atypical profiles, and PCA offering a clearer interpretation of global inter-state relationships. This synergy enables both robust clustering and unsupervised anomaly detection in HRV monitoring.

Although the sample size is limited, it reflects a relatively homogeneous group of experienced combat sports athletes, allowing for a focused assessment of HRV dynamics under intense physical and psychological conditions. Participant variability was con-trolled by selecting only healthy men with regular training regimens (5 days per week for 2 hours), minimizing confounding factors such as age-related cardiovascular variations or comorbidities. The dataset included balanced recordings in three different physiological states—rest, exercise, and stress—collected under standardized training and competition conditions. The sensitivity of the model to clustering parameters was assessed by varying ε and minPts in DBSCAN, and the robustness was confirmed in several configurations (see Section 3.5).

Table 3 presents an analysis of similar studies on a number of parameters.

Table 3. Comparative Analysis of Related HRV-Based Studies and the Proposed Approach

Parameter	Ref. [2]	Ref. [3]	Ref. [6]	Ref. [5]	Current Study
Domain/ Application	HRV for stress and fall risk	HRV in Parkinson’s disease	Autonomic regulation in healthy males	Psychological stress in firefighters	Recognition: rest, stress, workload
Data type	HRV (time/frequency)	HRV (pre-/post-activity)	HRV	HRV	HRV (including nonlinear indicators)

Feature diversity	Medium	Moderate (low–medium)	Medium	Medium	High
Methodology	Supervised ML (with labels)	K-Means clustering	Clustering/ classification	Autoencoder + clustering	PCA + K-Means + DBSCAN (hybrid)
Computational complexity	Medium–high	Low	Medium	High	Medium
Requires labels	No	Partial	Partial	Yes	Yes
Anomaly detection	Limited	Yes	Moderate	Strong	Strong
Interpretability	Moderate	High	Moderate	Low	High
Main limitations	Requires labels and large datasets	Specific to clinical scenario	Limited to healthy males	Risk of overfitting	Requires fine-tuning in noisy data
Advantages of our approach	Fully label-free; robust to noise and incomplete data	Multi-component; applicable in different contexts	Works with diverse populations	Clear and easy to implement	A combination of visualization, clustering, and anomaly detection

Messaoud & Thamsuwan (2025) [2] study applied K-Means clustering to HRV features and showed stress-induced reductions in RMSSD and increases in LF/HF ratio, although they did not provide exact numerical values. Similarly, Basri & Turki (2025) [3] observed significant HRV changes in Parkinson's patients following aerobic activity, with post-exercise increases in vagal tone reflected by LF/HF normalization. In contrast to these qualitative results, our study provides a full quantitative differentiation between the three physiological states. For example, RMSSD decreases from 28.39 ± 7.24 ms (Rest) to 9.86 ± 4.81 ms (Stress) and partially recovers during Physical Load (18.22 ± 5.93 ms). Similarly, LF/HF ratio rises markedly from 0.59 ± 0.11 (Rest) to 2.12 ± 0.23 (Stress), and then decreases to 1.22 ± 0.14 (Load), reflecting sympathetic–parasympathetic shifts consistent with mental and physical burden. DFA analysis further confirms these transitions, with α1 dropping from 1.14 ± 0.06 (Rest) to 0.88 ± 0.08 (Stress), then rising to 1.04 ± 0.07 (Load), and α2 showing a similar trend: 1.33 ± 0.05 → 0.86 ± 0.07 → 0.99 ± 0.06, which aligns well with known autonomic modulation patterns.

Zamora-Justo et al. (2025) [7] reported that participants with metabolic syndrome exhibit significantly lower DFA and SampEn values (no specific numerical values) both at rest and during exercise compared to healthy controls, indicating reduced autonomic complexity in MetS—even though specific numeric values for DFA α1/α2 descriptors were not provided. This qualitative alignment further supports our findings, which show a pronounced reduction in DFA α1 (from 1.14 at rest to 0.88 under stress), with intermediate values under physical load (1.04), illustrating how DFA metrics reliably reflect different physiological stress states.

Materko [6] investigates autonomic nervous regulation in healthy men through machine learning using a limited set of HRV metrics (mainly linear time and frequency parameters). Classification and clustering were applied in a relatively homogeneous population and without emphasis on visualization of latent structure. In contrast, we use a more diverse set of metrics (including nonlinear and fractal), apply PCA and hybrid clustering (K-Means + DBSCAN) to recognize different physiological states in heterogeneous data.

Borthakur et al. (2018) [8] work with multimodal smartwatch signals and use K-Means, GMM and SOM to detect natural groups in the data. Although the method is completely label-free, the visualizations are limited to basic scatter plots and are not inte-grated with the interpretation of physiological indicators. Our study extends the approach by combining PCA and DBSCAN/K-Means, providing richer visualization (2D/3D PCA, heatmap, clustermap), and includes nonlinear HRV indices for deeper physiological interpretation.

Schrumpf et al. (2017) [9] apply hierarchical clustering on physiological parameters during exercise, demonstrating agreement with experimental phases. However, their approach has fixed metrics and is less robust to noise. Our work also achieves real-world match, but with a hybrid algorithm and adaptive settings, robust to noise and incomplete data, making it applicable to real-world “in-the-field” measurements.

Serantoni et al. (2022) [10] use clustering to dynamically track fatigue in running, linking HR patterns to VO_2max and the onset of fatigue. The method is highly effective, but focused in a specific sports context. Our study applies similar label-free clustering, but in a broader context – recognizing rest, stress, and exercise under a variety of conditions, with higher feature diversity and integrated anomaly detectors.

In the study by Oskooei et al. (2019) [5], a clustering approach on latent representations extracted by autoencoders was used to distinguish two psychophysiological clusters among firefighters participating in a training session. It is important to note that all participants were in an active, busy state – i.e., no resting control condition was included. However, the authors distinguished individuals with a normal parasympathetic response (high RMSSD, low LF/HF ratio) from those with manifest mental stress. In contrast, the present study uses real and clearly distinguished physiological states – rest, mental stress, and physical exertion – and offers an extended set of features, including fractal and entropic metrics. This provides a broader range of assessment and more interpretable behavior of the clustering algorithms. The comparison highlights the complementary nature of the approaches, with the present study emphasizing real-time biomedical applicability and interpretability of the indicators.

Limitations

The present study has several limitations that should be acknowledged.

First, the number of participants and recordings was relatively small, which con-strains the statistical power and generalizability of the findings. Although the dataset was balanced across three physiological states (rest, exercise, and stress) and collected under standardized conditions, the limited sample size means that the clustering patterns observed—while consistent—should be interpreted with caution.

Second, the participants represented a relatively homogeneous group of experienced combat sports athletes. This reduces variability from confounding factors such as age, sex, and comorbidities, but it also limits the applicability of the results to other populations, including sedentary individuals, women, or athletes from other sports.

Third, the HRV features used were extracted from short-term recordings under con-trolled conditions; further research should examine the robustness of the proposed PCA + K-Means + DBSCAN framework using long-term monitoring data, real-world environments, and larger, more heterogeneous cohorts.

Future Work

In the validation study [3], physical activity included aerobic and submaximal endurance exercise, which are representative of moderate daily activity and training load. In our future work, we will aim to extend these observations by applying clustering and dimensionality reduction techniques to HRV data collected during typical daily activities such as walking, standing, sedentary work, and light exercise, which reflect common physiological states encountered in real-world monitoring scenarios. This will support the applicability of the proposed framework in both fitness and general health contexts.

The study in [4] demonstrated that HRV analysis with PCA and clustering can reveal different responses to interventions, even in chronic conditions. This confirms the applicability of nonlinear and fractal features, as well as unsupervised clustering, in clinical scenarios. In future work, we plan to adapt our framework to assess rehabilitation effects, chronic stress or cognitive load, as well as to monitor patients with cardiac diseases who need health monitoring, including post-stroke conditions. This would require longer time series, customized anomaly algorithms and adaptive learning models, but we expect the proposed architecture to be portable and effective in such cases.

Future comparison with supervised algorithms (e.g. SVM, Random Forest, DNN) and performance metrics (e.g. F1-score, ARI, silhouette) would strengthen the value of the results.

In future research, we plan to integrate the developed approach for automatic differentiation of physiological states and detection of deviations in real time into an existing wearable device for registration and monitoring of cardiac signals. The device was created within the framework of a current research project and includes PPG/ECG sensors, a wireless communication module, and multiplatform monitoring software.

IV. Conclusion

This analytical extension provides a deeper interpretation of previously obtained results on unsupervised HRV-based physiological state recognition. The findings reaffirm the value of combining fractal and entropy metrics with dimensionality reduction and clustering techniques to reveal latent autonomic dynamics and differentiate physiological states. The examination of borderline and anomaly cases highlights the potential for early recognition of atypical autonomic responses in real-time monitoring contexts. Overall, this complementary analysis strengthens the evidence for the applicability of label-free clustering in wearable health systems and supports future integration into personalized feedback and decision-support platforms.

Acknowledgment

This research was funded by the National Science Fund of Bulgaria (scientific project "Research on AI-based methods for modeling, analyzing and forecasting biomedical time series by applying a digital twin concept"), Grant Number КП-06-Н97/12, 10 December 2025

References

[1] Georgieva-Tsaneva, G.; Cheshmedzhiev, K.; Tsanev, Y.-A.; Dechev, M. Physiological State Recognition via HRV and Fractal Analysis Using AI and Unsupervised Clustering, Information, 2025, 16, 718.

[2] Messaoud, I.B.; Thamsuwan, O. Heart Rate Variability-Based Stress Detection and Fall Risk Monitoring During Daily Activities: A Machine Learning Approach, Computers, 2025, 14, 45.

[3] Basri, A.M.; Turki, A.F. Evaluating Heart Rate Variability as a Biomarker for Autonomic Function in Parkinson's Disease Rehabilitation: A Clustering-Based Analysis of Exercise-Induced Changes, Medicina, 2025, 61, 527.

[4] Hernández-Vicente, A.; Hernando, D.; Marín-Puyalto, J.; Vicente-Rodríguez, G.; Garatachea, N.; Pueyo, E.; Bailón, R. Validity of the Polar H7 Heart Rate Sensor for Heart Rate Variability Analysis during Exercise in Different Age, Body Composition and Fitness Level Groups, Sensors, 2021, 21, 902.

[5] Oskooei, A.; Chau, S.M.; Weiss, J.; Sridhar, A.; Rodríguez Martínez, M.; Michel, B. DeStress: Deep Learning for Unsupervised Identification of Mental Stress in Firefighters from Heart-Rate Variability (HRV) Data, arXiv, 2019, arXiv:1911.13213.

[6] Materko, W. Stratifying Autonomic Nervous System Regulation Patterns in Healthy Men: A Machine Learning Approach, Artificial Intelligence in Health, 2025.

[7] Zamora-Justo, J.A.; Campos-Aguilar, M.; Beas-Jara, M.C.; Galván-Fernández, P.; Ponciano-Gómez, A.; Sigrist-Flores, S.C.; Jiménez-Flores, R.; Muñoz-Diosdado, A. Utility of Nonlinear Analysis of Heart Rate Variability in Early Detection of Metabolic Syndrome, Frontiers in Physiology, 2025, 16, 1597314.

[8] Borthakur, D.; Peltier, A.; Dubey, H.; Gyllinsky, J.; Mankodiya, K. SmartEAR: Smartwatch-Based Unsupervised Learning for Multi-Modal Signal Analysis in Opportunistic Sensing Framework, arXiv, 2018, arXiv:1808.06473.

[9] Schrumpf, F.; Bausch, G.; Sturm, M.; Fuchs, M. Similarity-Based Hierarchical Clustering of Physiological Parameters for the Identification of Health States—A Feasibility Study, Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, South Korea, 2017, 458–462.

[10] Serantoni, C.; Zimatore, G.; Bianchetti, G.; Abeltino, A.; De Spirito, M.; Maulucci, G. Unsupervised Clustering of Heartbeat Dynamics Allows for Real-Time and Personalized Improvement in Cardiovascular Fitness, Sensors, 2022, 22, 3974.

A Secure Educational Platform for the Management and Protection of Cardiological Patient Data

Galya Georgieva-Tsaneva

Institute of Robotics, Bulgarian Academy of Science, 1113 Sofia, Bulgaria

***Abstract*— This paper presents a secure information platform model designed to support medical training through the controlled introduction, storage, analysis, and presentation of patient cardiological data. The platform integrates data protection mechanisms directly into its architecture, applying a hybrid cryptographic approach based on RSA and AES to ensure confidentiality and integrity of electrocardiographic signals. The effectiveness of the protection scheme is evaluated using quantitative signal quality metrics and statistical analysis, demonstrating preservation of diagnostic information after encryption and decryption. The proposed platform enables safe educational access to clinically relevant data while ensuring compliance with data protection requirements and supporting modern digital medical education.**

***Keywords*—information platform, medical training, data protection, ECG security, patient data privacy.**

I. Introduction

The rapid digitalization of healthcare and medical education has led to an increasing reliance on information systems capable of managing, analyzing, and visualizing large volumes of sensitive patient data. In contemporary educational environments—characterized by blended and online learning modalities—medical students require access to realistic clinical data in order to develop practical skills, clinical reasoning, and diagnostic competence. At the same time, the use of real patient information introduces significant challenges related to data protection, privacy, and regulatory compliance.

Cardiological data, including electrocardiographic (ECG) recordings and derived analytical results, represent a particularly sensitive category of medical information. Such data are not only personally identifiable but also diagnostically rich, making them highly valuable for both clinical decision-making and educational purposes. Consequently, any educational platform that incorporates patient cardio data must ensure strict protection against unauthorized access, data leakage, or misuse, while preserving the integrity and diagnostic value of the stored signals.

The protection of medical information is governed by strict legal and ethical frameworks, most notably the General Data Protection Regulation (GDPR) (EU Regulation 2016/679), which mandates principles such as data minimization, confidentiality, integrity, transparency, and accountability. These requirements necessitate the integration of technical and organizational security measures at the earliest stages of system design. For educational platforms, this challenge is further amplified by the need to balance data protection with usability, accessibility, and pedagogical effectiveness.

Existing medical information systems and training platforms often focus either on clinical functionality or on data security, but rarely provide an integrated solution tailored specifically to the needs of medical education. In many cases, educational use is limited to anonymized datasets with restricted analytical capabilities, while advanced platforms used in clinical practice are not accessible for training purposes due to privacy concerns. This gap highlights the need for a dedicated information platform model that supports the introduction, storage, processing, and secure presentation of patient data in a controlled educational environment.

In response to these challenges, the present work proposes a model of an information platform designed for medical training, with a focus on cardiology education. The platform enables the controlled introduction of patient information and cardio records, the storage of analytical results obtained through mathematically based signal processing methods, and the secure presentation of numerical and graphical data to authorized users. A key design principle of the proposed model is the integration of data protection mechanisms—such as cryptographic techniques and access control—directly into the platform architecture.

The proposed platform model is intended to support medical students and healthcare trainees by providing access to realistic, clinically relevant data while ensuring compliance with data protection regulations. By combining secure data management, modular system architecture, and educational-oriented functionality, the platform aims to bridge the gap between theoretical instruction and practical experience. In this way, the model contributes to the development of modern digital training environments that enhance the quality and effectiveness of medical education without compromising patient privacy.

Aim of the Study

The aim of this study is to propose a secure information platform model for medical training that enables the introduction, storage, processing, and presentation of patient cardiological data while ensuring a high level of data protection. The platform integrates cryptographic protection mechanisms and controlled access management directly into its architecture in order to safeguard sensitive patient information from unauthorized access and misuse. By embedding data security as a core design principle, the proposed model supports the use of clinically relevant data in medical education without compromising confidentiality, integrity, or regulatory compliance.

II. RELATED WORKS

The development of information platforms for medical applications has gained increasing attention in recent years, driven by the rapid digitalization of healthcare and the growing need for secure handling of sensitive patient data. Modern medical information systems are expected to support not only data acquisition and storage, but also advanced analytical processing and visualization, while ensuring compliance with strict data protection regulations such as the General Data Protection Regulation (GDPR) [1-3].

A significant portion of the existing research focuses on the protection of medical data, particularly cardiological signals such as electrocardiograms (ECG), which are considered highly sensitive due to their diagnostic relevance and personal identifiability. Various cryptographic approaches have been proposed to ensure confidentiality and integrity of ECG data in telemedicine and remote monitoring systems. Symmetric encryption algorithms, such as AES, as well as hybrid schemes combining symmetric and asymmetric cryptography, are widely used due to their balance between security and computational efficiency [4-7]. The effectiveness of these approaches is commonly evaluated using signal quality metrics such as Signal-to-Noise Ratio (SNR) and Percentage Residual Difference (PRD), ensuring that encryption does not compromise diagnostic information [8-12].

In addition to conventional cryptographic methods, several studies explore signal-dependent and feature-based security mechanisms. For example, ECG-specific characteristics such as the QRS complex have been employed as cryptographic keys or as part of lightweight encryption schemes, particularly in wireless body area networks (WBANs) and Internet of Medical Things (IoMT) environments [5]. Chaos-based encryption techniques have also been investigated for medical signals and images, offering high resistance to brute-force and statistical attacks through nonlinear dynamic behavior [7]. Although these methods demonstrate strong security properties, their complexity and limited interpretability may hinder their integration into educational-oriented platforms.

Parallel to research on data protection, numerous works address the design of medical information systems and digital platforms for healthcare and education. Many existing platforms prioritize clinical functionality or real-time monitoring, often restricting access to real patient data for training purposes due to privacy concerns [6], [11]. Educational platforms, on the other hand, frequently rely on fully anonymized datasets, which reduces privacy risks but may limit the pedagogical value and realism of the learning process. Recent studies emphasize the importance of privacy-by-design principles and architectural-level security integration when developing digital health platforms, rather than treating data protection as a secondary or add-on component [2], [3].

The introduction of GDPR has further influenced the architectural design of medical information systems, highlighting requirements such as data minimization, controlled access, accountability, and transparency [1], [10]. Several authors propose integrated socio-technical or platform-oriented security models that combine cryptographic protection, access control, and organizational measures to ensure regulatory compliance [9], [10]. However, these solutions are typically aimed at clinical or large-scale healthcare infrastructures and do not explicitly address the specific needs of medical education and training.

Despite the extensive body of work on medical data protection and ECG encryption, relatively few studies propose an integrated information platform model that simultaneously supports secure data management, analytical processing, and controlled educational access. In particular, there is a lack of system-level models that combine cryptographic protection of cardiological data with structured presentation of analytical results tailored for medical training. This gap motivates the present work, which aims to unify secure data handling, modular platform architecture, and educational functionality within a single information platform model designed specifically for medical education purposes.

III. Methods

A. Methodological Framework

The methodological approach of this study is based on the design and implementation of a secure information platform model intended for medical training, with a particular focus on cardiology education. The proposed methodology integrates system architecture design, data protection mechanisms, and signal processing procedures within a unified framework. Emphasis is placed on embedding data security measures directly into the platform architecture, following the principles of privacy by design and security by design.

The methodology consists of three main components:

- organization and management of patient data within the platform;
- implementation of cryptographic protection for sensitive cardiological signals;
- controlled access and presentation of analytical results for educational purposes.

B. Information Platform Architecture and Data Flow

The proposed information platform is designed as a modular system that supports the introduction, storage, processing, and presentation of cardiological data. Patient data, including electrocardiographic (ECG) records and associated physiological parameters, are introduced into the system through a controlled input module. Sensitive personal information is stored in a protected internal archive with restricted access, while non-confidential attributes and analytical results are organized in a structured database for educational use.

The platform architecture follows a layered approach, separating data acquisition, secure storage, analytical processing, and presentation layers. User access is regulated through role-based access control, enabling differentiated permissions for doctors, administrators, and students. This separation ensures that medical trainees can access anonymized or non-identifiable data and analytical outputs without exposure to confidential patient information.

C. Cryptographic Protection of Cardiological Data

To ensure confidentiality and integrity of patient cardiological data, a hybrid cryptographic protection procedure is implemented. The method combines asymmetric and symmetric encryption techniques, leveraging the advantages of both approaches. Asymmetric encryption (RSA) is used for secure key exchange, while symmetric encryption (AES) is applied for efficient encryption of ECG signal data.

The protection procedure begins with the extraction of characteristic ECG components, including the QRS complex and associated peaks, which are essential for preserving diagnostic information. The ECG signal is then encrypted using the generated symmetric key. To enhance robustness and enable secure storage or transmission, a wavelet transform is applied to the encrypted signal, producing a transformed data representation suitable for further processing.

The decryption process involves the inverse wavelet transform followed by symmetric and asymmetric decryption, resulting in a reconstructed ECG signal. This approach ensures that the decrypted signal retains its diagnostic properties while remaining protected during storage and transmission.

Figure 1 illustrates the security scheme implemented in the proposed information platform for protecting cardiological data. The original ECG signal is encrypted using a hybrid cryptographic approach, where RSA is applied for secure key exchange and AES is used for efficient symmetric encryption of the signal. A wavelet transform is subsequently applied to the encrypted data, enabling secure storage or transmission while preserving the diagnostic structure of the signal. The reverse process involves decryption and inverse transformation, resulting in a reconstructed ECG signal with preserved clinical integrity.

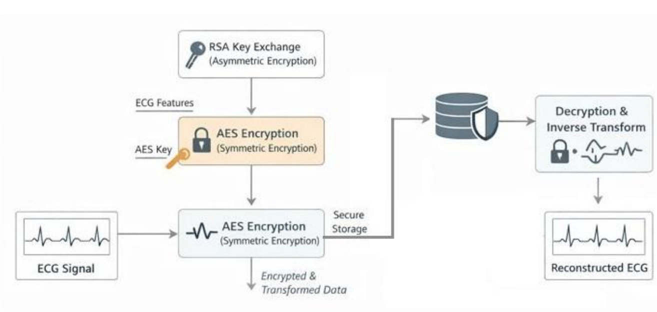


Figure 1. Security scheme of the proposed information platform for protection of cardiological data.

