# LLM-Based Automated Audit of Holter ECG Reports for Guideline Adherence and Quality Control

Mikhail Chmelevsky[1,2,3], Aleksey Popov[4], Margarita Budanova[1], Maria Sorokina[3], Ulyana Voronina[3], Alexey Tregubov[3], Andrey Kozlenok[1]

[1]Almazov National Medical Research Centre, Saint Petersburg, Russia
[2]Pavlov First Saint Petersburg State Medical University, Saint Petersburg, Russia;
[3]Saint Petersburg State University, Saint Petersburg, Russia
[4]Moscow Clinical Research and Practical Center, City Clinical Hospital No. 52, Moscow Department of Health, Russia

Interpretation of long-term ambulatory electrocardiographic monitoring reports is a cognitively demanding expert task characterized by high information density, multi-parameter integration and susceptibility to diagnostic, terminological and logical inconsistencies. Manual quality control is labor-intensive and poorly scalable. We developed a web-based platform for automated audit of Holter ECG reports using large language models within a formalized quantitative framework for report quality assessment.

The study aimed to evaluate the feasibility of an LLM-based system for automated assessment of Holter ECG report quality with respect to report completeness, adherence to guidelines, internal logical consistency and clinically relevant interpretation errors.

The platform was designed for automated analysis of free-text Holter ECG reports. The pipeline combined document parsing, section segmentation, semantic extraction of arrhythmological findings, rule-based validation of mandatory report fields and LLM-based expert-level assessment of interpretation quality. The language-model layer performed contextual semantic normalization, clinical entity extraction, contradiction detection, guideline-constrained reasoning, and structured error attribution.

For each report i, the number of detected errors was denoted by $n_i$, the presence of at least one error by $E_i \in \{0,1\}$ and the severity grade of error j by $g_{ij} \in \{1,\ldots,5\}$. Quality was quantified by the proportion of error-free reports, $P_0 = N_{\text{no error}}/N$, mean errors per report, $\bar{e} = (1/N)\sum_i n_i$, mean errors among erroneous reports, $\bar{e}_+ = (\sum_i n_i)/(\sum_i E_i)$, and a normalized Quality Index, $QI = 1 - (\sum_i \sum_j g_{ij})/M_{max}$, integrating both frequency and severity of errors. This framework transformed expert qualitative assessment into reproducible quantitative metrics suitable for continuous audit and between-physician comparison.

In an automated audit of 100 Holter ECG reports, complete processing and quality evaluation were performed in 3 min 42 s, whereas manual expert review was estimated to require approximately 4-5 h. The dominant error class was diagnostic misinterpretation of arrhythmias (about 60-65% of detected issues) followed by quantitative omissions such as pauses or episode duration (30-35%), terminological errors (25-30%), logical inconsistencies between presented events and final conclusions (20-25%) and technical defects including unreadable ECG fragments (15-20%). LLM-based automated audit of Holter ECG reports is feasible and enables rapid, scalable and clinically meaningful quality control.