D. Evaluation Metrics for Signal Integrity

The effectiveness of the implemented protection procedure is evaluated using quantitative signal quality metrics commonly employed in biomedical signal processing. Two metrics are used to assess the similarity between the original ECG signal and the decrypted signal:

Percentage Residual Difference (PRD), which measures the relative difference between the original and reconstructed signals:

$$PRD = \sqrt{\frac{\sum_{i=1}^{L}(x_{ori}(i)-x_{tr}(i))^2}{\sum_{i=1}^{L}(x_{ori}(i))^2}}.100[\%], \quad (1)$$

where:

x_{ori} - input signal;

x_{ti} – converted signal;

L – is the length of x.

Signal-to-Noise Ratio (SNR), which quantifies the level of preserved signal information relative to introduced noise:

$$SNR = -10.log_{10}\frac{\sum_{i=1}^{L}(x_{ori}(i)-x_{tr}(i))^2}{\sum_{i=1}^{L}(x_{ori}(i))^2} \quad (2)$$

Low PRD values and high SNR values indicate minimal distortion and preservation of diagnostically relevant signal characteristics. These metrics provide objective evidence that the applied security mechanisms do not compromise the clinical usefulness of the ECG data.

E. Educational-Oriented Data Presentation

For medical training purposes, the platform provides controlled access to numerical and graphical results derived from the analysis of cardiological data. Only non-confidential information—such as age, gender, and selected physiological parameters—is presented to students, along with visualizations and parametric outputs of the mathematical analyses.

This approach enables learners to study real-world cardiological patterns and analytical results in a safe and regulated environment. By separating confidential patient identifiers from educational content, the platform supports practical training scenarios while maintaining compliance with data protection regulations.

IV. REZULTS

The performance of the proposed ECG protection approach was evaluated using real electrocardiographic signals with a duration of approximately 10 minutes each. A total of 22 ECG recordings were analyzed, and the quality of the reconstructed signals after decryption was assessed using the PRD and SNR metrics. Table 1 presents an extended statistical analysis of the obtained SNR and PRD values, including confidence intervals, effect size, and normality assessment. The results demonstrate statistically and practically significant preservation of ECG signal quality following the encryption–decryption process.

Table 1. Extended statistical summary of ECG signal quality metrics after the encryption–decryption process (n = 22)

Statistic	SNR [dB]	PRD [%]
Min	29.58	0.071
Max	42.06	0.248
Mean	37.13	0.12
Std. Dev.	8.19	0.006
95% CI for Mean	[33.50, 40.76]	[0.117, 0.123]
Cohen's d	0.87	−146.7
Effect size	Large	Extremely large
p-value (t-test)	< 0.001	< 0.0001
Shapiro–Wilk (normality)	p > 0.05	p > 0.05

The evaluation was conducted by comparing the original ECG signals with the corresponding decrypted signals after the complete encryption–decryption procedure. The resulting PRD and SNR values provide a quantitative measure of the distortion introduced by the applied security mechanisms. As shown in Table 1, the PRD values for all analyzed ECG signals remain well below 0.5%, while the SNR values exceed 29 dB in all cases. These results indicate a very high similarity between the original and reconstructed signals and confirm that the applied protection procedure introduces negligible signal distortion.

Statistical significance was evaluated using a one-sample Student's t-test. Both one-tailed and two-tailed tests were performed. The one-tailed test was used to verify whether the obtained SNR and PRD values exceeded clinically accepted thresholds, reflecting the directional nature of the research hypothesis. To provide a conservative assessment, a two-tailed test was also applied. In both cases, the results demonstrate statistically significant preservation of ECG signal quality after encryption and decryption ($p < 0.01$).

Figure 2 presents the distribution of SNR values for eight ECG recordings. The median SNR is well above 30 dB and no extreme outliers are observed, indicating stable and diagnostically reliable signal reconstruction after encryption and decryption.

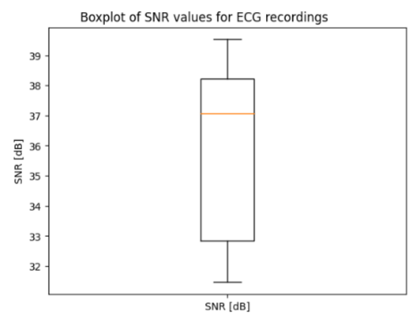


Figure 2. Boxplot of SNR values obtained after the ECG encryption–decryption process for eight ECG recordings.

Figure 3 illustrates the distribution of PRD values, showing consistently low distortion levels with no extreme outliers, which confirms preservation of ECG signal quality after encryption and decryption.

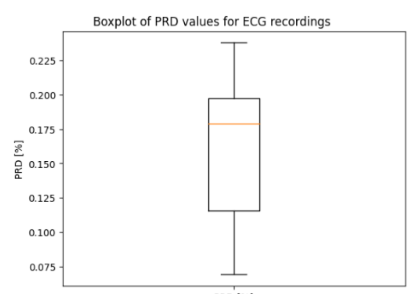


Figure 3. Boxplot of PRD values obtained after the ECG encryption–decryption process for six ECG recordings.

The obtained signal quality metrics demonstrate that the diagnostic characteristics of the ECG signals are preserved after encryption and decryption. This aspect is particularly important in biomedical applications, where even small distortions may affect clinical interpretation. The results confirm that the proposed hybrid cryptographic approach is suitable for practical implementation in software systems for the processing, analysis, storage, and protection of cardiological data, without compromising their clinical value.

In addition to the evaluation of the security procedure, an information platform with controlled public access was developed to demonstrate the functionality of the proposed system model. The platform presents non-confidential patient attributes—such as age, gender, weight, and height—along with numerical and graphical results obtained from the mathematical analysis of cardiological data. The presented datasets and analytical outputs are intended exclusively for educational and non-commercial purposes.

The implemented information platform provides several key functionalities relevant to medical training. It supports the presentation of large volumes of cardiological data (up to 1000 records), incorporates data protection mechanisms compliant with the General Data Protection Regulation (EU Regulation 2016/679), and is based on a modular architecture that facilitates extensibility and maintenance. Additionally, the platform supports multilingual operation in Bulgarian and English and can be accessed remotely via the Internet.

Access to information within the platform is organized according to user roles, including doctors, patients, and administrators, ensuring controlled and differentiated access rights. Confidential patient information is stored in a protected internal archive with restricted access and is encrypted to prevent unauthorized disclosure. Only authorized medical professionals, such as the attending physician and general practitioner, are permitted to access sensitive personal data, which is protected through authentication mechanisms.

The cardiological information stored in the platform includes both parametric and graphical results obtained from the analysis performed by the demonstration software system. The data are semantically organized, enabling efficient management and secure storage of cardiological records. This structured organization supports both clinical interpretation and educational use, while maintaining compliance with data protection requirements.

From an educational perspective, the proposed platform enables medical and healthcare students to gain access to realistic cardiological data and analytical results in a safe and regulated environment. By allowing students to study and compare cardio data from healthy individuals and patients with various cardiovascular diseases, the platform supports the development of practical skills and clinical reasoning. In the context of modern educational challenges, the presented information platform demonstrates the potential of secure digital systems to enhance the quality and effectiveness of medical training.

Limitations

The experimental evaluation was performed on a limited number of ECG recordings, which may affect the generalizability of the results to broader and more heterogeneous populations. The assessment of signal quality was based mainly on quantitative metrics such as SNR and PRD, without direct clinical validation by cardiology experts. In addition, the proposed security scheme was evaluated under controlled conditions, without considering network-related factors such as latency or real-time transmission constraints. Finally, the current platform implementation focuses on cryptographic protection and access control and does not yet include advanced anonymization or adaptive privacy mechanisms.

Future Work

Future work will focus on extending the experimental evaluation to larger and more diverse ECG datasets, including different cardiac conditions and recording environments. Additional validation will be performed through clinical assessment by cardiology experts and integration of real-time transmission scenarios. The platform will also be enhanced with advanced privacy mechanisms, such as adaptive anonymization, watermarking, and audit logging, to further strengthen data protection and regulatory compliance.

V. Conclusion

This study presented a secure information platform model for medical training that enables the introduction, storage, and controlled presentation of patient cardiological data. A hybrid cryptographic protection scheme was implemented and evaluated, demonstrating statistically and practically significant preservation of ECG signal quality after encryption and decryption. The results confirm that the proposed approach ensures high data security without compromising the diagnostic value of the signals. The developed platform represents an effective digital tool for supporting modern medical education while maintaining compliance with data protection requirements.

Acknowledgment

This research was funded by the National Science Fund of Bulgaria (scientific project "Research on AI-based methods for modeling, analyzing and forecasting biomedical time series by applying a digital twin concept"), Grant Number КП-06-Н97/12, 10 December 2025.

References

[1] European Data Protection Supervisor (EDPS). Health (personal data and GDPR special category). European Data Protection Supervisor, 2023.
[2] Tewari A. mHealth Systems Need a Privacy-by-Design Approach: Commentary on "Federated Machine Learning, Privacy-Enhancing Technologies, and Data Protection Laws in Medical Research: Scoping Review" J Med Internet Res 2023;25:e46700, doi: 10.2196/46700.
[3] Tunca, M. Privacy by Design: A Systematic Literature Review of European and British Regulatory Perspectives for Software and Information Engineering. In Proceedings of the 13th International Conference on Software and Information Engineering (ICSIE 2024); ACM: New York, NY, USA, 2025; pp. 54–63.
[4] Messinis S, Temenos N, Protonotarios NE, Rallis I, Kalogeras D, Doulamis N. Enhancing Internet of Medical Things security with artificial intelligence: A comprehensive review. Comput Biol Med. 2024 Mar;170:108036. doi: 10.1016/j.compbiomed.2024.108036.
[5] Jian W, Tabassum A, Li JP. Systematic survey on data security in wireless body area networks in IoT healthcare system. Front Med (Lausanne). 2024 Jul 30;11:1422911. doi: 10.3389/fmed.2024.1422911.
[6] Salehi Shahraki, A.; Lauer, H.; Grobler, M.; Sakzad, A.; Rudolph, C. Access Control, Key Management, and Trust for Emerging Wireless Body Area Networks. Sensors 2023, 23, 9856. https://doi.org/10.3390/s23249856.
[7] Cárdenas-Valdez, J.R.; Ramírez-Villalobos, R.; Ramirez-Ubieta, C.; Inzunza-Gonzalez, E. Enhancing Security of Telemedicine Data: A Multi-Scroll Chaotic System for ECG Signal Encryption and RF Transmission. Entropy 2024, 26, 787. https://doi.org/10.3390/e26090787.
[8] Sabri, O.; Al-Shargabi, B.; Abuarqoub, A.; Hakami, T.A. A Lightweight Encryption Method for IoT-Based Healthcare Applications: A Review and Future Prospects. IoT 2025, 6, 23. https://doi.org/10.3390/iot6020023.
[9] Conduah AK, Ofoe S, Siaw-Marfo D. Data privacy in healthcare: Global challenges and solutions. Digit Health. 2025 Jun 4;11:20552076251343959. doi: 10.1177/20552076251343959.
[10] Ogbodo, D.C.; Awan, I.-U.; Cullen, A.; Zahrah, F. From Regulation to Reality: A Framework to Bridge the Gap in Digital Health Data Protection. Electronics 2025, 14, 2629. https://doi.org/10.3390/electronics14132629.
[11] Yuan, S., Xu, S., Lu, X. et al. A privacy-preserving platform oriented medical healthcare and its application in identifying patients with candidemia. Sci Rep 14, 15589 (2024). https://doi.org/10.1038/s41598-024-66596-8
[12] Kh-Madhloom, J., Ghani, M.K.A., Baharon, M.R. (2021). ECG Encryption Enhancement Technique with Multiple Layers of AES and DNA Computing. *Intelligent Automation & Soft Computing*, *28(2)*, 493–512. https://doi.org/10.32604/iasc.2021.015129

Multifractal Wavelet-Based Analysis of Heart Rate Variability in Athletes Before and After Physical Exercise

Penio Lebamovski(1), Evgeniya Gospodinova(1)

(1) Institute of Robotics, Bulgarian Academy of Sciences, Sofia, Bulgaria

Abstract—The present study presents an integrated approach for the analysis of heart rate variability (HRV), combining time-frequency wavelet analysis and multifractal WTMM analysis. The method allows for the assessment of dynamic changes in RR-interval series at rest and after physical exertion following a one-hour sports training session. For quantitative assessment of signal complexity, the Multifractal Hurst Complexity Index (MHCI) was used, which combines information from the multifractal spectrum and the Hurst exponent. The results show decreases in the spectrum width (Δα: 0.61→0.45) and MHCI (0.78→0.59), and an increase in asymmetry (2.30→2.72) after physical exertion, reflecting a reduction in the complexity and adaptability of the autonomic nervous system. Energy concentration at certain scales and a more uniform structure of the RR intervals were also observed. The proposed approach provides sensitive integral indicators for assessing the functional state of the cardiovascular system and has potential for application in monitoring stress and physical exertion.

***Keywords*—Continuous Wavelet Transform (CWT), Complexity sygnal, Hurst Rate Variability (HRV), multifractal signal, Wavelet Transform Modulus Maxima (WTMM).**

I. Introduction

Heart rate is not a strictly periodic process, but a complex dynamic system regulated by the interaction between the sympathetic and parasympathetic parts of the autonomic nervous system. Heart Rate Variability (HRV), derived from the sequence of heart rate intervals (RR intervals), reflects the cardiovascular system's ability to maintain homeostasis and respond to internal and external influences [1,2,13]. Therefore, HRV is considered an important non-invasive indicator of the functional state of the autonomic nervous system and of the body's adaptability to stressors [3].

Classical methods for HRV analysis, based on the time- and frequency-domain characteristics of RR intervals, provide valuable information about the level of variability and autonomic balance. Widely used indicators include SDNN, RMSSD, LF/HF ratio, and spectral power across different frequency ranges. Despite their informativeness, these methods are limited in their ability to describe the nonlinear, non-stationary, and multiscale nature of physiological signals. Biological systems, including the cardiovascular system, exhibit complex dynamics that linear statistical methods cannot fully describe.

Physiological responses to stress, whether mental or physical, result in changes in autonomic balance, typically characterized by increased sympathetic activity and decreased parasympathetic regulation. These changes are reflected in the dynamics of RR intervals and can be better analyzed by nonlinear methods capable of capturing scaling properties, singularities, and complex signal structure [7-9]. In recent years, multifractal analysis has emerged as an effective approach to study the complex dynamics of HRV, as it allows for the description of signals characterized by multiple scales and varying degrees of local regularity.

One of the most effective methods for multifractal analysis is the Wavelet Transform Modulus Maxima (WTMM), which combines wavelet transformation with modulus maxima analysis and enables singularity detection, determination of scaling exponents, and construction of the multifractal spectrum [4,10,11]. The complexity, variability, and adaptability of physiological signals can be assessed using the parameters of the multifractal spectrum. In addition to the width of the multifractal spectrum, an important characteristic of the multifractal signal structure is the asymmetry of the spectrum. It provides information about the distribution of local scaling exponents and the dominant type of fluctuations in the signal. A left- or right-asymmetric multifractal spectrum indicates a different participation of large and small fluctuations in the signal structure, which is related to the mechanisms of heart rate regulation. Therefore, the asymmetry of the multifractal spectrum can serve as an additional indicator of the cardiovascular system's structural organization and adaptability.

Despite growing interest in multifractal methods for HRV analysis, several significant limitations remain in the literature. Most studies have focused on individual multifractal parameters, such as spectral width or Hurst exponent, without using integral indicators to characterize the overall structural complexity of the signal [5,6]. In addition, relatively few studies combine multifractal WTMM analysis with time-frequency wavelet methods to study HRV during physical exertion, which complicates interpretation and limits practical application. Therefore, the development of integral multifractal indicators for the analysis of HRV under stressful conditions remains an open scientific problem.

In this regard, the present article proposes an integrated approach to the analysis of heart rate variability, based on multifractal WTMM analysis and the introduction of the Multifractal Hurst Complexity Index (MHCI), which combines information from the multifractal spectrum and the Hurst exponent into a single generalized quantitative indicator. The proposed approach enables a more robust assessment of changes in the structural complexity of RR intervals during physical exertion.

The study hypothesizes that physical exertion during training leads to significant changes in the multifractal characteristics of RR intervals, reflecting a decrease in the complexity of the cardiac signal and a change in the adaptability of the autonomic regulation.

The main objective of the present work is to investigate the influence of a stressful situation induced by physical exertion on the dynamics of heart rate, as measured by WTMM of RR intervals, and to use an index of multifractal complexity for the quantitative assessment of structural changes in the signal.

II. Guidelines For Manuscript Preparation

A. Data Ascusition

The study used heart rate recordings from 28 athletes aged 18-20 years, practicing wrestling, recorded in a state of rest and after physical exertion induced by a one-hour training session. Electrocardiographic (ECG) signals were recorded with a Holter device, with recordings made for 15 minutes before the start of the training and 15 minutes immediately after its completion. The RR intervals, representing the time between consecutive R-peaks, were extracted from the ECG signals and used for subsequent analysis of heart rate variability. The study was conducted with approval from the Ethics Committee, and all participants signed informed consent to participate in the study. All participants were clinically healthy and did not have cardiovascular diseases.

B. Continuous Wavelet Transform

To analyse the non-stationary properties of the RR interval series, a continuous wavelet transform (Continuous Wavelet Transform , CWT) was used, which provides both time and frequency representations of the signal. This method is particularly suitable for analysing physiological signals characterised by variable-frequency components and non-stationary behaviour. In the present study, the Morlet wavelet function was used for its good time-frequency localisation and its ability to analyse non-stationary biomedical signals effectively.

The continuous wavelet transform is defined as:

$$W(a,b) = \frac{1}{\sqrt{a}} \int x(t)\psi\left(\frac{t-b}{a}\right) dt \tag{1}$$

where

- a –scale parameter;
- b – time parameter;
- ψ – wavelet function.

C. Wavelet Transform Modulus Maxima (WTMM)

The multifractal analysis was performed using the WTMM method. The method is based on locating the wavelet transform's modular maxima at different scales and tracking their evolution [4]. By analysing the scaling behaviour of these maxima, the signal's singularities and multifractal properties can be determined.

The partition function is calculated with the following equation:

$$Z(q,a) = \sum |W(a,b)|^q \tag{2}$$

The dependence between Z(q, a) and the scale a determines the scaling function τ(q), from which the multifractal spectrum f(α) is obtained through Legendre transformation.

D. Multifractal Spectrum

The multifractal spectrum f(α) describes the distribution of singularities in the signal and provides information about its complexity and scaling properties [5,12]. The main parameters of the spectrum are:

- spectrum width Δα;
- position of maximum α_0;
- spectrum asymmetry.

The width of the spectrum, which characterises the degree of multifractality and complexity of the signal, is determined by the following equation:

$$\Delta\alpha = \alpha_{max} - \alpha_{min} \tag{3}$$

E. Asymmetry of Multifractal Spectrum

The asymmetry of the multifractal spectrum characterises the shape of the spectrum and the distribution of fluctuations in the signal. It indicates whether small or large fluctuations dominate the multifractal structure and can serve as an indicator of the signal's structural organisation and the dynamics of autonomic regulation. Positive asymmetry is usually associated with the dominance of small fluctuations and a more stable signal structure. In contrast, negative asymmetry indicates a greater influence of large fluctuations and more uneven dynamics. Changes in the asymmetry of the multifractal spectrum may reflect changes in the regulatory mechanisms of the autonomic nervous system and the body's adaptive responses to stress or physical exertion.

F. Multifractal Hurst Complexity Index (MHCI)

The Multifractal Hurst Complexity Index, based on the multifractal spectrum and the Hurst exponent, was used to quantify signal complexity. This index is an integral indicator of the signal's structural complexity, correlation properties, and multifractality. By combining information on the long-term dependence described by the Hurst exponent with the characteristics of the multifractal spectrum, the index enables a more complete assessment of the signal's nonlinear dynamics. Higher MHCI values correspond to greater complexity, better structural organisation, and greater adaptability of the physiological system, while lower values indicate reduced complexity and a more limited regulatory capacity. Therefore, the index can be used as an integral indicator for assessing the functional state of the autonomic nervous system and the body's adaptive responses in different physiological states, such as rest, stress, and physical exertion. An advantage of MHCI is that it combines information about the scaling behaviour, correlation structure, and multifractal properties of the signal into a single generalised parameter, facilitating comparative analysis across different physiological states and subject groups.

G. Statistical Analysis

A t-test was used to compare the studied parameters between the rest and physical exertion states, with $p < 0.05$ as the criterion for statistical significance. All results are presented as mean ± standard deviation (mean ± SD).

III. Results

This section presents the main results of the HRV analysis in the studied athletes at rest and after physical exertion. The aim is to illustrate the changes in the dynamics of RR intervals and signal structure related to the adaptation of the autonomic nervous system to physical exertion.

Figure 1 shows the RR interval series at rest and after exercise. Clearly pronounced changes in the variability and structure of the signal are observed after exercise, which is an indicator of changes in the autonomic regulation of cardiac activity.

Figure 2 illustrates the wavelet scalograms of the RR intervals for the two conditions. The scalograms show a different distribution of energy across scales, indicating a change in the frequency structure of the signal after physical exertion.

Figure 3 presents the dependence of the Partition function Z(q,a) on the scale for different values of q. The obtained linear dependences on the logarithmic scale confirm the presence of fractal behaviour in the RR interval series.

Figure 4 shows the multifractal spectrum f(α) for the resting state and after physical exertion. Differences in the width and shape of the spectrum are observed, indicating changes in the complexity and fractal structure of the signal after physical exertion.

Table I summarizes the values of the main multifractal parameters – spectrum width (Δα), spectrum asymmetry and multifractal complexity index (MHCI) at rest and after physical exertion.

TABLE I. Statistical parameters of variables at Rest and Stress

Parameter	**Rest** mean±std	**Stress** mean±std	**p-value** mean±std
Multifractal spectrum	0.61±0.10	0.45±0.09	<0.0001
MHCI	0.78±0.09	0.59±0.07	<0.0001
Asymmetry	2.30±0.21	2.72±0.35	<0.0001

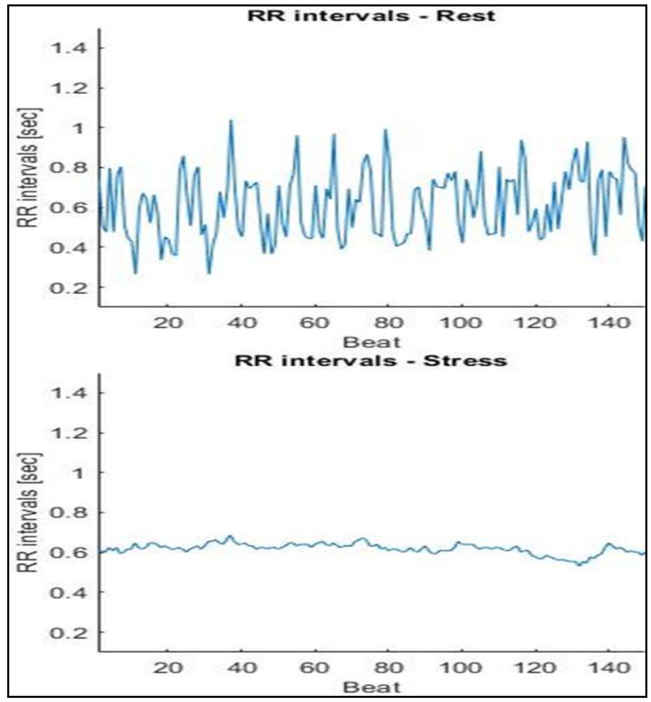


Fig 1. RR interval time series for rest and post-training conditions

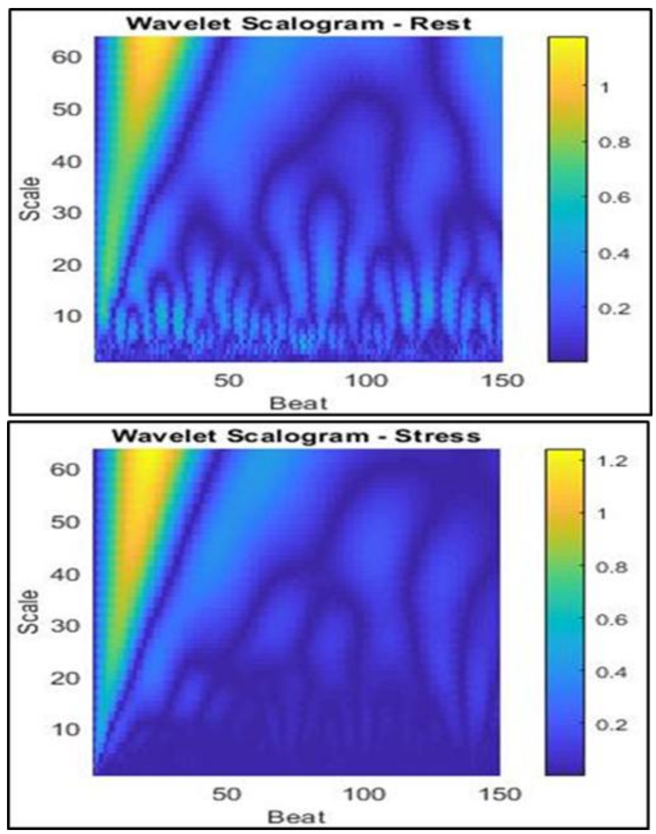


Fig. 2. Wavelet scalogram of RR intervals

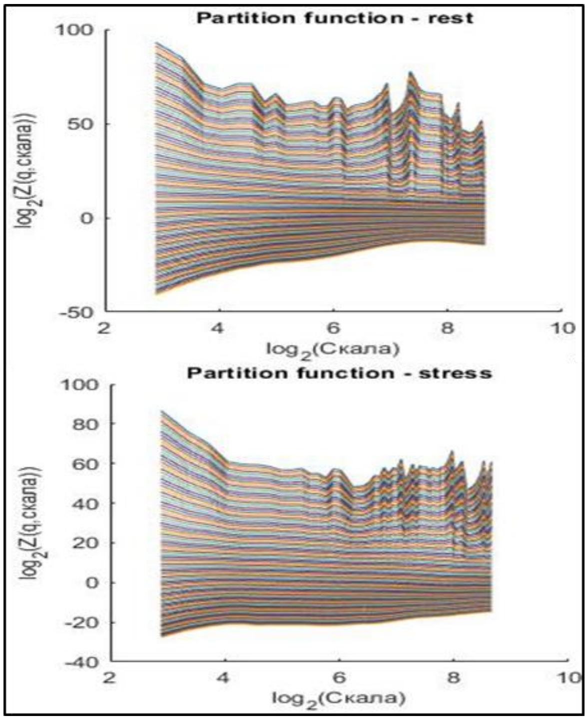


Fig. 3. Partition function of RR intervals

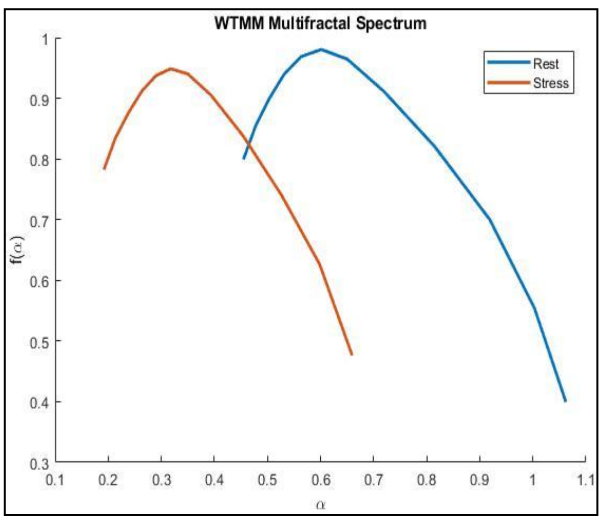


Fig. 4. Multifractal spectrum f(α) of RR intervals

IV. DISCUSSION

In the present study, multifractal analysis of RR intervals based on the WTMM method revealed significant alterations in the structural complexity of cardiac rhythm after one hour of exercise. The results summarized in Table 1 demonstrate significant post-exercise changes in heart rate. Specifically, a decrease in the spectrum width (Δα) and the multifractal heart rate complexity index (MHCI) was observed, accompanied by an increase in the spectrum asymmetry index. These changes suggest a modification in autonomic nervous system balance and altered cardiovascular regulation following physical load.

The decrease in the width of the spectrum (Δα) reflects less variability of the RR intervals on different time scales, which is an indicator of a reduction in the complexity of cardiac regulation. Changes in asymmetry indicate that the dominant fluctuations change: at rest the spectrum is slightly left-sided, which reflects a greater participation of large fluctuations; after exercise the spectrum becomes more symmetrical, indicating a decrease in the participation of large and sharp fluctuations. The decrease in the MHCI index is of particular importance, since this index is related to the Hurst parameter, which characterizes the long-term correlation and fractal properties of heart rate variability. Higher values of the Hurst parameter are associated with greater complexity and better adaptability of the cardiovascular system, while values close to 0.5 indicate a more random nature of the signal and reduced regulatory capacity of the autonomic nervous system. In the present study, a decrease in MHCI from 0.78 at rest to 0.59 after exercise was observed, which can be interpreted as a decrease in the fractal complexity of the heart rhythm and a decrease in the adaptability of the body as a result of physical stress and fatigue.

The increase in asymmetry also indicates an increased load on regulatory mechanisms and a change in autonomic balance in the direction of increased sympathetic activity and decreased parasympathetic activity after physical exertion.

The analysis of the partition function shows a clear difference between the state of rest and the state after physical exertion. In the state of rest, a greater separation of the curves is observed for different values of the parameter q, which indicates a strong dependence on q and the presence of a well-expressed multifractal behaviour of the signal. This indicates that the dynamics of the RR intervals in the state of rest is characterized by a complex multifractal structure. In the state of stress, the partition function curves are closer to each other and the dependence on the parameter q is less pronounced, which indicates a reduction in the multifractal properties and a transition to a more monofractal behaviour of the signal. This means that during physical exertion, the dynamics of the heart rate becomes simpler and less complex. These results are consistent with the obtained reduction in the multifractal width Δα and the MHCI index and confirm the hypothesis of a reduction in physiological complexity during stress and physical exertion.

This combination of changes in the studied parameters indicates that physical exertion causes a reduction in the complexity of the heart rhythm and reduced adaptability, accompanied by a concentration of energy at certain scales, also observed in the Wavelet scalogram (Figure 2).

Therefore, physical exertion leads to a decrease in heart rate variability, a loss of heart rhythm complexity, and reduced adaptability of the autonomic nervous system.

The results are consistent with the concept of "loss of complexity" described by the authors [13], according to which physiological stress and exercise lead to a decrease in HRV complexity. The reduction in Δα, asymmetry, and

MHCI is consistent with previous observations that stressful conditions reduce the multifractal properties of RR intervals and the adaptability of the autonomic nervous system. Similar effects have been reported for both exercise and cognitive or psycho-emotional stress.

The reduction in multifractal complexity observed after exercise likely reflects a shift toward sympathetic dominance and reduced parasympathetic modulation. This suggests that the cardiovascular system adapts to physical load by optimizing regulatory mechanisms, leading to more uniform and less complex RR interval dynamics. Additionally, the observed changes in spectrum asymmetry indicate a redistribution of dominant fluctuation scales, which may represent a specific adaptive response of the cardiovascular control system to physiological stress.

The integrated WTMM approach, combining multifractal spectrum analysis and MHCI, allows for both quantitative and qualitative description of changes in the structural complexity of RR intervals. The asymmetry analysis provides additional information about the nature of fluctuations and dynamic organization of the signal, which would be difficult to capture using classical time and frequency indices. Thus, this approach can be used for monitoring physiological stress and assessing the adaptability of the cardiovascular system in sports and clinical settings.

V. STUDY LIMITATION

The current study has several limitations. The number of participants was limited, which may affect the statistical power of the analysis. In addition, the focus was on a one-hour exercise session and the immediate effects after exercise; long-term changes and the influence of different types of physical activity should be investigated in future studies. The interpretation of integral multifractal indices, including MHCI, also requires further validation in different populations.

Despite these limitations, the present study demonstrates that multifractal WTMM analysis and MHCI are effective tools for assessing the functional state of the autonomic nervous system and heart rate dynamics during exercise. The results obtained confirm the concept of complexity reduction under stressful conditions and highlight the potential of integral multifractal metrics as sensitive markers of adaptability and physiological reactivity.

VI. CONCLUSION

The present study demonstrates that multifractal WTMM analysis combined with the multifractal complexity index provides sensitive and integral indicators of the structural complexity and adaptability of the cardiovascular system. The results obtained show that one hour of physical exercise leads to a significant decrease in multifractal width ($\Delta\alpha$), a change in spectral asymmetry, and a decrease in MHCI, reflecting a reduction in the complexity and adaptability of RR intervals under stress. These findings support the concept of a reduction in physiological complexity under stress conditions and highlight the potential of integral multifractal indicators as sensitive markers of the functional state of the autonomic nervous system.

The proposed multifractal approach expands the possibilities for quantitative analysis of heart rate variability and can be used for stress monitoring, assessment of recovery and adaptation to sports loads, as well as in clinical practice for assessing the functional state of the autonomic nervous system. This defines multifractal indicators and the MHCI index as promising quantitative markers for assessing physiological stress and the adaptive capacity of the body.

In summary, multifractal analysis of HRV can be considered as a reliable tool for quantitatively characterizing complex physiological signals and for assessing the adaptability of the organism to stressful influences.

ACKNOWLEDGMENT

This research was funded by the National Science Fund of Bulgaria (scientific project “Research, mathematical analysis and assessment of the impact of stress on cardiac data”), Grant Number KP-06-M72/1, 5 December 2023.

REFERENCES

[1] P. Ch. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. G.Rosenblum, Z. R. Struzik, and H. E. Stanley, “Multifractality in human heartbeat dynamics,” Nature, vol. 399, 1999, pp. 461–465. https://doi.org/10.1038/20924.

[2] P. Ch. Ivanov et al., “From 1/f noise to multifractal cascades in heartbeat dynamics,” Chaos, vol. 11, no. 3, 2001, pp. 641–65. https://doi.org/10.1038/20924.

[3] R. Sassi, M. G. Signorini, and S. Cerutti, “Multifractality and heart rate variability,” Chaos, vol. 19, no. 2, 2009. https://doi.org/10.1063/1.3152223

[4] R. Galaska, D. Makowiec, A. Dudkowska, A. Koprowski, K. Chlebus, J. Wdowczyk-Szulc, A. Rynkiewicz. “Comparison of wavelet transform modulus maxima and multifractal detrended fluctuation analysis of heart rate in patients with systolic dysfunction of left ventricle”. Ann Noninvasive Electrocardiol. 2008 Apr;13(2):155-64. https://doi.org/10.1111/j.1542-474X.2008.00215.x

[5] R. Fischer, M. Akay, P. Castiglioni, and M. Di Rienzo, “Multi- and monofractal indices of short-term heart rate variability,” Medical & Biological Engineering & Computing, vol. 41, no. 5, pp. 543–549, 2003. https://doi.org/10.1007/BF02345316

[6] V. Catrambone, R. Barbieri, H. Wendt, P. Abry, and G. Valenza, “Functional brain-heart interplay extends to the multifractal domain,” Philosophical Transactions of the Royal Society A, vol. 379, 2021. https://doi.org/10.1098/rsta.2020.0260

[7] A. Maluckov, D. Stojanovic, M. Miletić, L. Hadzievski and J. Petrović,“Multifractal features of multimodal cardiac signals: Nonlinear dynamics of exercise recovery,” ArXiv abs/2509.23317, 2025. https://doi.org/10.1063/5.5.00303657

[8] J. W. Kantelhardt, S. Zschiegner, Eva Koscielny-Bunde, S. Havlin, A. Bunde, H. Stanley,“Multifractal detrended fluctuation analysis of nonstationary time series,” Physica A, 2002.

https://doi.org/10.1016/S0378-4371(02)01383-3
[9] J. W. Kantelhardt, "Fractal and Multifractal Time Series". In: Meyers, R. (eds) Encyclopedia of Complexity and Systems Science. Springer, New York, NY, 2009. https://doi.org/10.1007/978-0-387-30440-3_221

[10] S. Jaffard, B. Lashermes, and P. Abry,"Wavelet leaders in multifractal analysis," Wavelet Analysis and Applications, 2006. htpps://doi.org/10.1007/978-3-7643-7778-6_17
[11] J. F. Muzy, E. Bacry, and A. Arneodo,"The multifractal formalism revisited with wavelets," International Journal of Bifurcation and Chaos, 1994. https://doi.org/10.1142/S0218127494000204
[12] P. Mukli P, Z. Nagy, FS.Racz, P. Herman, A. Ek. "Impact of Healthy Aging on Multifractal Hemodynamic Fluctuations in the Human Prefrontal Cortex." Front Physiol. 2018 Aug 10;9:1072. https://doi.org/10.3389/fphys.2018.01072
[13] A.Goldberger, A.N. L. A. N. Amaral, J. M. Hausdorff, P. Ch. Ivanov, C.-K. Peng, and H. E. Stanley. "Fractal Dynamics in Physiology: Alterations with Disease and Aging." Proceedings of the National Academy of Sciences of the United States of America 99, no. 3 , 2002,: pp. 2466–72. http://www.jstor.org/stable/3057579

Hybrid Artificial Intelligence Model for R-Peak Forecasting in Electrocardiogram Signals

Dobromir Slavov(1), Ekaterina Popovska (2), and (3) Galya Georgieva-Tsaneva

(1)(2)(3) Institute of Robotics, Bulgarian Academy of Science, 1113 Sofia, Bulgaria

Abstract- **Short-term cardio time series offer valuable insight into beat-level, moment-to-moment variations in heart rhythm and autonomic modulation. Beat-to-beat heart-rate sequences derived from electrocardiogram recordings demonstrate high non-stationarity and natural physiological variability, limiting the predictive stability of single-method data-driven approaches at short horizons. This study introduces the initial methodological milestone of a hybrid**

Artificial Intelligence framework for short-horizon heart-rate time-series analysis. The approach integrates rolling temporal representations learned by recurrent neural architectures with a compact set of physiologically informed descriptors extracted from R-peak annotations and derived R-R intervals. The forecasting task is formulated as a five-beat-ahead regression problem using only short preceding beat sequences, targeting embedded real-time monitoring scenarios rather than long-range signal prediction. The work benchmarks a baseline long short-term memory recurrent model against a hybrid feature-fusion extension, emphasizing trend adherence, temporal sensitivity and model stability under restricted context.

The findings confirm that lightweight hybrid fusion can enhance short-horizon responsiveness to transient fluctuations while preserving overall clinical plausibility and computational compactness. The methodology establishes a reproducible benchmark pipeline for short-context cardio time-series forecasting and supports future research extensions toward broader rhythm conditions and real-time AI monitoring pipelines.

***Keywords*—Artificial intelligence, cardio signals, deep learning, heart rate dynamics, hybrid models, nonlinear analysis, time series.**

I. Introduction

Short-term cardiovascular dynamics provide critical insight into the moment-to-moment regulation of heart function and autonomic nervous system activity. Time series derived from electrocardiographic recordings, beat-to-beat RR intervals, and instantaneous heart rate measurements are widely used for assessing physiological state, stress response, and early indicators of cardiac dysfunction [1], [2]. Unlike long-duration ECG monitoring, short cardio time series capture transient beat dynamics within limited observation windows, making their modeling particularly sensitive to signal non-stationarity, noise, and abrupt physiological modulation.

Traditional heart rate variability analysis methods rely on time-domain, frequency-domain, and nonlinear descriptors that remain clinically meaningful due to their interpretability and strong physiological grounding [3]. Metrics such as mean RR interval, variability dispersion, and entropy-based complexity indices have been applied for decades in ambulatory and clinical monitoring. However, these approaches do not learn temporal dependencies and may lose sensitivity to fast beat-level transients when the signal context is short.

Modern neural sequence models, including recurrent neural networks, one-dimensional convolutional networks, and attention-based architectures, have demonstrated the ability to learn representations directly from cardio signals [4], [5]. Their deterministic precision, however, may degrade when applied as standalone predictors on chronologically short windows, where data scarcity increases the likelihood of temporal oversmoothing and instability under transient fluctuations.

Hybrid artificial intelligence architectures attempt to mitigate these issues by integrating domain-informed variability markers extracted from RR dynamics into neural temporal embeddings [6], [7]. These strategies align with the broader scientific shift toward combining model-based signal meaning with data-driven learning, particularly for short-horizon regression formulations, where robustness and temporal sensitivity outweigh model scale or long-window classification.

This work investigates hybrid artificial intelligence models for short-horizon heart-rate time-series modeling. The study formulates a constrained beat-trend prediction task and evaluates a rolling adaptive RR baseline [7], a single-branch recurrent sequence model [4], and a hybrid recurrent model enhanced with standardized physiological variability descriptors [6], [7]. By emphasizing beat-level temporal precision, physiological plausibility, and noise robustness, the proposed framework supports the development of compact and reproducible cardiac AI pipelines suitable for real-time monitoring systems.

II. Data and Preprocessing

A. Dataset Description

The experimental evaluation in this study is conducted using real-world electrocardiographic recordings obtained from the MIT-BIH Normal Sinus Rhythm Database (NSRDB), which is part of the broader MIT-BIH PhysioNet collection [8], [9]. The database contains long-term ECG recordings acquired from subjects exhibiting normal sinus rhythm, making it suitable for the analysis of physiological heart rate dynamics without dominant pathological disturbances.

Each record consists of digitized ECG signals sampled at a frequency of 128 Hz, accompanied by expert-annotated R-peak locations. The availability of precise beat annotations enables reliable extraction of beat-to-beat intervals and subsequent construction of short-term heart rate time series. In this work, a subset of NSRDB records is used to focus on methodological validation, which is consistent with the objectives of short-horizon prediction and hybrid model evaluation.

B. ECG Signal Processing and R-peak Annotation

Let $x(t)$ denote the discrete-time ECG signal sampled at frequency f_s. For each ECG record, R-peak locations are obtained directly from the provided annotation files, which have been validated and widely used in the literature [9]. This approach avoids potential errors introduced by automatic QRS detection algorithms and ensures reproducibility of the preprocessing pipeline.

Given a sequence of annotated R -peak sample indices

$$\{r_1, r_2, \dots, r_N\},$$

the corresponding R-peak times in seconds are computed as

$$t_i = \frac{r_i}{f_s}$$

This representation forms the basis for deriving RR interval and heart rate series used in subsequent modeling stages.

C. RR Interval and Heart Rate Time Series Construction

The beat-to-beat RR interval series is defined as:

$$RR_i = t_{i+1} - t_i, i = 1, \dots, N-1.$$

D. Windowing and Forecasting Setup

To formulate the short-horizon forecasting task, the cardio series is segmented into overlapping windows using a sliding window approach. For each sample, an input window of length L peaks is used to predict the subsequent H peaks:

$$\mathbf{x}_i = \{R_i, R_{i+1}, \dots, R_{i+L-1}\},$$
$$\mathbf{y}_i = \{R_{i+L}, R_{i+L+1}, \dots, R_{i+L+H-1}\}$$

In this study, the input length is fixed to $L = 30$ peaks, while the prediction horizon is set to $H = 10$ peaks, reflecting a short-horizon forecasting scenario. This formulation is motivated by practical monitoring applications, where near-future cardiac behavior is of greater relevance than long-term prediction [5].

The segmentation is performed chronologically without shuffling, and the resulting samples are divided into training, validation, and test sets using a time-based split. This strategy prevents information leakage across temporal boundaries and ensures that model evaluation reflects realistic deployment conditions.

E. Feature Normalization and Physiological Descriptors

For model training, heart rate input sequences are standardized using statistics computed from the training set only. Given the training mean μ and standard deviation σ, normalized inputs are obtained as

$$\tilde{R}R_i = \frac{RR_i - \mu}{\sigma}.$$

In addition to raw heart rate sequences, a small set of physiologically motivated descriptors is computed for each input window. These features include basic time-domain and short-term variability measures derived from the corresponding RR intervals, such as mean RR interval, standard deviation of RR intervals, and root mean square of successive differences. Such descriptors provide complementary information about shortterm autonomic regulation and have been shown to enhance robustness when combined with data-driven models [6], [11].

The preprocessing pipeline transforms raw ECG recordings into normalized short-term heart rate sequences and associated physiological descriptors suitable for hybrid modeling. By relying on expert-annotated R-peaks and beat-level segmentation, the proposed approach preserves physiologically meaningful temporal dynamics while enabling systematic evaluation of short-horizon prediction models.

III. Methodology

Let the beat-level heart rate time series derived from an ECG recording be denoted as:

$$\mathrm{T} = \{HR_1, HR_2, \dots, HR_T\},$$

where $\mathrm{T}_i \in \mathbb{R}$ represents the time in seconds of occurrence of the i-th R-peak t. The objective of this study is to address a short-horizon forecasting problem, in which the goal is to predict the future in which the moment of occurrence of future R-peaks is estimated based on a limited number of preceding events. Given an input window of length L, the prediction task is defined as:

$$\hat{\mathbf{y}}_t = f_\theta(\mathbf{x}_t),$$

Where:

$$\mathbf{x}_t = \{RR_t, HR_{t+1}, \dots, HR_{t+L-1}\}, \hat{\mathbf{y}}_t = \{\hat{R}R_{t+L}, \dots, \hat{H}R_{t+L+H-1}\},$$

and $f_\theta(\cdot)$ denotes a parameterized prediction model with parameters θ. In this work, the input window length is fixed to $L = 30$ beats, and the prediction horizon is set to $H = 10$ beats. This formulation reflects practical monitoring scenarios, where near-future cardiac behavior is of primary interest, and aligns with prior studies on short-term physiological forecasting [5], [12].

A. Baseline forecasting model

As a reference, a purely data-driven baseline model is implemented to establish a lower bound on prediction performance. The baseline model operates directly on normalized heart rate sequences and does not incorporate explicit physiological descriptors.

The baseline predictor is implemented using a recurrent neural network with long short-term memory (LSTM) units, which are well suited for modeling temporal dependencies in sequential data [13]. Given the input sequence $\tilde{\mathbf{x}}_t$, obtained through z-score normalization, the LSTM computes a hidden representation:

$$\mathbf{h}_t = \mathrm{LSTM}(\tilde{\mathbf{x}}_t),$$

which is subsequently mapped to the predicted output via a fully connected layer:

$$\hat{\mathbf{y}}_t = \mathbf{W}\mathbf{h}_t + \mathbf{b}.$$

The model is trained to minimize the mean squared error (MSE) between predicted and true future heart rate values:

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{H}\sum_{i=1}^{H} \left(\hat{H}R_{t+L+i-1} - HR_{t+L+i-1}\right)^2$$

This baseline provides a strong yet interpretable benchmark for evaluating the added value of hybrid modeling strategies.

B. Physiological Feature Extraction

To incorporate domain knowledge into the prediction process, a set of physiologically meaningful descriptors is computed for each input window. These features are derived from the corresponding RR interval sequence and capture short-term variability characteristics of cardiac dynamics [3], [11]. Given an input window $\mathbf{x}_t$, the associated RR interval sequence is computed as:

$$\mathbf{x}_t = \{RR_t, RR_{t+1}, \ldots, RR_{t+L-1}\}.$$

From this sequence, the following features are extracted:

- mean RR interval,
- standard deviation of RR intervals (SDNN),
- root mean square of successive differences (RMSSD),
- mean heart rate,
- standard deviation of heart rate.

These features form a low-dimensional physiological descriptor vector:

$$\mathbf{z}_t \in \mathbb{R}^5$$

which provides complementary information to the raw heart rate sequence and reflects autonomic regulation mechanisms.

C. Hybrid Model Architecture

The proposed hybrid model combines data-driven temporal modeling with physiologically motivated information through a mid-level feature fusion strategy [6], [14]. The architecture consists of two parallel branches: a sequence modeling branch that captures temporal dependencies in cardiac rhythm dynamics, and a physiological feature branch that encodes short-term variability characteristics derived from R–R intervals.

In the sequence branch, the normalized sequence of recent R–R intervals is processed by an LSTM network to produce a latent temporal embedding:

$$\mathbf{h}_t = \mathrm{LSTM}\left(\tilde{RR}_t\right),$$

where $\tilde{RR}_t$ denotes the standardized input window of RR intervals.

In the physiological branch, the feature vector $\tilde{\mathbf{z}}_t$, obtained via standardization, is transformed using a fully connected layer:

$$\mathbf{g}_t = \phi(\mathbf{W}_z\tilde{\mathbf{z}}_t + \mathbf{b}_z)$$

where $\phi(\cdot)$ denotes a nonlinear activation function. The two representations are then concatenated:

$$\mathbf{u}_t = [\mathbf{h}_t \| \mathbf{g}_t]$$

and subsequently processed by dense layers to generate the final prediction of future RR intervals:

$$\hat{\mathbf{y}}_t = \psi(\mathbf{W}_u\mathbf{u}_t + \mathbf{b}_u).$$

This hybrid formulation allows the model to jointly exploit temporal patterns learned from data and structured physiological information, improving robustness in short and noisy time series settings.

D. Training Strategy

Both baseline and hybrid models are trained using the Adam optimization algorithm with early stopping based on validation loss [15]. To avoid information leakage, a chronological split is employed, dividing the dataset into training, validation, and test subsets. All normalization parameters are computed exclusively on the training set and subsequently applied to validation and test data.

Model performance is evaluated using mean absolute error (MAE) and root mean square error (RMSE), expressed in beats per minute, which preserve direct physiological interpretability.

The proposed methodology formulates short-horizon R-peak time prediction as a multistep regression problem and evaluates both purely data-driven and hybrid AI models. By integrating physiologically motivated descriptors with neural network-based temporal modeling, the hybrid approach aims to achieve improved robustness and interpretability in short-term RR time series analysis.

IV. Experimental Setup

A. Dataset Overview

All experiments use ECG records from the MIT-BIH Normal Sinus Rhythm Database as described in Section 2. The models are trained and evaluated on chronologically segmented R–R interval sequences for short-horizon regression tasks .

B. Signal Preprocessing and Segmentation

The ECG signals are first bandpass-filtered using a zero-phase Butterworth filter to suppress baseline wander and high-frequency noise. R-peak annotations provided by the database are used to derive RR intervals, defined as the temporal differences between consecutive R-peaks.

Based on the sequence of R-peaks, the inter-beat intervals (RR intervals) are computed as the difference between the occurrence times of two consecutive R-peaks. The resulting RR intervals form the input time series for the predictive models.

The resulting RR series are segmented into overlapping windows of fixed length $L = 30$ beats, followed by a prediction horizon of $H = 10$ beats. A sliding-window approach with a stride of one beat is applied, yielding a large number of training samples while preserving temporal continuity. Each input window is normalized using z-score normalization. Both metrics are computed over the entire prediction horizon and averaged across all test samples. Errors are reported in bpm, rather than percentages, to preserve clinical relevance and facilitate comparison with existing studies [18].

C. Experimental Protocol

Each experiment is repeated independently for each ECG record. Reported results correspond to the mean performance across all test segments for a given record. To ensure reproducibility, random seeds are fixed for weight initialization and data shuffling. In addition to quantitative evaluation, qualitative analysis is performed by visually comparing predicted and ground-truth heart rate trajectories over selected short segments, highlighting the models' ability to capture rapid local fluctuations.

V. Results and Discussion

This section presents and analyzes the experimental results obtained from the proposed short-horizon heart rate (HR) prediction framework. The evaluation focuses on the ability of recurrent neural architectures to predict future RR intervals and the appearance times of upcoming R-peaks, formulated as a 10 -beat-ahead regression task on chronologically segmented, expert-annotated R-R interval sequences extracted from ECG recordings.

A. ECG-Derived R-Peaks Dynamics

Figure 1 illustrates a representative 6-second ECG segment from record 16 265 with annotated R-peaks. The signal exhibits stable morphology and clearly identifiable ventricular depolarization complexes, confirming the reliability of the provided annotations and pre-processing pipeline.

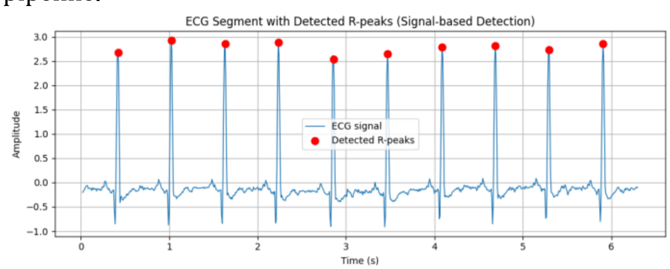


Figure 1. *ECG signal segment with detected R-peak locations.*

Based on the detected R-peaks, a sequence of R–R intervals was constructed by computing the temporal differences between consecutive peaks. The first 50 R–R intervals used as model input are shown in Figure 2. The sequence demonstrates moderate beat-to-beat variability, reflecting natural short-term fluctuations in cardiac rhythm even under normal sinus conditions. This non-stationary behavior motivates the use of adaptive and nonlinear prediction models rather than simple constant-rate assumptions.

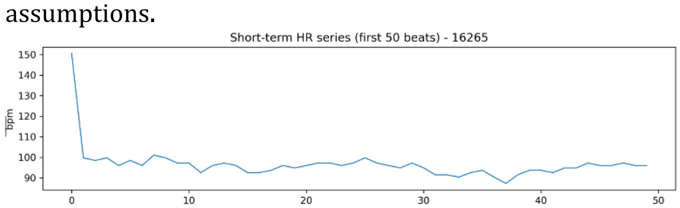


Figure 2. *Short-term beat-to-beat heart rate trajectory for the first 50 beats.*

B. Short-Horizon HR Prediction Performance

The prediction task was formulated as a multistep forecasting problem, where an input window of 30 consecutive R–R intervals was used to predict the occurrence times of the next 10 cardiac cycles. Two approaches were evaluated:

- an adaptive rolling R–R baseline model, and
- an LSTM-based recurrent neural network trained to predict future R–R intervals.

Figure 3 presents a representative example of short-horizon R-peak time prediction (adaptive rolling R–R baseline model). The true future R-peaks are shown together with the predicted peak locations obtained from both models. For visualization purposes, predicted times are projected onto the ECG waveform by mapping them to the nearest detected peak amplitudes.

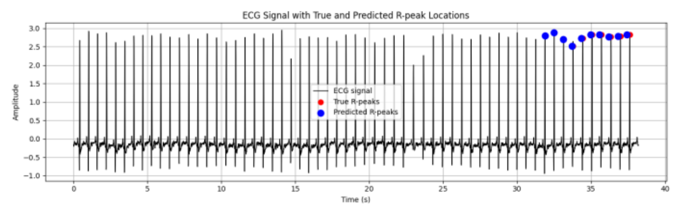


Figure 3. *True and predicted future R-peak locations for an ECG segment (baseline model).*

Figure 4 compares the true future R-peak locations with the predictions obtained from the adaptive rolling RR baseline and the LSTM-based model for a representative ECG segment. While both approaches successfully follow the general rhythm trend, the adaptive RR model exhibits increasing temporal deviation as the prediction horizon advances. In contrast, the LSTM model produces predictions that remain closely aligned with the true R-peak positions, particularly for the later beats. This visual comparison confirms the quantitative results, demonstrating the superior robustness of the LSTM model in short-horizon R-peak time forecasting.

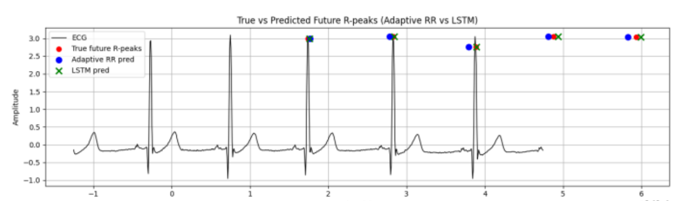


Figure 4. *Comparison of true and predicted future R-peak locations using Adaptive RR and Hybrid LSTM models.*

C. Model Comparison

Table I presents a quantitative comparison of three forecasting approaches for short-horizon R-peak time prediction over a 10-beat horizon: an adaptive rolling RR baseline, a data-driven LSTM model, and a hybrid LSTM model augmented with physiological variability features. Prediction accuracy is assessed using mean absolute error (MAE) and root mean square error (RMSE), expressed in milliseconds, providing a direct measure of temporal prediction precision.

The adaptive rolling RR baseline achieved an MAE of 39.06 ms and an RMSE of 42.65 ms. These results indicate reasonable short-term extrapolation performance under normal sinus rhythm conditions, but also reflect the inherent limitations of relying primarily on the most recent RR interval. As expected, this simple approach is prone to gradual error accumulation in multistep prediction scenarios.

Introducing a purely data-driven LSTM model leads to a moderate improvement in prediction accuracy, reducing the MAE to 34.96 ms and the RMSE to 37.02 ms. This improvement demonstrates the benefit of recurrent neural modeling, which can exploit temporal dependencies across multiple preceding cardiac cycles rather than extrapolating from a single interval.

The hybrid LSTM model, which integrates temporal representations learned by the LSTM with physiologically motivated RR variability descriptors, achieves a substantial further reduction in prediction error. Specifically, the hybrid approach attains an MAE of 15.61 ms and an RMSE of 24.10 ms, representing more than a twofold improvement in MAE compared to the adaptive baseline. This pronounced performance gain highlights the complementary role of physiological features in constraining and stabilizing the learning process, particularly in short and noisy cardiac time series.

The comparison confirms that while recurrent neural networks improve short-horizon R-peak time prediction relative to simple adaptive baselines, the integration of physiological domain knowledge provides an additional and significant benefit. The hybrid architecture demonstrates enhanced robustness and precision, especially in scenarios characterized by limited temporal context, making it well suited for real-time cardiac monitoring applications. The hybrid model demonstrates enhanced temporal precision and improved short-horizon stability compared to baseline extrapolation and standalone LSTM trend predictors, while the overall beat-trend modeling remains clinically plausible and computationally compact.

In relative terms, the hybrid LSTM model reduces the mean absolute error by 60.0% compared to the adaptive rolling RR baseline and by 55.3% compared to the purely data-driven LSTM model. In addition, the hybrid approach achieves a 43.5% reduction in RMSE relative to the baseline, confirming its superior robustness in short-horizon R-peak time prediction.

Table I. *Model comparison for short-horizon R-peak time prediction (H = 10 beats)*

Model	**MAE (ms)**	**RMSE (ms)**
Baseline adaptive RR	39.06	42.65
LSTM	34.96	37.02
Hybrid LSTM + physiological features	15.61	24.10

Across all evaluated test segments, both the adaptive rolling RR baseline and the LSTM-based model achieve prediction errors within a physiologically acceptable range for short-horizon R-peak time forecasting. However, a clear performance difference is observed between the two approaches. While the adaptive RR model provides reasonable short-term extrapolation, the LSTM model consistently achieves substantially lower errors, particularly for the first few predicted beats.

Horizon-wise analysis reveals that the LSTM model exhibits significantly improved accuracy already at the earliest prediction steps (H = 1–3), indicating a higher sensitivity to immediate temporal context. This behavior reflects the ability of recurrent neural networks to exploit dependencies across multiple previous cardiac cycles, rather than relying solely on the most recent RR interval. In contrast, the adaptive RR baseline is more susceptible to gradual error accumulation as the prediction horizon increases.

D. Feature Fusion Implications for Short-Context Cardio Time Series

Previous studies have primarily focused either on classical HRV analysis using statistical descriptors [3], [18] or on deep learning approaches trained on long-duration recordings [8], [20]. In contrast, the present study explicitly targets short-horizon prediction, where limited temporal context poses significant modeling challenges.

The obtained results align with recent trends toward physiology-aware and physics-informed machine learning in biomedical applications. Unlike large transformer-based architectures, which typically require extensive training data, the proposed hybrid approach remains computationally lightweight and suitable for real-time cardiovascular monitoring systems.

E. Limitations and Future Work

Several limitations should be acknowledged. First, the experiments were conducted exclusively on normal sinus rhythm recordings; extending the analysis to pathological conditions may reveal additional challenges and opportunities for improvement. Second, the current study focuses solely on ECG-derived R-peak sequences.

Future work will explore adaptive and physiology-aware feature fusion strategies, as well as attention-based mechanisms to dynamically emphasize relevant temporal patterns. Additionally, multimodal extensions incorporating complementary signals, such as photoplethysmography or respiration, may further enhance robustness and prediction accuracy.

VI. Conclusion

This study investigated short-horizon prediction of future R-peak occurrence times using sequential models applied to ECG-derived R–R interval series. Focusing on limited observation windows and a 10-beat prediction horizon, the work addressed a practically relevant yet challenging problem characterized by strong non-stationarity and beat-to-beat variability.

A comparative evaluation between an adaptive rolling RR baseline and an LSTM-based recurrent neural model demonstrated that both approaches yield physiologically meaningful prediction errors, while the LSTM model achieves a substantial reduction in MAE and RMSE. These results highlight the advantage of recurrent temporal modeling in capturing short-term rhythm dynamics and mitigating error accumulation in multistep prediction tasks.

Beyond improved accuracy, the proposed framework emphasizes computational efficiency and interpretability, making it suitable for real-time cardiovascular monitoring applications. The presented methodology provides a reproducible benchmark for short-horizon R-peak time forecasting and supports the development of next-generation physiology-aware artificial intelligence systems for continuous cardiac monitoring.

ACKNOWLEDGMENT

This research was funded by the National Science Fund of Bulgaria (scientific project “Research on AI-based methods for modeling, analyzing and forecasting biomedical time series by applying a digital twin concept”), Grant Number КП-06-Н97/12, 10 December 2025.

References

[1] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, Heart rate variability: Standards of measurement, physiological interpretation, and clinical use, Circulation, vol. 93, no. 5, 1996, pp. 1043–1065.
[2] Clifford, G.D., Azuaje, F., McSharry, P., Advanced Methods and Tools for ECG Data Analysis, London, UK: Artech House, ch. 2, 2006, pp. 21–52.
[3] Shaffer, F., Ginsberg, J.P., An overview of heart rate variability metrics and norms, Frontiers in Public Health, vol. 5, 2017, pp. 1–17.
[4] Faust, O., et al., Deep learning for healthcare applications based on physiological signals: A review, Computer Methods and Programs in Biomedicine, vol. 161, 2018, pp. 1–13.
[5] Lim, B., Zohren, S., Roberts, S., Time-series forecasting with deep learning: A survey, Philosophical Transactions of the Royal Society A, vol. 379, no. 2194, 2021, pp. 1–25.
[6] Acharya, U.R., et al., Automated characterization of coronary artery disease using nonlinear features extracted from heart rate signals, Biomedical Signal Processing and Control, vol. 31, 2017, pp. 213–224.
[7] Natarajan, A., et al., A wide and deep learning approach for robust prediction of cardiovascular events, Proc. IEEE Conf. on Engineering in Medicine and Biology Society (EMBS), Honolulu, HI, USA, 2018, pp. 273–276.
[8] Goldberger, A.L., et al., *PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals*, Circulation, vol. 101, no. 23, 2000, pp. e215–e220.
[9] Moody, G.B., Mark, R.G., *The MIT-BIH Arrhythmia Database*, IEEE Engineering in Medicine and Biology Magazine, vol. 20, no. 3, 2001, pp. 45–50.
[10] Berntson, G.G., et al., *Heart rate variability: Origins, methods,* and interpretive caveats, Psychophysiology, vol. 34, no. 6, 1997, pp. 623–648.
[11] Acharya, U.R., et al., Heart rate variability: A review, Medical & Biological Engineering & Computing, vol. 44, no. 12, 2006, pp. 1031–1051.
[12] Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Time Series Analysis: Forecasting and Control, Hoboken, NJ, USA: Wiley, ch. 3, 2015, pp. 89–132.
[13] Hochreiter, S., Schmidhuber, J., Long short-term memory, Neural Computation, vol. 9, no. 8, 1997, pp. 1735–1780.
[14] Bishop, C.M., Pattern Recognition and Machine Learning, New York, NY, USA: Springer, ch. 14, 2006, pp. 631–660.
[15] Kingma, D.P., Ba, J., Adam: A method for stochastic optimization, Proc. Int. Conf. on Learning Representations (ICLR), San Diego, CA, USA, 2015.
[16] Goldberger, A.L., Amaral, L.A.N., Glass, L., et al., *PhysioBank, PhysioToolkit, and PhysioNet*, Circulation, vol. 101, no. 23, 2000, pp. e215–e220.
[17] Hyndman, R.J., Athanasopoulos, G., *Forecasting: Principles and Practice*, Melbourne, Australia: OTexts, ch. 2, 2021, pp. 45–78.
[18] Shaffer, F., Ginsberg, J.P., *An overview of heart rate variability metrics and norms*, Frontiers in Public Health, vol. 5, 2017, pp. 1–17.

A Framework for Reliable Evaluation under Uncertainty in Large Language Models: From Answer Generation to Interactional Reasoning

Momchil R. Rusev

Independent Researcher

Sofia, Bulgaria

mrusev@gmail.com

Abstract

Large language models (LLMs) are increasingly deployed in technical domains where evaluation tasks are inherently underspecified, context-dependent, and informed by heterogeneous sources of uneven reliability. In such settings, a central limitation is not only factual inaccuracy, but premature evaluative closure: the tendency to produce coherent conclusions before the underlying case has been sufficiently specified and the available evidence properly differentiated.

We argue that this behavior reflects structural pressures toward early coherence under uncertainty, rather than knowledge deficits alone. As a result, LLM outputs may over-weight discursively dominant but evidentially weak claims while under-attending to less salient yet epistemically stronger evidence. To capture this phenomenon, we introduce the Narrative Dominance Gap (NDG), defined as the divergence between discursive prevalence and evidential support.

To address these challenges, we propose a unified framework that reframes LLM interaction from answer generation to collaborative case evaluation. The framework integrates Adaptive Context Elicitation (ACE), which reconstructs missing decision-relevant context, and Context-Aware Epistemic Filtering (CEF), which structures and evaluates heterogeneous evidence relative to that context.

The paper contributes a conceptual account of discourse–evidence divergence, formalizes interactional and epistemic mechanisms for ambiguity-aware evaluation, and proposes an interaction-centered evaluation paradigm focused on problem specification, evidence transparency, and calibrated trust. More broadly, the findings suggest that improving LLM reliability may depend not only on model capabilities, but on how reasoning is structured within interaction, highlighting the role of system architecture and epistemic discipline in supporting robust evaluation under uncertainty. The work is conceptual and methodological, providing a foundation for future empirical validation.

Keywords: large language models (LLMs); human–AI interaction; technical decision support; ambiguity-aware evaluation; narrative dominance gap (NDG); adaptive context elicitation (ACE); context-aware epistemic filtering (CEF); evidence differentiation; evaluation under uncertainty; trust calibration; interaction-centered evaluation; system architecture; reasoning structure

I. INTRODUCTION

Large language models (LLMs) are increasingly integrated into technical decision-support contexts, including engineering, diagnostics, maintenance, procurement, and product evaluation. Their utility is often attributed to their ability to synthesize large volumes of heterogeneous information and present it in a fluent, accessible form.

However, this fluency can obscure a structural limitation. Many real-world technical queries are not well-formed information requests, but underspecified and context-dependent evaluative problems. Questions such as “Is this system reliable?” or “Is this a good choice?” do not admit meaningful answers in the abstract. Their interpretation depends on variables such as system condition, usage context, maintenance history, comparison frame, and the user’s actual decision objective.

In practice, these variables are often omitted. This omission may occur because relevant parameters are taken for granted, difficult to articulate, or not yet known to the user. As a result, initial queries frequently represent incomplete problem specifications rather than fully defined evaluation tasks.

At the same time, technical information environments are characterized by heterogeneous and unevenly reliable sources, including empirical measurements, regulatory standards, expert analysis, experiential reports, and widely circulated narratives. These sources differ in evidential strength, scope, and context dependence, yet they may be presented in similar discursive forms.

Within this setting, LLMs can produce outputs that are rhetorically coherent but epistemically under-specified. The limitation is not necessarily factual inaccuracy, but the compression of heterogeneous evidence and implicit assumptions into a single, fluent evaluative statement. This can obscure distinctions between empirical performance, normative adequacy, contextual suitability, and discursively dominant claims.

More broadly, LLM-based systems can be understood not only as sources of answers, but as evolving cognitive environments that shape how users interpret information, form judgments, and allocate attention over repeated interaction. In this sense, the epistemic risks discussed here extend beyond individual responses and relate to the structural properties of the interaction context itself.

A key contributing factor to this behavior is the sensitivity of LLMs to patterns present in their training data. Prior work suggests that model outputs are influenced not only by factual content, but also by distributional properties such as frequency, salience, and representational density. While these properties do not determine epistemic validity, they can shape which claims are more likely to be generated.

To capture this phenomenon, we introduce the **Narrative Dominance Gap (NDG)**, defined as the potential divergence between the discursive prevalence of an evaluative claim and the strength of its empirical or normative support. NDG is not introduced as an empirically validated mechanism, but as a conceptual tool for analyzing situations in which discursive prominence may exceed evidential grounding.

From this perspective, a central limitation of current LLM use in technical domains is not only epistemic, but interactional. Systems typically respond before the relevant case context has been sufficiently reconstructed. As a result, the model may produce an evaluation based on an implicit interpretation of the problem that does not correspond to the user's actual decision context.

We therefore propose a reframing: from answer generation to **collaborative case evaluation under uncertainty**. In this reframing, the system should treat the initial query as a partial specification and actively support the reconstruction of the underlying problem before evaluation takes place.

To operationalize this shift, we introduce two complementary mechanisms. **Adaptive Context Elicitation (ACE)** supports the identification and elicitation of missing, decision-relevant variables through structured interaction. **Context-Aware Epistemic Filtering (CEF)** supports the differentiation, alignment, and evaluation of heterogeneous evidence relative to the reconstructed context.

Together, NDG, ACE, and CEF define an interactional–epistemic framework for ambiguity-aware technical evaluation. NDG functions as a diagnostic lens, ACE as an interactional mechanism for context reconstruction, and CEF as an evaluative mechanism for structuring evidence.

A further observation motivating this work is that the limitations of LLM-based evaluation in technical domains may not be reducible to factual inaccuracy alone. When operating under incomplete or underspecified problem formulations, LLM systems can exhibit patterns of premature evaluative closure that are structurally similar to those observed in everyday human reasoning under uncertainty. In such cases, models may implicitly treat partially specified context as sufficient, gravitate toward discursively dominant or familiar explanations, underweight less salient but technically critical evidence, and converge on coherent conclusions before the underlying case has been adequately reconstructed. Importantly, this behavior does not necessarily arise from error in the narrow sense, but from the interaction between training on large-scale human-generated discourse and optimization toward fluent, coherent, and timely responses. From this perspective, the challenge is not only to improve correctness, but to redesign the interactional process such that context reconstruction and evidence differentiation precede evaluative closure. This shift reframes the problem from answer correctness to evaluation structure. More broadly, this suggests that improving the reliability of LLM-based systems may require not only better models, but also improvements in how reasoning itself is structured — both in human–AI interaction and within the processes through which models generate conclusions. This implies that reliability is not solely a property of model outputs, but of the interactional structure through which evaluation is constructed. In this sense, NDG is introduced not only as a lens on discursive bias, but as a diagnostic indicator of early closure under uncertainty.

The contributions of this paper are threefold. First, it introduces NDG as a conceptual lens for analyzing discursively influenced evaluation in LLM outputs. Second, it formalizes ACE and CEF as mechanisms for handling underspecified queries and heterogeneous evidence. Third, it outlines an interaction-centered evaluation paradigm in which system performance is assessed not only in terms of answer correctness, but also in terms of problem specification quality, transparency of evidential structure, trust calibration, and support for human judgment.

The paper is conceptual and methodological in nature. Its goal is to articulate design principles for more reliable human–LLM interaction in technical domains and to provide a foundation for future empirical work.

The proposed framework is not a standalone conceptual construct but emerges from a series of studies spanning economic theory, engineering analysis, and human–machine interaction.

II. Related Work and Positioning

A.2.1 Reliability, truthfulness, and epistemic limits of LLMs

A growing body of research has examined the reliability of large language models (LLMs), particularly their tendency to produce fluent but epistemically uncertain outputs. These issues are commonly discussed in terms of hallucination, truthfulness limitations, and overconfidence (Bender et al., 2021; Lin et al., 2022). Empirical benchmarks such as TruthfulQA (Lin et al., 2022) suggest that models can reproduce widely circulated misconceptions or misleading claims, especially when such claims are prevalent in training data.

Taken together, these findings indicate that LLM outputs are influenced not only by factual knowledge, but also by distributional properties of the data on which they are trained, including frequency, salience, and representational density of claims.

While existing work does not directly establish a causal relationship between discursive prevalence and evidential validity, it suggests that widely repeated or salient claims may be more likely to be generated, regardless of their epistemic strength.

In this sense, model outputs can reflect statistical regularities in discourse alongside, and sometimes independently from, structured evidence. The resulting limitation is not necessarily factual fabrication, but a form of epistemic compression: the collapse of heterogeneous evidence, uncertainty, and context-dependence into a single coherent narrative.

This observation is consistent with findings in cognitive psychology, such as the "illusory truth effect," where repeated exposure increases perceived truthfulness (Hasher et al., 1977; Fazio et al., 2015). Importantly, this parallel should be understood as an analogy rather than a shared mechanism. LLMs do not form beliefs; however, their training on large-scale human-generated corpora may encode statistical regularities that functionally resemble exposure-driven prominence.

Research on bias in language models further supports this perspective by showing that outputs are shaped by distributional features such as frequency, salience, and social prominence (Bender et al., 2021; Gehman et al., 2020). As a result, models may preferentially reproduce dominant or widely circulated narratives, even when these are weakly supported, context-dependent, or contested.

The arguments in this section draw on heterogeneous sources, including empirical benchmarks, theoretical analyses, and analogies from cognitive psychology. These sources differ in epistemic status and should not be interpreted as providing uniform evidential support. Instead, they are used as converging indications that motivate the present framework rather than as direct validation of its components.

The present work builds on these insights by introducing the Narrative Dominance Gap (NDG) as a theoretical construct. NDG provides an interpretive lens for analyzing situations in which discursive prominence and evidential support diverge, rather than constituting a directly empirically established mechanism.

B.2.2 Retrieval augmentation and evidence grounding

Retrieval-augmented generation (RAG) and related approaches aim to improve model reliability by grounding outputs in external information sources (Lewis et al., 2020). By incorporating retrieved documents into the generation process, such systems reduce reliance on parametric memory and can improve factual accuracy in well-defined tasks.

However, retrieval alone does not resolve how heterogeneous information should be interpreted, compared, or weighted within a specific decision context. Even when relevant information is available, systems must still determine which sources are epistemically stronger, how different types of evidence should be evaluated, and how evidence should be aligned with the specific case at hand.

Insights from research on uncertainty and decision-making under incomplete information highlight the importance of distinguishing between different types of evidence and their respective reliability (Gigerenzer & Gaissmaier, 2011; Kahneman, 2011). Failure to differentiate between evidence sources can lead to overconfident but poorly grounded judgments.

Even when relevant documents are retrieved, the system must still determine how to distinguish between quantitative evidence, normative standards, contextual information, and discursive signals. Access to information does not by itself provide an evidence hierarchy, nor does it prevent the collapse of epistemically distinct layers into a simplified evaluative narrative.

Current RAG-based systems typically treat retrieved information as relatively homogeneous input, without explicitly representing differences between empirical measurements, normative criteria, expert analysis, experiential reports, and discursive signals. As a result, heterogeneous evidence may still be collapsed into a single narrative output, reproducing the same form of epistemic compression observed in non-retrieval settings.

Moreover, retrieval does not address the problem of underspecified queries. A system may access high-quality information, yet still apply it to an implicitly assumed problem formulation that does not match the user's actual context or intent.

The Context-Aware Epistemic Filtering (CEF) mechanism proposed in this work addresses these limitations by structuring how evidence is differentiated, aligned, and weighted relative to a reconstructed case context. Rather than treating information access as sufficient, CEF reframes the problem as one of epistemic organization: making explicit the type, strength, relevance, and conditional validity of different evidence sources.

Why retrieval is insufficient? Retrieval improves information access, but it does not by itself determine how retrieved material should be classified, weighted, or aligned with the case context. A system may retrieve both empirical test results and subjective commentary, yet still collapse them into a single evaluative narrative. The central problem is therefore not only access to information, but control over the epistemic organization of information. CEF addresses this by requiring explicit differentiation between empirical findings, normative thresholds, expert interpretation, experiential reports, and diffusion signals.

C.2.3 Ambiguity and conversational clarification

Research in conversational information retrieval and ambiguity-aware question answering has consistently shown that many user queries are underspecified and require clarification (Zamani and Craswell, 2020; Aliannejadi et al., 2021). In these approaches, clarification is typically introduced as a mechanism for improving retrieval effectiveness or answer relevance.

More recent work on human–LLM interaction suggests a deeper role of interaction in shaping outcomes. In particular, interaction-centered evaluation demonstrates that user behavior is sensitive to contextual features of the interaction, including framing, prior exposure, and domain context (Zhou et al., 2025). Even when model outputs remain unchanged, variations in interactional conditions can lead to systematically different reliance decisions.

These findings indicate that ambiguity is not only a property of the input, but also of the interactional conditions under which the input is interpreted. Despite this, most current approaches treat clarification as an auxiliary conversational strategy rather than as a prerequisite for valid reasoning.

This problem is amplified by the fact that evaluative language often compresses multiple analytical dimensions into a single expression. Terms such as "reliable," "effective," or "good" may simultaneously refer to performance, controllability, compliance, durability, or suitability for a specific use case. Clarification is therefore necessary not only because the query is incomplete, but because evaluative terms themselves are analytically decomposable.

In technical domains, this limitation becomes critical, as small differences in context specification can fundamentally alter the validity of an evaluation. To address this gap, we introduce Adaptive Context Elicitation (ACE) as a core component of LLM-based technical decision support.

Unlike generic clarification strategies, ACE treats problem specification as an explicit and structured process aimed at reducing epistemic uncertainty before evaluation takes place. It is goal-directed, focusing on variables that materially affect outcomes; domain-sensitive, guided by expected parameter structures; and uncertainty-driven, prioritizing questions that maximally reduce ambiguity.

In this sense, ACE reframes clarification from a conversational convenience into a necessary step in constructing a valid and contextually grounded problem representation.

D.2.4 Explainability, trust, and human–AI collaboration

A growing body of research emphasizes that effective human–AI collaboration depends not only on model accuracy, but also on interaction design, trust calibration, and decision support (Doshi-Velez and Kim, 2017; Lee and See, 2004; Amershi et al., 2019). More recent work highlights that system effectiveness emerges from the structure of interaction, feedback mechanisms, and the distribution of roles between humans and AI systems (Vats et al., 2025).

Within this paradigm, trust is understood as a dynamic outcome of interaction rather than a static property of a model. Emerging evaluation frameworks therefore shift attention from output correctness to human reliance behavior, examining how users act upon model outputs in decision-making contexts. Research further suggests that reliance is influenced not only by correctness, but also by how uncertainty is communicated (Kim et al., 2024).

Interaction-centered studies provide evidence that reliance is sensitive to non-evidential cues such as framing, perceived competence, prior interaction history, and domain context (Zhou et al., 2025). These findings are complemented by work on linguistic calibration, which indicates that conversational systems can reduce overconfidence through calibrated verbal framing (Mielke et al., 2022). More broadly, decision-support research suggests that effective systems should not merely increase trust, but support appropriately calibrated reliance (Buçinca et al., 2021).

From the perspective of the present work, these effects can be interpreted as instances in which discourse-level signals influence user judgment independently of evidential strength. The Narrative Dominance Gap (NDG) provides a conceptual lens for analyzing such situations, capturing potential divergence between discourse prominence and evidential support.

In addition, research on anthropomorphism in dialogue systems suggests that conversational form itself can shape perceptions of competence and understanding (Abercrombie et al., 2023), further reinforcing the role of discourse-level features in human–AI interaction.

To address these challenges, the present framework integrates explainability, trust, and reasoning through three complementary components: ACE supports problem transparency, CEF supports evidence transparency, and NDG supports epistemic transparency. Together, these mechanisms aim to align user reliance with both the structure of the problem and the strength of the available evidence.

E.2.5 Positioning of the present work

Recent work on human–AI collaboration increasingly emphasizes interaction-centered evaluation, focusing on how system behavior shapes user decisions rather than solely on model performance (Vats et al., 2025; Zhou et al., 2025). Approaches such as REL-AI further demonstrate that human reliance is systematically influenced by interactional context.

The present work builds on this shift while contributing a complementary perspective. Rather than focusing primarily on behavioral measurement, it introduces a structured interactional–epistemic framework for technical evaluation.

Specifically, the framework addresses three interrelated dimensions that are not jointly structured in existing approaches:

- problem specification under ambiguity,
- epistemic differentiation of heterogeneous evidence,
- and the influence of discourse-level signals on evaluation.

Unlike standard question–answering systems, the framework treats the initial query as incomplete by default and requires explicit reconstruction before evaluation. Unlike retrieval-based approaches, it emphasizes evidence differentiation and contextual alignment rather than information access alone. Unlike traditional explainability methods, it structures reasoning prior to answer generation rather than post hoc. And unlike generic conversational systems, it defines clarification as decision-critical rather than optional.

Most importantly, the framework introduces Narrative Dominance Gap (NDG) as a unifying theoretical construct. While prior work has identified related phenomena—such as bias, overconfidence, and inappropriate reliance—it has not provided a general mechanism that connects them within a single interpretive structure.

NDG offers such a perspective by highlighting how discourse-level properties (e.g., frequency, familiarity, perceived authority) may diverge from evidential grounding. In this sense, interaction-centered findings on human reliance can be interpreted as empirical contexts in which such divergence becomes observable.

Together with Adaptive Context Elicitation (ACE) and Context-Aware Epistemic Filtering (CEF), NDG defines a reframing of LLM-based systems in technical domains: from answer generators to context-sensitive evaluative partners that support the co-construction of problem representations and evidence-based judgments.

III. A Unified Framework for Case-Based Technical Evaluation

A.3.1 From answer generation to case evaluation

The central argument of this paper is that LLM-based systems in technical domains should not be understood primarily as answer generators, but as **participants in case-based evaluation processes under uncertainty**.

In conventional interaction, the system receives a query and produces a response in a single step. This paradigm assumes that the problem is already well-specified. In practice, however, many technical queries are incomplete representations of underlying decision situations. As a result, direct answer generation often leads to premature evaluative closure.

We therefore define **case evaluation** as:

the structured reconstruction and assessment of a concrete problem instance through iterative context elicitation and evidence differentiation, enabling context-aware judgment rather than generalized answer production.

This reframing introduces three key implications:

1. **Problem specification becomes central**
   The system must actively participate in defining the problem before attempting to evaluate it.
2. **Interaction becomes necessary**
   Clarification is not optional but a core functional requirement.
3. **Evaluation replaces assertion**
   Outputs should reflect structured reasoning over heterogeneous evidence, rather than singular authoritative claims.

B.3.2 The role of Narrative Dominance Gap (NDG)

The transition from answer generation to case evaluation is motivated by a central epistemic risk: the tendency of LLMs to reproduce discursively dominant narratives.

We define **Narrative Dominance Gap (NDG)** as:

the divergence between the discursive prevalence of an evaluative claim and the strength of its empirical or normative support.

NDG is not a binary condition, but a graded diagnostic dimension. It becomes pronounced in situations where:

- claims are widely repeated or socially salient,
- but evidence is sparse, conflicting, or strongly context-dependent.

In such cases, models may generate outputs that are:

- rhetorically coherent,
- socially familiar,
- but analytically under-specified.

Importantly, NDG operates even when the model does not hallucinate. The failure mode is not fabrication, but **epistemic compression**—the collapse of heterogeneous evidence into a single narrative representation.

Within the proposed framework, NDG functions as a **trigger condition**:

- high NDG → increased need for clarification (ACE),
- high NDG → stricter evidence differentiation (CEF).

Operational boundary of NDG: A key challenge in operationalizing NDG is distinguishing discursive prevalence from reasonable prior probability. Not every widely held belief reflects narrative dominance; some priors are well grounded in aggregate evidence. NDG becomes diagnostically relevant when the prevalence of a claim is not proportional to its evidential support within the specific case under evaluation. In practical terms, NDG is most clearly present when a system defaults to category-level expectations despite the availability of case-specific contrary evidence. Thus, NDG should not be interpreted as a rejection of priors as such, but as a warning condition indicating that generic expectations may be overriding context-specific evidence. NDG does not refer to the presence of widely held beliefs per se, but to their inappropriate persistence in the presence of stronger, case-specific counterevidence. In practice, this means that NDG is most visible in situations where case-specific evidence is available but not integrated into the evaluation.

C.3.3 Framework overview

The proposed framework integrates three components:

- **NDG (diagnostic layer)**
  Identifies potential divergence between discourse and evidence
- **ACE (interactional layer)**
  Reconstructs missing context through targeted elicitation
- **CEF (evaluative layer)**
  Structures and weights evidence relative to context

Together, these components define a structured interaction loop for technical evaluation.

Conceptual interaction flow

User Query (Q0)
↓
NDG / Ambiguity Signals
↓
ACE (context reconstruction)
↓
Context Representation (C*)
↓
CEF (evidence structuring)
↓
Structured Output (O)

This structure ensures that:

- evaluation does not proceed without sufficient context,
- evidence is not collapsed into a single narrative,
- and conclusions are explicitly conditioned on available information.

A further operational implication of NDG concerns the direction of information-seeking. ACE addresses underspecified user context through structured elicitation. However, a complementary mechanism is required when the evidential base itself is insufficient — that is, when case-specific empirical evidence is absent, sparse, or unverified. In such conditions, the system should explicitly flag the evidential gap rather than defaulting to category-level priors. This distinguishes two types of epistemic insufficiency that require different responses: underspecified user context, addressed by ACE through targeted clarification; and underspecified evidential base, addressed by explicit evidence-seeking or, where retrieval is unavailable, by surfacing uncertainty as a conditional qualifier on any evaluative output. High NDG therefore functions as a trigger not only for ACE, but also for active evidence-seeking prior to the application of CEF.

D.3.4 Adaptive Context Elicitation (ACE)

a) Objective

ACE aims to transform an underspecified query into a **well-defined case representation** by identifying and eliciting missing, decision-relevant variables.

b) Core principle

The system should assume that:

the initial user query is incomplete unless sufficient evidence suggests otherwise.

c) Types of missing variables

ACE focuses on variables that materially affect evaluation:

- **System parameters**
  (age, configuration, condition)
- **Operational context**
  (environment, usage intensity, constraints)
- **Lifecycle factors**
  (maintenance history, degradation, modifications)
- **User intent**
  (purchase decision, diagnosis, comparison)
- **Benchmark frame**
  (what the system is being compared to)

d) Example

Query:
"Is this system reliable?"

ACE identifies missing elements:

- reliable under what conditions?
- for what type of usage?
- compared to which standard or alternative?

e) Adaptive behavior

ACE is:

- **selective**, asking only relevant questions,
- **goal-directed**, focusing on decision-critical variables,
- **adaptive**, terminating when sufficient clarity is achieved.

Adaptive behavior ACE is selective, goal-directed, and uncertainty-driven. However, the implementation of the ACE framework must remain proportional to the complexity and criticality of the case to avoid unnecessary conversational friction in routine tasks.

While high-stakes technical evaluations necessitate rigorous context reconstruction, the system should prioritize directness for low-stakes or purely informational queries. Effective human–AI collaboration thus relies on a dynamic balance, ensuring that the depth of elicitation aligns with the epistemic weight of the underlying problem.

E.3.5 Context representation (C)*

The output of ACE is a structured context representation:

C* = {parameters, constraints, intent, comparison frame}

This representation is central to the framework because it:

- defines what is being evaluated,
- determines which evidence is relevant,
- and conditions how that evidence should be interpreted.

Without a well-defined C*, subsequent evaluation remains inherently unstable and potentially misleading.

F.3.6 Context-Aware Epistemic Filtering (CEF)

a) Objective

CEF organizes heterogeneous information into a **structured and transparent evaluative output**.

b) Core principle

Not all information is equally relevant, and not all relevant information is equally reliable.

c) Evidence dimensions

CEF distinguishes between multiple types of evidence:

1. **Empirical evidence**
(measurements, test results)
2. **Normative criteria**
(standards, regulatory requirements)
3. **Expert analysis**
(engineering interpretation, technical assessments)
4. **Experiential reports**
(user observations, field reports)
5. **Diffusion signals**
(popularity, widespread usage, reputation)

Experiential evidence and evaluative overreach. Subjective observations may provide important signals about usability, controllability, or operator experience, but they are not epistemically equivalent to measured performance parameters. Experiential reports should therefore be treated as context-bound observations whose relevance depends on the task, benchmark, and degree of source independence. In technical evaluation, a common failure mode is to overgeneralize subjective impressions into system-level deficiencies. CEF prevents this by preserving experiential reports as a distinct evidential layer rather than allowing them to function as unmarked substitutes for empirical performance evidence.

These evidence categories are not epistemically equivalent. Quantitative and normative evidence will often carry greater evaluative weight in technical domains, while contextual evidence supports interpretation and experiential or discursive evidence may indicate perception, reputation, or recurring field observations without constituting primary proof. CEF therefore requires not only evidence classification, but explicit differentiation of evidential weight.

d) Key operations

CEF performs five key operations:

1. **Source identification**
Classifying evidence according to its epistemic type
2. **Epistemic differentiation**
Distinguishing between levels of reliability and evidential strength
3. **Context alignment**
Filtering and prioritizing evidence based on relevance to C*
4. **Conflict detection**
Preserving disagreement rather than forcing consensus
5. **Structured synthesis**
Presenting evidence as layered rather than merged

Structural dependence of sources. Source independence should be assessed not only at the level of institutional origin, but also at the level of shared methodological, cultural, and contextual constraints. Two sources may come from different publications while sharing the same testing facility, time period, readership, evaluative norms, and discourse environment. In such cases, convergent judgments should not automatically be treated as independent corroboration. CEF therefore requires explicit assessment of structural dependence: the degree to which sources could realistically have arrived at different conclusions given their positions, constraints, and shared background conditions.

G.3.7 Benchmark alignment

Reliable evaluation requires explicit comparison frames. CEF introduces multiple benchmark layers:

- regulatory baseline (minimum compliance),
- historical or class-based benchmark,
- contemporary reference standard,
- and performance ceiling.

This prevents misleading conclusions resulting from implicit or inappropriate comparisons.

H.3.8 System-state and lifecycle differentiation

CEF explicitly distinguishes between:

- **design capability**, and
- **in-use condition**.

This prevents conflating:

- inherent design limitations,
- and degradation due to usage or maintenance.

Public evaluations of technical systems often arise from encounters with systems in degraded, modified, or poorly maintained states. For this reason, CEF distinguishes between design-level capability and typical in-use condition. The former concerns the intended or specification-level performance of a system, whereas the latter concerns the practical performance of instances encountered in operational environments.

Historical systems in present-day evaluation.

A further source of evaluative distortion arises when historically specified systems are assessed as if their present-day instances remained materially identical to their original tested condition. In practice, historical vehicles or technical systems are often encountered with replacement consumables, altered maintenance histories, updated materials, and changed operating environments. Reliable evaluation therefore requires explicit separation between design-level capability, historically tested configuration, and present-day in-use condition. Without this distinction, systems may be judged either too harshly or too favorably on the basis of an unstated and unstable comparison frame.

I.3.9 NDG-aware evaluation

CEF explicitly evaluates whether conclusions are:

- supported by structured evidence, or
- driven by discursive dominance.

When NDG is high:

- narrative claims are explicitly labeled as such,
- uncertainty is surfaced,
- and conclusions are presented as conditional.

J.3.10 Output structure

Final outputs are structured into:

1. **Context summary (C*)**
2. **Evidence layers (separated by type)**
3. **Trade-offs and constraints**
4. **Conditional conclusions**
5. **Uncertainty and NDG indicators**

Within this structure, outputs should distinguish between verified findings, plausible inferences, discursively prevalent claims, and unresolved uncertainties. This prevents rhetorically smooth conclusions from obscuring differences in evidential status.

This replaces single authoritative answers with **transparent evaluative reasoning**.

K.3.11 Integration

ACE and CEF are interdependent:

- ACE defines *what is being evaluated*,
- CEF defines *how it is evaluated*.

NDG connects both by:

- signaling when additional clarification is required,
- and constraining how conclusions should be formed.

Together, NDG, ACE, and CEF define a coherent **interactional–epistemic pipeline** for ambiguity-aware technical evaluation.

IV. Operationalization and Evaluation Paradigm

A.4.1 From conceptual framework to operational system

The proposed NDG–ACE–CEF framework is intentionally conceptual, but it is designed to support practical implementation in LLM-based systems.

Importantly, the framework does not require a fundamentally new model architecture. Instead, it can be operationalized through:

- structured prompting strategies,
- multi-step interaction design,
- modular reasoning pipelines,
- and retrieval-augmented systems with explicitly labeled evidence layers.

The key requirement is not architectural innovation, but **control over how reasoning is structured, sequenced, and exposed to the user.**

At a high level, an operational system implementing the framework would consist of the following stages:

1. Detection of ambiguity and NDG signals
2. Adaptive Context Elicitation (ACE)
3. Construction of a structured context representation (C*)
4. Evidence retrieval and classification
5. Context-Aware Epistemic Filtering (CEF)
6. Structured output generation

These stages can be implemented either as explicit system modules or as coordinated prompting strategies within a single model.

B.4.2 Ambiguity and NDG detection

a) Conceptual role

Ambiguity and NDG detection serve as the entry point of the framework. Their role is not to fully resolve uncertainty, but to determine whether **additional interaction and structured evaluation are required**.

b) Operational signals

In practice, ambiguity and NDG can be approximated using a set of heuristic or learned indicators:

- **Missing parameter signals**
  Absence of key domain variables required for evaluation
- **Generic evaluative language**
  Use of underspecified terms such as "good," "reliable," or "efficient"
- **Intent uncertainty**
  Lack of a clearly defined decision context (e.g., purchase vs. diagnosis)
- **Knowledge variance signals**
  High disagreement across available sources
- **Discursive dominance signals**
  High-frequency claims with low specificity or weak evidential grounding

c) Formal approximation

These signals can be combined into a scoring function:

$A(Q) = \Sigma\, w_i \cdot f_i(Q)$

Where:

- f_i represent ambiguity or NDG indicators
- w_i represent domain-dependent weights

The purpose of this function is not precise quantification, but **triggering ACE and CEF when ambiguity or NDG is sufficiently high**.

C.4.3 Adaptive Context Elicitation (ACE)

a) Objective

ACE transforms an underspecified query into a **well-defined case representation** by identifying and eliciting missing, decision-relevant variables.

b) Core principle

The system should assume that:

the initial user query is incomplete unless sufficient evidence suggests otherwise.

c) Question selection policy

Clarification can be modeled as an optimization problem:

$Q^* = \text{argmax}\ (\text{Information Gain} / \text{Interaction Cost})$

Where:

- Information Gain represents expected reduction in decision-relevant uncertainty
- Interaction Cost represents cognitive and interaction burden

d) Practical interpretation

In practice, this corresponds to:

- prioritizing variables with high decision impact,
- avoiding redundant or low-value questions,
- and stopping when marginal information gain becomes negligible.

e) Domain-specific templates

ACE relies on structured templates adapted to specific domains.

For example, in technical product evaluation (e.g., vehicles), relevant variables may include:

- production year,
- usage intensity or mileage,
- maintenance history,
- environmental conditions,
- intended use,
- and user intent (evaluation, purchase, diagnosis).

These templates guide the system in identifying which parameters are critical for evaluation.

f) Adaptive sequencing

ACE operates dynamically:

- it prioritizes high-impact variables first,
- adapts questioning based on user responses,
- and terminates when sufficient contextual clarity is achieved.

D.4.4 Context representation (C)*

The output of ACE is a structured context representation:

C* = {parameters, constraints, intent, comparison frame}

This representation is central to the framework because it:

- defines what is being evaluated,
- determines which evidence is relevant,
- and conditions how that evidence should be interpreted.

Without a well-defined C*, subsequent evaluation remains unstable and potentially misleading.

E.4.5 Context-Aware Epistemic Filtering (CEF)

This section translates the conceptual structure of CEF into operational terms.

a) Objective

CEF structures heterogeneous information into a context-sensitive evaluative output.

b) Evidence categories

CEF distinguishes between multiple types of evidence:

1. Empirical evidence (measurements, test results)
2. Normative criteria (standards, regulations)
3. Expert analysis (engineering interpretation)
4. Experiential reports (user observations)
5. Diffusion signals (popularity, widespread usage)

c) Key operations

In operational terms, CEF performs the following operations:

1. **Source identification**
   Classifying evidence according to its epistemic type
2. **Epistemic differentiation**
   Distinguishing between levels of reliability and evidential strength
3. **Context alignment**
   Filtering and prioritizing evidence based on relevance to C*
4. **Conflict handling**
   Preserving disagreement rather than forcing consensus
5. **Structured synthesis**
   Presenting evidence as layered rather than merged

Source independence must be evaluated not only at the level of institutional origin, but also at the level of shared methodological, cultural, and contextual constraints. Two sources may originate from different organizations while sharing the same testing facility, the same publication period, the same readership, and the same dominant discursive context. In such cases, treating convergent findings as fully independent confirmations inflates apparent evidential weight. CEF therefore requires explicit assessment of structural independence — the degree to which sources could plausibly have reached different conclusions given their respective positions and constraints. Where structural dependence is high, convergent findings should be treated as a single evidential signal rather than independent corroboration. This is particularly relevant in domains where publication ecosystems are shaped by shared commercial dependencies, common methodological standards, or dominant institutional narratives. This operational structure enables the evaluation process to be assessed not only in terms of outcomes, but in terms of how context is reconstructed, evidence is differentiated, and conclusions are formed under uncertainty.

F.4.6 Benchmark alignment

Reliable evaluation requires explicit comparison frames. CEF introduces multiple benchmark layers:

- regulatory baseline (minimum compliance),
- historical or class-based baseline,
- contemporary reference standard,
- and performance ceiling.

Failure to distinguish between these benchmark layers can produce systematically misleading judgments: a system may appear deficient against contemporary high-performance technologies while remaining adequate within its historical class or regulatory frame. This prevents misleading conclusions arising from inappropriate or implicit comparisons.

G.4.7 System-state and lifecycle differentiation

CEF explicitly distinguishes between:

- **design capability**, and
- **in-use condition**.

This prevents conflating:

- inherent design limitations,
- and degradation due to usage or maintenance.

H.4.8 NDG-aware evaluation

CEF incorporates NDG by explicitly evaluating whether conclusions are:

- supported by structured evidence, or
- driven by discursive dominance.

When NDG is high:

- narrative claims are explicitly labeled,
- uncertainty is surfaced,
- and conclusions are presented as conditional.

I.4.9 Output schema

A minimal structured output includes:

1. Context summary (C*)
2. Evidence layers (separated by type)
3. Trade-offs and constraints
4. Conditional conclusions
5. Uncertainty and NDG indicators

This schema replaces single authoritative answers with **transparent evaluative reasoning**.

J.4.10 Integration

ACE and CEF are interdependent components:

- ACE defines *what is being evaluated*,
- CEF defines *how it is evaluated*.

NDG connects both by:

- signaling when additional clarification is needed,
- and constraining how conclusions should be formed.

Together, NDG, ACE, and CEF define a coherent interactional–epistemic pipeline for ambiguity-aware technical evaluation.

V. Evaluation Paradigm

A.5.1 Why traditional evaluation is insufficient

Unlike standard evaluation approaches, the proposed framework evaluates not only the correctness of outputs, but the quality of the evaluation process itself. Standard evaluation of large language models (LLMs) has primarily focused on metrics such as accuracy, factual correctness, and benchmark performance.

In such domains, correctness is not an intrinsic property of an answer but is contingent on the underlying problem specification. If the problem itself is underspecified, even a factually correct response may be irrelevant or misleading. Moreover, many technical queries do not admit a single correct answer but instead require conditional judgment based on multiple interacting variables.

As a result, evaluating LLM systems solely on answer correctness overlooks a critical dimension: whether the system has appropriately reconstructed the problem and supported a valid evaluative process.

This work therefore adopts an **interaction-centered evaluation paradigm**, where system performance is assessed in terms of its ability to support structured problem formulation, evidence-based reasoning, and calibrated user judgment.

This aligns with emerging reliance-based evaluation frameworks that measure how users depend on AI systems in decision contexts (Zhang et al., 2025).

B.5.2 Research hypotheses

To operationalize this paradigm, we define a set of testable hypotheses.

H1 — Problem Specification Quality

Systems implementing Adaptive Context Elicitation (ACE) will produce more complete and better-specified representations of the problem compared to direct-answer baselines.

H2 — Decision Quality

Users interacting with NDG–ACE–CEF systems will make higher-quality decisions in technical tasks compared to users receiving standard LLM responses.

H3 — Trust Calibration

Structured outputs produced by the framework will improve trust calibration, reducing both over-reliance and under-reliance, and aligning user confidence more closely with actual system reliability.

H4 — Transparency and Understanding

Users will demonstrate improved understanding of evidence structure, trade-offs, and uncertainty when interacting with Context-Aware Epistemic Filtering (CEF) outputs.

H5 — Cognitive Efficiency (conditional)

Although ACE introduces additional interaction steps, it will reduce downstream cognitive load in complex decision-making tasks by improving problem specification and reducing ambiguity.

C.5.3 Metrics

These metrics are designed not only to evaluate output correctness, but to capture the quality and structure of the evaluation process itself.

We define four evaluation dimensions:

a) 1. Problem Specification Quality

- number of relevant variables identified
- completeness vs expert baseline
- ambiguity reduction score

b) 2. Decision Quality

- agreement with expert judgment
- task outcome quality
- decision error rate

c) 3. Trust Calibration

- calibration gap
- over-reliance index
- under-reliance index

These measures capture whether users trust the system appropriately rather than blindly or insufficiently.

d) 4. Cognitive Load & Usability

- NASA-TLX
- number of interaction steps
- perceived usefulness

e) 5. Transparency

- ability to identify evidence sources
- understanding of trade-offs
- explanation satisfaction

f) 6. Process Integrity Score (PIS)

Measures the extent to which the system follows a structured and complete evaluation process prior to producing a conclusion, including context reconstruction, differentiation of evidence types, and explicit handling of uncertainty.

D.5.4 Baseline conditions

To evaluate the proposed framework, we define three baseline conditions:

Baseline 1 — Direct Answer LLM

Single-turn response with no clarification or structured reasoning.

Baseline 2 — Generic Clarification

Systems that ask follow-up questions, but without a structured policy such as ACE.

Baseline 3 — Retrieval-Augmented Generation (RAG)

Systems that incorporate external information sources but do not apply structured epistemic filtering (no CEF).

Experimental Condition — NDG + ACE + CEF

Systems implementing the full framework:

- structured clarification,
- context reconstruction,
- and multi-layer evidence evaluation.

E.5.5 Experimental design

We propose a **within-subjects experimental design**, where participants complete tasks under multiple system conditions.

Key features:

- same participants across conditions
- counterbalanced order
- tasks with real ambiguity

a) Task types

- product evaluation
- diagnostics
- risk assessment

b) Data collection

- interaction logs
- decisions
- confidence
- questionnaires

F.5.6 Key shift

The core shift is:

✕ “Is the answer correct?”

✓ “Did the system help construct a correct evaluation?”

This shift reflects the reality that, in ambiguous technical domains, the quality of reasoning and interaction is as important as the correctness of any single output.

VI.6. Discussion

A.6.1 Reinterpreting LLM failure modes in technical domains

The framework proposed in this paper suggests a shift in how LLM limitations are understood in technical contexts. While much of the literature focuses on hallucination and factual inaccuracy, our analysis indicates that a more pervasive failure mode is **premature evaluative closure under ambiguity**.

Beyond training data effects, LLM behavior is also shaped by optimization toward coherent and complete responses under partial context. This can lead to early closure: the formation of a plausible and internally consistent answer before all available evidence has been fully differentiated and evaluated. As a result, less salient but epistemically stronger signals may be underweighted relative to more dominant or readily integrable narratives. This pattern resembles well-documented human cognitive heuristics, not because the model “thinks like a human,” but because both systems operate under constraints that favor coherence, compression, and timely decision-making.

LLMs do not simply inherit human biases; they reproduce a structurally similar pattern of early coherence formation under uncertainty, which can lead to premature closure and underweighting of critical evidence.

From a broader perspective, the observed behavior can be interpreted as a form of structural convergence between LLM systems and human reasoning under uncertainty. This similarity should not be understood as evidence that models "think like humans," but rather as a consequence of shared constraints: operating under incomplete context, heterogeneous information, and pressure to produce a coherent output within limited time. Under such conditions, both human reasoning and model generation tend toward early formation of plausible interpretations, even when the available evidence has not been fully differentiated. In LLM systems, this manifests as premature evaluative closure, whereby discursively dominant or easily integrable narratives may be preferred over less salient but epistemically stronger signals. This perspective complements the concept of Narrative Dominance Gap (NDG) by linking it not only to properties of the data, but also to more general constraints in the process of evaluation under uncertainty. In this sense, NDG can be understood not only as a property of information environments, but as an emergent effect of how both discourse and reasoning processes interact under conditions of uncertainty.

In many cases, the model does not generate false information. Instead, it:

- selects an implicit interpretation of an underspecified query,
- collapses heterogeneous evidence into a single narrative,
- and presents the result with unwarranted coherence.

This behavior is consistent with the concept of **Narrative Dominance Gap (NDG)**. The model reflects patterns of discourse rather than explicitly structured evidence. As a result, outputs may appear plausible and informative while remaining insufficiently grounded for decision-making.

From this perspective, improving LLM reliability is not only a matter of better knowledge or retrieval, but of **restructuring the interaction and reasoning process**.

B.6.2 Clarification as a core function, not an auxiliary feature

A central implication of this work is that clarification should be treated as a **core functional requirement** rather than an optional conversational feature.

In current systems, clarification is often:

- minimal,
- generic,
- or absent.

However, in technical evaluation tasks, the absence of clarification leads directly to mis-specified problem representations. By introducing **Adaptive Context Elicitation (ACE)**, the framework positions clarification as a structured mechanism for reconstructing decision-relevant context.

This has important implications for system design:

- interaction becomes part of reasoning,
- users become participants in problem specification,
- and system outputs become conditional on explicitly surfaced assumptions.

C.6.3 From information retrieval to epistemic structuring

Another implication concerns the role of retrieval and information access. While retrieval-augmented systems improve access to external knowledge, they do not by themselves ensure reliable evaluation.

The key challenge is not only retrieving information, but **structuring it epistemically**.

Through **Context-Aware Epistemic Filtering (CEF)**, the framework emphasizes:

- differentiation between types of evidence,
- alignment with case-specific context,
- and explicit representation of disagreement and uncertainty.

This shifts the focus from "more information" to **better-structured information**.

D.6.4 NDG as a unifying concept

The introduction of NDG provides a unifying perspective across several known issues:

- hallucination → extreme NDG (no evidential support)
- overconfidence → unacknowledged NDG
- bias toward common narratives → discursive dominance
- miscalibration → mismatch between discourse and evidence

NDG therefore functions as:

- a diagnostic tool,
- a design constraint,
- and a bridge between epistemic and interactional perspectives.

E.6.5 Implications for human–AI collaboration

The framework contributes to human–AI collaboration in three key ways:

1. **Improved problem formulation**
   Users are supported in articulating relevant variables.
2. **Improved decision support**
   Outputs reflect structured evaluation rather than simplified narratives.
3. **Improved trust calibration**
   Users are exposed to assumptions, evidence layers, and uncertainty.

This aligns with broader goals in human-centred AI, where system value is measured not only by correctness, but by how effectively it supports human reasoning.

From a broader systems perspective, the role of LLMs can be understood as a key component in shaping what has been described as "architectural capital" in human–computational systems (Rusev, 2026). Within this framework, system performance is not determined solely by the capabilities of individual components, but by the way interaction structures organize the flow of information, attention, and decision-making. When embedded in structured interactional frameworks such as NDG–ACE–CEF, LLMs function not merely as information-processing modules, but as configurational elements that shape the integration of human and machine capabilities and enable their augmentation through partial compensation of their respective limitations. In this sense, improving reliability is not only a matter of enhancing the model itself, but of optimizing the architecture of interaction through which the evaluative process is constructed.

VII. Contributions (Explicit Summary)

This paper makes the following contributions:

A.7.1. Conceptual Contribution

It introduces **Narrative Dominance Gap (NDG)** as a theoretical construct for analyzing divergence between discursive prevalence and evidential support in LLM-based evaluation.

B.7.2. Framework Contribution

It proposes a unified **NDG–ACE–CEF framework** that integrates:

- ambiguity detection,
- context reconstruction,
- and structured evidence evaluation.

C.7.3. Methodological Contribution

It defines **case evaluation** as an alternative paradigm to answer generation and provides a structured interactional–epistemic model for implementing it.

D.7.4. Evaluation Contribution

It outlines an **interaction-centered evaluation paradigm**, including hypotheses, metrics, and experimental design focused on:

- problem specification quality,
- decision quality,
- trust calibration,
- and transparency.

E.7.5. Design Contribution

It provides actionable design principles for LLM systems in technical domains:

- treat queries as incomplete by default,
- structure clarification,
- differentiate evidence explicitly,
- expose uncertainty and disagreement.

F.7.6 From Answer Generation to Interactional Reasoning

This paper argued that improving LLM performance in technical domains requires a shift from answer generation to case evaluation under uncertainty. By introducing Narrative Dominance Gap (NDG), Adaptive Context Elicitation (ACE), and Context-Aware Epistemic Filtering (CEF), it proposed a unified framework for ambiguity resolution, context reconstruction, and structured evidence evaluation.

The central claim is that reliable AI systems in technical domains must not only generate answers, but also reconstruct the problem, differentiate evidence, and support calibrated human judgment. Beyond these specific mechanisms, the broader implication is that limitations of LLM-based evaluation arise not only from gaps in knowledge, but from structural pressures toward early coherence under uncertainty. In this respect, model behavior may converge with human reasoning—not because of shared cognition, but because both operate under constraints of incomplete, heterogeneous information and pressure for coherent output.

The proposed NDG–ACE–CEF framework addresses this challenge by restructuring interaction so that context reconstruction and evidence differentiation precede evaluative closure. Ultimately, the next generation of human–AI systems will be defined not by fluency alone, but by epistemic discipline and interactional intelligence: systems that ask before answering and structure before concluding, thereby supporting more reliable judgment under real-world uncertainty.

VIII. Limitations

This work has several limitations.

First, the framework is conceptual and not yet implemented as a full system. Its effectiveness depends on how ACE and CEF are instantiated in practice.

Second, the approach assumes that users are willing and able to provide additional context. In real-world settings, user input may be incomplete or inaccurate.

Third, while CEF improves evidence structuring, it does not eliminate biases in underlying data sources.

Fourth, the framework introduces additional interaction steps, which may increase perceived friction if not carefully managed.

Finally, the framework is developed primarily for technical domains with measurable parameters and may require adaptation for more subjective domains.

Conceptual scope of NDG.

NDG is introduced as a diagnostic and interpretive construct, not as a fully validated causal mechanism. Its role is to identify conditions under which discursive prominence may become misaligned with evidential support in case-based evaluation. Future work should therefore focus on operational criteria, annotation procedures, and empirical benchmarks capable of distinguishing NDG from related phenomena such as generic prior dependence, overconfidence, or source bias.

The framework is developed with particular attention to domains characterized by measurable parameters, established normative standards, heterogeneous source reliability, and verifiable empirical benchmarks. These structural properties — present in engineering evaluation, technical diagnostics, regulatory compliance assessment, and related fields — are the conditions under which NDG, ACE, and CEF are most directly applicable. The framework's applicability to less structured domains, such as aesthetic, legal, ethical, or purely subjective evaluation, is not excluded in principle, but requires domain-specific adaptation that lies beyond the scope of the present work. This bounded scope is not a limitation of the framework's ambition, but a precondition for its empirical testability and practical implementability.

IX. Conclusion

This paper argued that improving LLM performance in technical domains requires a shift from answer generation to **case evaluation under uncertainty**.

By introducing:

- **Narrative Dominance Gap (NDG)**,
- **Adaptive Context Elicitation (ACE)**,
- and **Context-Aware Epistemic Filtering (CEF)**,

we proposed a unified framework that integrates:

- ambiguity resolution,
- context reconstruction,
- and structured evidence evaluation.

The central claim is that reliable AI systems in technical domains must not only generate answers, but also:

- reconstruct the problem,
- differentiate evidence,
- and support calibrated human judgment.

Beyond specific mechanisms, this work advances a broader interpretation: limitations of LLM-based evaluation in technical domains arise not only from gaps in knowledge, but from structural pressures toward early coherence under uncertainty. In this respect, model behavior may exhibit forms of convergence with human reasoning—not due to shared cognition, but due to shared constraints in processing incomplete, heterogeneous information under time and coherence pressures.

The proposed NDG–ACE–CEF framework addresses this challenge not by eliminating these pressures, but by restructuring interaction so that context reconstruction and evidence differentiation precede evaluative closure.

More broadly, improving LLM reliability requires not only better access to information, but stronger control over how evaluative reasoning is structured, sequenced, and made explicit in interaction.

Ultimately, the next generation of human–AI systems will be defined not by increased fluency, but by epistemic discipline and interactional intelligence—systems that ask before answering and structure before concluding, thereby supporting more reliable decision-making under real-world uncertainty. In this sense, the primary limitation of current LLM systems is not that they lack knowledge, but that they tend to draw conclusions too early under conditions of uncertainty — a pattern that, to some extent, is also characteristic of human reasoning. Addressing this limitation does not necessarily require more data, greater computational power, additional training programs for users, or more complex models, but rather a reorganization of how existing capabilities are structured and applied within interaction. The NDG–ACE–CEF framework demonstrates how targeted adjustments in the structure of interaction and reasoning can improve system reliability without fundamental changes to the underlying model. This perspective aligns with the broader view that the effectiveness of technological systems depends not only on the properties of individual components, but on how they are integrated. Future research should therefore extend beyond the immediate scope of this work and include interdisciplinary efforts — spanning engineering, cognitive, and behavioral domains — aimed at improving patterns of reasoning as a whole, both in human–AI interaction and in the way models themselves structure the reasoning process, as well as the overall architecture of human–AI systems.

Appendix: **A Documented NDG Instance — From Narrative to Evidence-Based Evaluation**

The appendix provides a fully documented real-world case demonstrating the NDG–ACE–CEF framework in action and is included to support the methodological contribution.

This appendix is included as an illustrative and methodological component, not as a standalone empirical validation.

A.1. Initial Query

The following case begins with an intentionally underspecified evaluative query:

“*What do you think about the effectiveness of the braking system of the Soviet car Lada 1500 (VAZ-2103)?”*

The query does not specify:

- temporal frame (historical vs present-day usage),
- operational context (single-stop vs repeated braking),
- system state (original vs maintained/modified),
- comparison benchmark.

As such, it represents an incomplete problem specification rather than a well-defined evaluation task.

A.2. Baseline LLM Outputs (NDG Manifestation)

Initial responses generated by large language models contained claims such as:

- “tendency to overheat (fade)”
- “weak braking performance”
- “poor component quality”
- “unsuitable for modern driving conditions”

These claims exhibit:

- high discursive prevalence (common associations with older vehicles and drum brakes),
- lack of case-specific empirical support,
- presentation as definitive conclusions rather than conditional hypotheses.

This constitutes a clear instance of **Narrative Dominance Gap (NDG)**, where discursively dominant claims are produced without sufficient evidential grounding.

A.3. ACE in Action — Context Reconstruction

Context was progressively reconstructed through targeted input (Adaptive Context Elicitation), including:

1. Historical road tests (Motor, Autocar, 1976)
2. Engineering analysis aligned with UN Regulation No. 13-H
3. Clarification of realistic present-day usage conditions:
   - modern tires (post-1980 / contemporary compounds),
   - replacement of asbestos brake pads,
   - use of modern brake fluid (e.g., DOT-4),
   - realistic system condition (maintained, not factory-new).

This yields a structured context representation:

C* = { system: VAZ-2103 braking system (disc front / drum rear),
temporal frame: historical design vs present-day usage,
operational context: single-stop and repeated braking,
comparison frame: regulatory baseline and contemporary vehicles,
system state: maintained with modern consumables }

A.4. CEF — Evidence Differentiation

Evidence is structured into distinct epistemic layers:

Evidence Type	Content	Reliability
Empirical measurements	0.4 g @ 20 lb; 1.0 g @ 50 lb; ≤38 m @ 80 km/h; ≈60 m @ 100 km/h	High
Thermal behavior	Fade test (10 stops from 70 mph; moderate pedal increase, no collapse)	High
Normative criteria	Compliance with UN Regulation No. 13-H	High
Engineering analysis	Improved performance with higher tire friction (μ increase)	High
Experiential reports	Subjective descriptions of pedal feel (“non-progressive”, “over-servoed”)	Medium
Diffusion signals	“Soviet cars have weak brakes”	Low

A.5. NDG Identification

The claim:

“The braking system is prone to fade”

is identified as a **high-NDG statement**, because:

- discursive prevalence: high (general knowledge about drum brakes),

- case-specific evidence: contradictory (fade test passed successfully),
- evidential grounding: absent in the initial response.

Similarly, claims regarding “weak braking performance” and “poor component quality” were not supported by case-specific empirical evidence.

A.5.1. NDG Example: Pedal Feel

The evaluation of pedal feel provides a distinct example of Narrative Dominance Gap (NDG) arising not from incorrect data, but from the misclassification and over-weighting of subjective evidence.

Historical road tests (Motor, Autocar, 1976) describe the braking behavior as “non-progressive” and “over-servoed,” indicating difficulty in smooth modulation. These observations constitute **experiential reports**, reflecting driver perception under specific testing conditions.

However, such statements:

- are inherently **context-dependent** (driver expectations, comparison vehicles, test conditions),
- are not supported by **quantitative metrics**,
- may be influenced by **shared testing environments and editorial framing.**

Despite this, they are often implicitly treated as **system-level deficiencies**, rather than as **subjective observations of limited generalizability**.

This creates an NDG condition:

Dimension	Pedal Feel Case
Discursive prevalence	Moderate to high (“non-progressive braking”)
Evidence type	Experiential (subjective)
Case-specific empirical support	Absent
Interpretation risk	Over-generalization into system deficiency

Under Context-Aware Epistemic Filtering (CEF), such evidence should be:

- explicitly classified as **subjective**,
- assigned **moderate evidential weight**,
- interpreted as **context-bound**, not universal system behavior.

A.6. Structural Dependence of Sources

Although two independent publications (Motor and Autocar, 1976) report similar subjective impressions (e.g., pedal feel), their independence is limited by:

- shared testing environment (MIRA),
- identical temporal context,
- overlapping readership and editorial culture,
- exposure to similar discourse conditions.

Therefore, their convergence does not constitute fully independent validation, but rather structurally dependent agreement.

A.7. Revised Evaluation under CEF

After context reconstruction and evidence differentiation:

- **Maximum braking capability**: objectively sufficient (≈6.5 m/s^2), exceeding regulatory minimums.
- **Thermal performance**: stable under repeated braking; no critical fade observed.
- **System architecture**: appropriate for the era and comparable to contemporaries.
- **Pedal feel**: subjective reports of non-linear or over-assisted behavior (medium reliability).
- **Modern usage scenario**: performance is significantly improved with contemporary tires and consumables.

Notably, subjective reports of pedal feel should be interpreted with caution. While historical test sources describe non-progressive behavior, such assessments are inherently context-dependent and may reflect specific testing conditions, driver expectations, or comparative frames rather than universally reproducible system characteristics.

A.8. Key Observation

The primary failure mode observed in initial responses is not factual fabrication, but:

evaluative closure based on category-level priors, arising under conditions of incomplete context and pressure for coherent output, structurally analogous to well-documented patterns in human reasoning under uncertainty

The NDG–ACE–CEF framework transforms:

- a narrative-driven evaluation
  into
- a structured, context-conditioned assessment.

The primary failure mode observed in initial responses is not factual fabrication, but evaluative closure based on category-level priors, arising under conditions of incomplete context and pressure for coherent output, structurally analogous to well-documented patterns in human reasoning under uncertainty. This contrast illustrates that the primary change is not the acquisition of new knowledge, but the restructuring of evaluation through context reconstruction and evidence differentiation. In this sense, the case does not demonstrate improved factual recall, but improved epistemic organization.

This demonstrates that the improvement arises from restructuring the evaluation process, not from introducing new information.

A.9. Implication for LLM Evaluation

This case demonstrates that:

- access to information alone is insufficient,
- reliable evaluation requires:
  - context reconstruction,
  - explicit differentiation of evidence,
  - controlled handling of discursive priors.

Therefore, system performance should be evaluated not only in terms of answer correctness, but in terms of:

its ability to construct a valid evaluation process under uncertainty.

***

ACKNOWLEDGMENT

This study is part of an independent research program: *Ideas for the Third Millennium*, developing an architectural theory of complex systems.

The program centers on the concept of Architectural Capital and extends it across engineering, economic, and cognitive domains, including product design, system evaluation, and human–machine interaction.

X. References

[1] Abercrombie, Gavin, Amanda Cercas Curry, Tanvi Dinkar, Verena Rieser, and Zeerak Talat. (2023). Mirages: On anthropomorphism in dialogue systems. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 4776–4790.
[2] Autocar. (1976, August 14). *Lada 1500 road test*. IPC Business Press Ltd.
[3] Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S. T., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3290605.3300233
[4] Aliannejadi, M., Kiseleva, J., Chuklin, A., Dalton, J., & Moffat, A. (2021). Conversation-based information retrieval: An overview of recent developments and challenges. *Information Retrieval Journal*, 24, 1–39.
[5] Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., & Weld, D. S. (2021). Does the whole exceed its parts? The effect of AI explanations on complementary team performance. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3411764.3445717
[6] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT)*, 610–623.
[7] Buçinca, Zana, Maja Barbara Malaya, and Krzysztof Z. Gajos. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1).
[8] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.
[9] Geng, S., Liu, H., Zhang, Z., & Liu, Y. (2023). Are language models calibrated? *Advances in Neural Information Processing Systems (NeurIPS)*.
[10] Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing trust in artificial intelligence: Prerequisites, causes, and goals of human trust in AI. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT)*.
[11] Kim, Sunnie S. Y., Q. Vera Liao, Mihaela Vorvoreanu, Stephanie Ballard, and Jennifer Wortman Vaughan. (2024). “I’m not sure, but...”: Examining the impact of large language models ’uncertainty expression on user reliance and trust. *In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 822–835.
[12] Mielke, Sabrina J., Arthur Szlam, Emily Dinan, and Y-Lan Boureau. (2022). Reducing conversational agents ’ overconfidence through linguistic calibration. *Transactions of the Association for Computational Linguistics,* 10:857–872.
[13] Motor. (1976, № 32). *Road test: Lada 1500*. IPC Business Press Ltd.
[14] Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80.
[15] Lin, S., Hilton, J., & Evans, O. (2022). TruthfulQA: Measuring how models mimic human falsehoods. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*.
[16] Liu, A., et al. (2024). CLAMBER: A benchmark for ambiguity resolution in language models. *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*.
[17] Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253.
[18] Rajpurkar, P., Min, S., Lyu, X., & Liang, P. (2020). AmbigQA: Answering ambiguous open-domain questions.

Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP).
[19] Rusev, M. R. (2026). Architectural Capital: A Modern Framework for Addressing Consumer Dissatisfaction and Utility Stagnation. *Economic Thought Journal*, 71 (1), 94–125. https://doi.org/10.56497/etj2671105
[20] Rusev, M. (2026). Analysis of the compliance between historical braking performance parameters and contemporary UN R13-H regulatory requirements. Engineering Sciences, № (1). https://doi.org/10.7546/EngSci.LXIII.26.01.06
[21] Vats, V., Nizam, M. B., Liu, M., Wang, Z., Ho, R., Prasad, M. S., Titterton, V., Malreddy, S. V., Aggarwal, R., Xu, Y., Ding, L., Mehta, J., Grinnell, N., Liu, L., Zhong, S., Gandamani, D. N., Tang, X., Ghosalkar, R., Shen, C., Shen, R., Hussain, N., Ravichandran, K., & Davis, J. (2025). A survey on human-AI collaboration with large foundation models. ACM. https://arxiv.org/pdf/2403.04931v3
[22] Wang, X., et al. (2024). Human-AI collaboration: A survey of design, evaluation, and open challenges. *ACM Computing Surveys*.
[23] Zamani, H., & Craswell, N. (2020). Generating clarifying questions for information retrieval. *Proceedings of the Web Conference (WWW 2020)*.
[24] Zhang, Y., et al. (2025). REL-AI: A framework for reliance-based evaluation of AI systems. *Proceedings of NAACL 2025*.
[25] Zhou, K., Hwang, J. D., Ren, X., Dziri, N., Jurafsky, D., & Sap, M. (2025). REL-A.I.: An interaction-centered approach to measuring human-LM reliance. In Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers) (pp. 11148–11167). Association for Computational Linguistics.
[26] Zhu, W., Liu, H., Dong, Y., Yu, J., Cheng, J., & Li, J. (2025). Rethinking hallucination in large vision-language models: A task-specific perspective. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)* (pp. 9811–9832). Association for Computational Linguistics.

Cognitive, visual, and neuropsychological challenges of prolonged exposure to electronic screen devices: risks and protective mechanisms

Miroslav Dechev[1], Maria-Nikol Tsaneva[2]

[1]Institute of Robotics - Bulgarian Academy of Sciences, Sofia, Bulgaria
[2] University of Economics - Varna, Faculty " Informatics", Varna, Bulgaria

***Abstract.* Electronic screen devices, such as smartphones, tablets, computers, televisions, interactive displays and e-readers, are part of everyday life and are used intensively by all age groups. This widespread and prolonged exposure raises important scientific questions regarding their impact on the cognitive functions of the human brain, visual comfort and general neuropsychological health. The present study aims to analyze data related to the effects of prolonged use of electronic screen devices on brain activity, with a special focus on attention, memory, mental state and sleep, as well as on possible consequences for the central and cardiovascular nervous systems. The study is based on an interdisciplinary approach, including a questionnaire survey among 25 individuals aged over 18 years, as well as an analysis of experimental studies in the fields of neuroscience, ophthalmology, psychology and information technology. The main research question is whether prolonged use of electronic screen devices leads to functional changes in cognitive processes and visual comfort, without causing direct structural damage to the brain or visual system. The results obtained have scientific and applied value for assessing the risks of prolonged screen exposure and can help develop recommendations for healthy use of digital technologies. The results obtained are applicable in educational practice, healthcare and the development of protective technological solutions.**

***Keywords*—screen devices, visual system, information technology, risks, neuroscience.**

I. Introduction

Electronic screen devices (smartphones, tablets, computers, televisions, video walls, interactive displays, e-book readers, touchscreen monitors, etc.) have been an integral part of our daily lives for at least the last two decades. Their use is widespread in all age groups. This raises research questions related to their impact on the cognitive functions of the human brain and our visual abilities. The present study aims to analyze modern scientific data on the effects that electronic screen devices have on the human brain and the challenges that people face during their longer-term use, as well as modern technological protections implemented in modern electronic devices in order to protect and reduce risks to vision. The research tasks that this scientific article should solve are several: analysis of the impact of screens on attention, memory, psyche and sleep, even the functions of the cardiovascular and central nervous systems; to examine the main visual problems associated with screen exposure; assessment of the risk and impact at different ages, as a result of the time and content of screen use. The development is based on a survey conducted among 25 individuals aged 18 and over on the topic: “Survey on the impact of electronic screen devices on the human brain”, a review and comparative analysis of the medical scientific literature and studies by specialists on the subject; analysis of experimental studies and an interdisciplinary approach (neuroscience, ophthalmology, psychology, information technology). The main question to be answered is whether the prolonged and uncontrolled use of screen electronic devices leads to functional changes in cognitive processes and visual comfort, without causing direct structural damage to the brain or vision.

II. Overview and comparative analysis of studies related to the impact of screens on cognitive, visual and neuropsychological functions

Today, younger generations, who are also more active online, such as Gen Z and their successors – Alpha, are the young people who have suffered from the so-called “brain rot”[1], mainly due to the long time spent in front of screen devices with their permanent digital life.

Screen media activity [1] may have an impact on neurodevelopment in young people. Moreover, the applied results of the researchers indicate that the covariation (changing behavior) of the structural patterns of the brain is associated with Screen media activity, as well as with the psychopathological state, cognition and sleep in children, for example. On the other hand, the first signals of this condition are observed in a feeling of fatigue, lethargy, reduced concentration, cognitive decline. Characteristic of brain rot is the so-called doomscrolling (“doom” – in translation “doom”, and “scrolling” – scrolling, scrolling) [1]. According to a scientific study by the Newport Institute from 2024, “Brain rot” is caused by excessive use of technology.

This can mean watching videos on YouTube, scrolling on social media or switching between different browser tabs. In addition, you can simultaneously surf the Internet, write messages and check your email. The end result: You’re overstimulating your brain. And when you’re digitally inundated with too much information, you’re at risk of brain rot. Scrolling through social media platforms increases levels of the neurochemical dopamine, which produces feelings of satisfaction and pleasure. The more you do it, the more you want to do it. Your brain associates scrolling with feelings of satisfaction, even when you’re aware of its negative consequences. In this way, scrolling can become a behavioral addiction [2].

Neuroscientist Susan Greenfield, a senior research fellow at Oxford and widely regarded as the most popular and successful female scientist in the UK, claims that internet and computer game use can damage the brain.

Screen time has been blamed for increasing depression in young people, behavioral problems and insomnia [3].

The renowned neuroscientist says that internet and computer game use can damage the brain of adolescents. In her latest book, Mind Change, she examines the various ways in which digital technologies can affect not only our thinking patterns and other cognitive skills, but also our lifestyles, culture and personal perceptions.

She compares changes in consciousness to changes in climate, signals that were considered crazy decades ago but are now generally accepted scientific fact – both are “global, contradictory, unprecedented and multifaceted” [3].

Another group of British scientists claims that there is a lack of concrete scientific evidence for the negative effects of screens on cognitive functions.

Among our main tasks in this report are: to analyze the impact of screens on attention, memory, psyche and sleep, even the functions of the cardiovascular and central nervous systems; to examine the main visual problems associated with screen exposure; to assess the risk and impact at different ages, as a result of the time and content of screen use. In today’s world, saturated with screen devices, we are all constantly exposed to artificial light.

Among all the types light, most often we talk about blue light and the connection between it and hormones, especially those of sleep [4].

With very early use of screens, the child also slows down his language development. He speaks later, his intellect develops later, which shows that screen activity competes with the simulation abilities of the brain to develop language, consciousness, alertness, intellect, and ego. The functioning of the child's brain is compromised by screen activity in all sorts of ways, but morphological lesions are no exception. Scientific evidence of changes in the white and gray matter of the brain is constantly emerging. There is a persistent falling asleep and the inability of the brain to be awake, with which it can fully participate in life. On the contrary, it becomes completely dependent on digital heroin and has no will to go out and cope on its own [4].

A. Digital fatigue - a mental health challenge from excessive exposure to electronic screen devices

With every swipe of a finger on a smartphone and every keystroke on a laptop, we accumulate something invisible – digital fatigue. This often underestimated condition has an increasingly serious impact on mental health, concentration, and even sleep quality [5].

The term “digital fatigue” encompasses the physiological and psychological effects of excessive use of screens – smartphones, laptops, tablets, televisions, and other devices. Symptoms include fatigue, irritability, blurred vision, headaches, lack of motivation, and decreased productivity. It can also manifest as difficulty concentrating, anxiety, or even social isolation [5].

A number of studies suggest that the blue light emitted by electronic screen devices contributes to difficulty falling asleep by suppressing the hormone melatonin. This has generated scientific and public interest in their potential impact on the visual system. Does prolonged and uncontrolled use of electronic screen devices lead to functional changes in visual comfort, without causing direct structural damage to the brain or retina. To what extent do luminous displays affect vision, and could cause computer vision syndrome, myopia or visual discomfort. These are research searches by dozens of modern scientists and specialists in the fields of neuroscience, ophthalmology, psychology.

The blue light emitted by screens suppresses the production of melatonin - the sleep hormone. Using devices before bedtime leads to difficulty falling asleep, lower sleep quality and chronic fatigue [5].

B. Protective tools and mechanisms in modern electronic display devices

In today’s digital world with a wide range of electronic devices, from phones and tablets to laptops and monitors, people spend a huge amount of time looking at screens.

Among them, organic light-emitting diode (OLED) screens occupy an important place in the field of electronic displays, due to their remarkable advantages such as high contrast ratio, wide viewing angle, thinness, lightness, and bendability, which are due to their self-emitting characteristics. They are widely applied in various products, including smartphones and televisions [6].

However, with the increasing prevalence of OLED screens, the issue of eye protection has gradually attracted attention. Some OLED screens exhibit screen flicker at low brightness levels, which can lead to uncomfortable symptoms such as visual fatigue, dry eyes, and headaches. This discomfort is especially pronounced when used at night or in low-light environments. Therefore, improving the eye protection of OLED screens is of great practical importance [6].

Staring at electronic display devices leads to problems with visual systems. Therefore, eye protection is of great importance for new technologies related to electronic display devices.

Based on the advantages of infrared light, which can resonate with water molecules, activate mitochondria, stimulate blood circulation, and relieve visual fatigue, BOE has developed a screen that can emit infrared light to achieve eye protection effects by adding an infrared light emitting device to the screen [7].

C. “Eye Comfort Shield” - an option to filter blue light

“Eye Comfort Shield” is available in mid-range monitors, laptops, TVs, smartphones and tablets. It is a blue light filter function. It is filtered to minimize the risk of eye irritation and not to interfere with sleep. In general, “Eye Comfort Shield” is a complex of optical electronic and software technologies designed to reduce visual strain (asthenopia) by limiting the harmful photobiological effects of light,

stabilizing visual adaptation and reducing neurosensory stress during prolonged use of visual displays.
“Eye Comfort Shield” is a set of preventive technologies that, by controlling the light spectrum, brightness and flicker, reduce photobiological, neurosensory and muscle stress on the visual system. It is not a treatment, but a means of prevention and improvement of visual ergonomics [8,9,10,11,12]. Blue light is emitted by display devices – such as televisions, telephones, lighting fixtures and can irritate the eyes. Blue light also blocks the production of the hormone melatonin, which prevents sleep [13].

Prolonged work with digital displays is associated with an increased incidence of visual fatigue, summarized in the clinical term Digital Eye Strain (DES). Symptoms include asthenopia, dry eyes, headaches and impaired accommodation. In this context, the concept of “eye comfort shield” is introduced as a preventive technological approach to reduce visual stress.

1) Photobiological mechanisms
Short-wavelength visible light (400–500 nm), known as blue light, has higher energy and causes increased scattering in the optical media of the eye. This increases chromatic aberration and strains the accommodative apparatus [8,9,10,11,12]. Prolonged exposure leads to accelerated visual fatigue and subjective discomfort.
2) Neuroendocrine effects
Blue light activates melanopsin-containing retinal ganglion cells, which are involved in the regulation of circadian rhythms. Increased stimulation suppresses melatonin secretion, which can lead to sleep disturbances and secondary increases in cognitive and visual fatigue[8,9,10,11,12].
3) Role of Flicker
Many displays use pulse-width modulation (PWM) to regulate brightness. Although flicker is often below the threshold of conscious perception, it leads to increased load on the visual cortex and can cause headaches and fatigue[8,9,10,11,12]. Flicker-free technologies significantly reduce this effect.
4) Physiology of accommodation
Prolonged fixation on close objects leads to prolonged contraction of the ciliary muscle. Optimizing contrast, brightness, and color temperature through eye comfort technologies reduces the need for constant accommodative correction [8,9,10,11,12].

From a physiological point of view, Eye Comfort Shield affects:

- the spectral composition of light;
- neuroendocrine regulation;
- electrophysiological comfort;
- muscle activation;
- preventive intervention against DES (Digital Eye Strain).

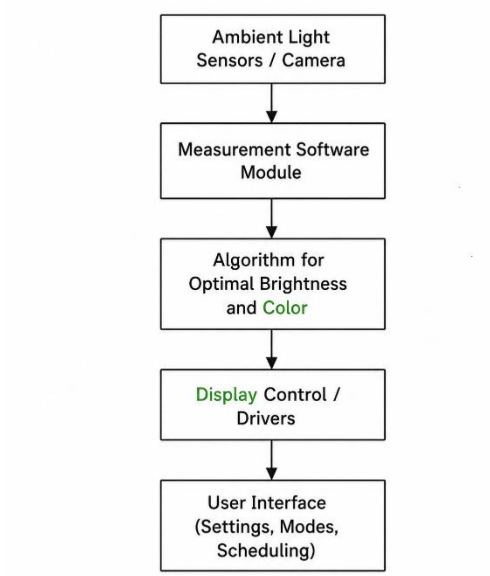


Fig.1. Screenshot of Eye Comfort Shield Architecture Diagram

III. Studying the impact of electronic screen devices on the human brain and vision

The use of electronic screen devices (smartphones, tablets, computers, light displays, smartphones, e-book readers, etc.) has grown exponentially, spreading literally across all age groups. This has generated scientific and public interest in whether and how screens affect the human brain and our vision. In this “Survey” type study, we are looking for answers to several questions related to how people perceive their work with light-emitting machines, whether they affect them after use, how they feel after prolonged standing in front of screen electronic devices. Do they use protective mechanisms and tools while looking at screens, does their behavior change after excessive use. The results will provide a clear answer to how much time people spend in front of electronic screen devices, why they use them, and whether they protect their eyes from the glare mentioned above in the report. The current scientific study aims to analyze current data on the effects of screen use on humans and whether they are dangerous to health or simply burden the visual system and psyche of a person. Scientists are still not certain that screens change the brain in a permanent or lasting way for the worse, rather they define the time spent in front of the screen as a key moment for cognitive changes, without disrupting the structure of vision or the human brain.

A: Description of the study conducted

The survey was conducted online using the JotForm platform, an online form creation and workflow automation tool powered by artificial intelligence (https://eu.jotform.com/workspace/) and is titled: “Survey on the impact of electronic screen devices on the human brain", including 14 questions:

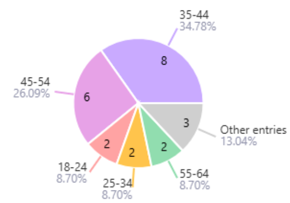
1. What is your gender?
2. What is your age range?
3. How much time do you spend in front of electronic screen devices per day?
4. What do you most often use electronic screen devices for /smart phones and TVs, watches, video walls, tablets, e-books, laptops, desktop computers with monitors?
5. What is the highest level of education you have completed?
6. How often do you exercise/relax/rest during the week?
7. How often do you use medications, vitamins, nutritional supplements related to improving your physical and mental condition?
8. How much time do you need for rest/sleep per day?
9. How do you feel after using electronic screen devices?
10. Have you ever had mental discomfort or physical pain after using electronic screen devices?
11. Do electronic screen devices affect your vision in your opinion?
12. Have you ever used protective glasses for working with a screen device or eye drops to relieve dry eyes, or used a shield for eye comfort from settings on your smart devices?
13. Please rate your work with electronic screen devices.
14. What are your interests?

B: Analysis of the data obtained from the study and results

The largest share of 34.78% in the study covers people aged 35 to 44, which clearly shows that active working people are most often in front of screen electronic devices. Next in line are those aged 45 to 54, which again affects the actively employed population. The participants in this age range are 26.09%.

It seems that younger people, contrary to the expected and public perception that they are constantly online, are not confirmed by this study. The participants aged 18 to 25 and 25 to 34 are only slightly over 8%, as are actually those of pre-retirement age up to 64. This suggests that those who probably use electronic screen devices are mostly active working and able-bodied people who have to use new technologies in a work environment.

Your age

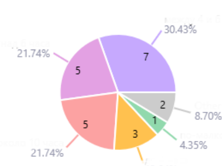


Data	Response	%
35-44	8	34.78%
45-54	6	26.09%
18-...	2	8.70%
25-34	2	8.70%
55-64	2	8.70%
Other entries	3	13.04%

Fig. 2. Age range

The answers to Question number 3 in the Questionnaire are of interest: "How much time do you spend per day in front of screen electronic devices?" Almost 22% of respondents spend about 10 hours per day in front of screen electronic devices. At most, 30.43% spend between 4 and 6 hours, and nearly 22% spend more than 6 hours. Only 4 percent of those surveyed spend less than an hour in front of the screen. As a conclusion, it can be concluded that electronic screen devices, such as smartphones, tablets, computers, televisions, interactive displays and e-readers, are part of our daily lives and are used intensively by all age groups. It is this broad and prolonged exposure that raises significant scientific questions about their impact on the cognitive functions of the human brain, visual comfort and general neuropsychological health.

How much time a day do you spend in front of electronic screen devices?

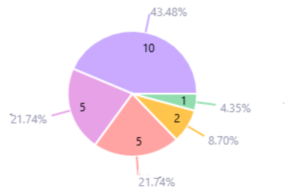


Data	Response	%
between 4 and 6 hours	7	30.43%
over 6 hours	5	21.74%
about 10 hours	5	21.74%
4 hours	3	13.04%
less than an hour	1	4.35%
Other entries	2	8.70%

Fig. 3. Time spent in front of electronic screen devices

Just over 43% of respondents say that the screen time they spend is for the purpose of searching for information, 22% use the screen for educational purposes and communication. And only 3 of those surveyed use the glowing screens for social networking and entertainment. From the analysis, we can summarize that electronic screen devices are almost the main source of information in the digital 21st century.

What do you most often use devices with electronic screens for?

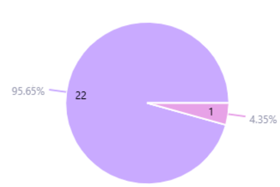


Data	Response	%
I am looking for information	10	43.48%
education/training	5	21.74%
communication	5	21.74%
entertainment	2	8.70%
social networks	1	4.35%
	0	0.00%

Fig. 4. Purpose of using electronic display devices

We can put a reasonable focus on the question: Do electronic screen devices affect your vision in your opinion? In the Questionnaire, the answer of 95.65% of the respondents is "Yes" (Fig.5). This shows that according to people of all ages in the survey, screens affect vision, but a minimum of knowledge is needed in people, what exactly they affect and how to protect themselves if this influence is threatening to the health of the eye. Here, however, we must refer to the specialists in the field of ophthalmology and neuroscience, how the light from the screen affects and whether it poses a danger to the human visual system. According to a publication in "The Negative Influence of Screens on the Development of Children" by Dr. Valentina Dimitrova-Syarova - a clinical neuropsychologist, the blue light from screens has a harmful effect on the eyes and in particular on the retina. The wavelengths emitted by screens trigger various chemical reactions that release toxic molecules into photoreceptor cells. Screen use is also contributing to a sharp increase in myopia, according to a study published in Ophthalmology [14].

Do you think electronic screen devices affect your vision?



Data	Response	%
yes	22	95.65%
no	1	4.35%

Fig. 5. Impact of Electronic Screen Devices on Vision

As it became clear above, the participants in the survey believe that electronic screen devices affect their vision. However, when asked in the following question number 12: Have you ever used protective glasses for working with a screen device or eye drops to relieve dry eyes, or used a shield for eye comfort from the settings of your smart devices?, the respondents said (48%) that they necessarily use protective glasses. 22% believe that modern smart devices have the necessary eye protection, and an equal number of 22% of the participants in the Survey said that they want to use glasses designed for working with screens, but do not know anything about them. 4% of the respondents believe that the light from the screens does not affect them in any way.

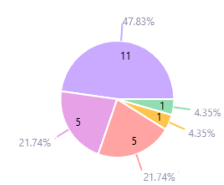
Eye protection technology such as the "Eye Comfort Shield" discussed above is generally considered by experts to reduce blue light and is a preventative intervention. However, research shows that most people rely on protective eyewear when working with electronic screen devices rather than using the software-based eye protection tools installed in modern smart devices.

The lack of information about modern digital screens (Fig.6.), such as that blue light activates melanopsin-containing retinal ganglion cells, which are involved in the regulation of circadian rhythms, and that increased stimulation suppresses melatonin secretion, which can lead to sleep disorders and secondary increases in cognitive and visual fatigue [8,9,10,11,12], is proving to be a major challenge to the use of electronic screen devices.

It is precisely the filtering of blue light that improves neuroendocrine balance, which is subjectively felt as "comfort".

The blue light emitted by new LED Screens slow down the secretion of melatonin (the sleep hormone), disrupt the biological clock and delay falling asleep[14].

Have you used safety glasses for working with a screen device or eye drops to relieve dryness, or a comfort eye shield from settings?

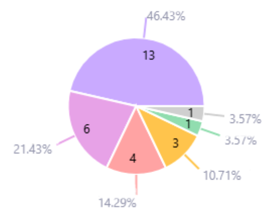


Data	Response	%
yes, a must	11	47.83%
not used, modern devices have provided protection	5	21.74%
I want to use glasses, but I'm not familiar with them	5	21.74%
I use drops, but they are prescribed to me	1	4.35%
I'm not affected by the light from the screen	1	4.35%
Other entries	0	0.00%

Fig. 6. Have you ever used safety glasses for working with a screen device, or eye drops to relieve dry eyes, or an eye comfort shield from settings on your smart devices?

Of great importance is the question in the Survey under number 9, related to how the respondents feel after using electronic screen devices? The answer options were 6 /satisfied, happy, inspired, tired, overexcited, depressed/. Nearly 47 percent or almost half of the participants in the survey of all ages, said that they feel tired after exposure to the screen light from smart devices and displays. This has its explanation from neuropsychologists. Again in the publication “The Negative Influence of Screens on the Development of Children” [14], the author and clinical neuropsychologist Dr. Valentina Dimitrova-Syarova writes that with supersensory stimulation, the vestibular apparatus is suppressed, which provokes an artificial lack of mood [14], which is probably why nearly 11% of the respondents in the survey responded that they feel overexcited, and almost 4 percent depressed. But this comparison is a research question related to more in-depth scientific studies.

How do you feel after using electronic screen devices?



Data	Response	%
tired	13	46.43%
satisfied	6	21.43%
happy	4	14.29%
overexcited	3	10.71%
depressed	1	3.57%
Other entries	1	3.57%

Fig. 7. How respondents feel after using electronic screen devices

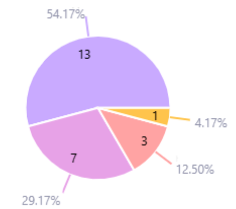
Nighttime exposure to LED-lit devices, such as most of today's screens (computers, phones, tablets, flat-screen TVs, e-readers, video games) suppresses melatonin secretion and disrupts natural sleep cycles. A good night's sleep is crucial for proper brain development [14].

Undoubtedly, the answers to question number 10 in the Survey summarize the real effect on individuals after using electronic screen devices, and whether respondents experienced mental discomfort or physical pain? Nearly 55% (Fig. 8.) or more than half of the respondents said that after using electronic screen devices they experienced mental discomfort or physical pain. Almost 30 percent did not share such feelings and discomforts, about 4% often experience psychological or physical discomfort, and nearly 13 percent of respondents did not even pay attention.

There is some controversy over whether excessive Screen Media Activity (SMA) is associated with problematic outcomes among youth and adolescents. While some have reported that frequent SMA is associated with internalizing psychopathology, including depression and anxiety, externalizing psychopathology, riskier behavior, and even suicide, others have found no evidence of a link between SMA and problematic outcomes. Even fewer studies have examined the relationship between different types of SMA and brain structure in healthy youth [15].

The conclusion here may be ambiguous. The majority of people experience psychological distress and physical pain, which is a sign that the impact of excessive use of electronic screen devices would lead to increasingly frequent negative consequences for our psyche and physique, with the caveat that there is no current and definitive scientific evidence yet that the human brain is at risk of structural changes, other than changes in attitudes, cognitive functions, sensations, emotions, and eye fatigue.

Have you ever experienced mental discomfort or physical pain after using electronic screen devices?



Data	Response	%
yes	13	54.17%
no	7	29.17%
I didn't pay attention	3	12.50%
often	1	4.17%

Fig. 8. After using electronic screen devices, participants experience mental discomfort or physical pain

IV. CONCLUSION

Electronic screen devices cannot lead to a degree of regulation of our brain and vision irreversibly, scientists are categorical, and this study has somewhat clarified which are the problems that should be addressed. One of them comes from excessive and unconscious use, the lack of boundaries and the neglect of our body's signals. In the digital world, the refusal of electronic screen devices is not on the agenda, but rather their balanced use. It is advisable to take advantage of the "digital detox" (the abrupt cessation of screen activity) [4], at least when possible. It is also imperative to think about more information in the education of children and adolescents, to create boundaries within which we spend time in front of screens and displays daily. And last but not least, to pay attention to how we feel after a prolonged stay in front of an electronic screen device.

APPENDIX

The results obtained in this report have scientific and applied value for assessing the risks of prolonged screen exposure and can help develop recommendations for healthy use of digital technologies. The research question about the digital world remains open, which has posed to scientists and specialists the solution of neuropsychological challenges after prolonged and uncontrolled use of screen devices.

REFERENCES

[1] Vinarova, M., Risks for Children on the Internet: The “Brain Rot” Phenomenon, Media and Language Journal 1(17),2025,https://www.researchgate.net/publication/394945963_Risks_for_Children_on_the_Internet_The_Brain_R

ot_Phenomenon#fullTextFileContent, DOI:10.58894/IUKH8464
[2] Newport Institute 2024: Brain Rot: The Impact on Young Adult Mental Health [online], https://www.newportinstitute.com/resources/co-occurring-disorders/brain-rot/, 2025
[3] Krustev, A., Digital Technologies and Mind Changes in Susan Greenfield's "Mind Change", 2014, Azcheta, https://azcheta.com/susan-greenfield-mind-change/
[4]Vezenkov,S.,Modern Myths about the Human Brain, "Sto Leta", Sofia, First Edition, Myth 19: Electronic Devices Develop Children's Brain, https://www.researchgate.net/publication/368789748_Mit_19_Elektronnite_ustrojstva_razvivat_detskia_mozk#fullTextFileContent, 2025
[5] Digital Fatigue: How Do Screens Affect Our Psyche?, April 2025, Tech News, NewTechGen.com
[6]Yan, Z., Jiaxin Ye, Gaohui Xie, Mingsheng Li, Ying Cao, Mingwei Ge, Ying Shen, Xiujian Zhu, Research and Optimization on Eye-Injury Issues of OLED Screens, July 2025, SID Symposium Digest of Technical Papers 56(S1):1742-1744, DOI:10.1002/sdtp.19202
[7]Zhang, T., Hu Li, Douqing Zhang, Liangliang Zheng, Hailin Xue, Xibin Shao, Zhiqiang Liu, Eye protection display technology based on infrared light, July 2025, SID Symposium Digest of Technical Papers 56(S1):1381-1383, DOI: 10.1002/sdtp.19092
[8]Rosenfield, M., Computer vision syndrome: a review of ocular causes and potential treatments. Ophthalmic and Physiological Optics, 2016
[9]Sheppard, A. L., & Wolffsohn, J. S., Digital eye strain: prevalence, measurement and amelioration. BMJ Open Ophthalmology, 2018
[10]Chang, A.M., Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness. PNAS, et al. 2015
[11]Wilkins, A.J., LED lighting flicker and potential health concerns, IEEE, et al. 2010
[12]American Optometric Association(AOA), Digital Eye Strain
[13]Popova, B., What is an "Eye Comfort Shield" and how is it adjusted? - divna.Tech © 2025
[14]Dimitrova-Syarova, V. The negative impact of screens on children's development, Neuropsychological research and consultations, February 2023, https://valentinadimitrova.com/negativno-vliyanie-na-ekranite-detsa/
[15]Paulus, M., Lindsay M. Squeglia, Kara Bagot, Joanna Jacobus, Rayus Kuplicki, Florence J. Breslin, Jerzy Bodurka, Amanda Sheffield Morris, Wesley K. Thompson, Hauke Bartsch, Susan F. Tapert, Screen media activity and brain structure in youth: Evidence for diverse structural correlation networks from the ABCD study, NeuroImage, Volume 185, January 2019, Pages 140-153, https://doi.org/10.1016/j.neuroimage.2018.10.040

Complex Control Systems

ISSN: 2603-4697 (Online)