

AI-Powered ECG interpretation: Only trusted, externally-validated and approved tools can turn innovation into practice

Description

Artificial Intelligence (AI) has gained rapid attention over the last years in cardiology scientific literature. One promising area of application is ECG automatic interpretation, as ECG signals are, by nature, more prone to be analyzed by machines than by humans [1]. The hopes are about enhanced efficiency-reducing time spent to analyzing visually complex signals-and improving diagnostic accuracy. Yet, despite a vast number of publications showing good performance, AI tools frequently fail to translate into clinical practice. Indeed, clinicians want more than results reported in scientific articles, they need to rely on regulatory-approved medical devices, for example with FDA clearance or CE marking, that they can use confidently in clinical practice. To reach this level of trust, demonstration of good performance through robust external validation is essential. These validations evaluate performance across independent patient populations, devices, and clinical sites not involved in training, ensuring real-world generalizability. Here we discuss why external validation is vital and highlight some recent studies reporting such results with regulatory-approved devices aiming at analyzing ECG recordings.

Regulatory approval is essential to bring AI to practice but not enough

Regulatory pathways serve as filters to ensure basic quality, safety, and clinical performance standards. In other words, products should perform as claimed by the manufacturer and should not harm the patients. However, especially for low-risk devices, substantial validation of AI methods in external settings is not mandatory and is rarely undergone. Many regulatory submissions are based on data handled by the manufacturer in a limited, often favorable environment, raising concerns around internal data biases. Performing external validation to test a trained model on entirely independent datasets from different sources is of paramount importance to demonstrate that device performance is as expected in different patients and environments. A lack of such validation remains a limiting factor for the hospital or physician to use a new medical device. In fact, the absence of true external validation is a common drawback in many academic AI-based cardiovascular models [2]. However, this should be tackled even more for products already in the market, in line with the regulators emphasizing that post-market evidence is essential. Finally, external validation results in the form of peer-reviewed articles enable a broader audience to assess credibility and transparency, which are core pillars of science, and should be among the main principles of the medical device business. It is worth mentioning that this quest for external validation is usually pursued whenever reimbursement or health technology assessment is targeted by manufacturers. Should we wait for this longer-term journey to perform external validation studies? We believe in including as best practice the development, execution

Manuel Martínez-Sellés^{1,2}, François De Guio^{3*}, Rubén Ruiz³, José Maria Lillo-Castellano^{3,4}, Manuel Marina-Breyse^{2,3,4}

¹Department of Cardiology, Universidad Complutense, Universidad Europea, Instituto de Investigación Sanitaria Gregorio Marañón, Hospital General Universitario Gregorio Marañón, Madrid, Spain

²Centro de Investigación Biomédica en Red. Enfermedades Cardiovasculares, Madrid, Spain

³Department of Cardiology, Idoven Research, Madrid, Spain

⁴Centro Nacional de Investigaciones Cardiovasculares, Myocardial Pathophysiology Area, Madrid, Spain

*Author for correspondence:
François De Guio, Department of Cardiology, Idoven Research, Madrid, Spain, E-mail: francois.deguio_ext@idoven.ai

Received date: 03-Apr-2025, Manuscript No. FMIC-25-167838;
Editor assigned: 05-Apr-2025, PreQC No. FMIC-25-167838 (PQ);
Reviewed date: 19-Apr-2025, QC No. FMIC-25-167838;
Revised date: 26-Apr-2025, Manuscript No. FMIC-25-167838 (R);
Published date: 05-May-2025, DOI: 10.37532/1755-5310.2025.17(3).1013

and publication of well-designed external validation studies since the start of the product ideation.

Regulatory-approved and externally validated ECG models: Some recent examples

Among thousands of works published every year describing AI models for ECG records automatic analysis, there are very few related to the external validation of regulatory-approved medical devices. The purpose here is not to systematically appraise existing studies but rather to consider a few examples to illustrate our statement about the need for external validation. External validation can be executed in several forms, depending on the product maturity and yielding different levels of rigor and clinical relevance. The most common method consists of validating the model on an entirely independent dataset from training, obtained in a different setting or time period, to evaluate its performance in a new population. Generalizability and potential overfitting to the original data can then be assessed along with the need to recalibrate initial models or not. One step further, more robust and clinically meaningful validation can be achieved through prospective designs, such as pragmatic Randomized Controlled Trials (RCTs), where the model is deployed in real-world clinical workflows and its impact is compared against standard care. These RCTs not only deal with models' accuracy but also evaluate their effectiveness in improving clinical outcomes, making the evidence even more powerful for clinical adoption.

Attia, et al. evaluated an AI model designed to detect left ventricular systolic dysfunction in a completely independent cohort from Russia [3]. Trained on Mayo Clinic data, the model was externally tested using the "Know Your Heart" dataset, achieving an area under the curve of 0.82 but with low sensitivity (26.9%). Interestingly, the study raised the need for population-specific threshold recalibration, emphasizing that even when a model meets regulatory standards, its performance may vary across different clinical environments. Validation cannot rely solely on internal test sets but must include diverse and representative populations to ensure generalizability.

We recently published an assessment of a CE-marked AI model for atrial fibrillation detection using over 8,500 publicly available single-lead ECGs from the PhysioNet Challenge [4]. High performance, including over 96% accuracy and strong F1-scores, was achieved despite the heterogeneity and noise inherent in real-world ECG recordings. These results were obtained without any new training or recalibration of the AI model. To note, the model was not previously trained on data from the same ECG hardware. As a post-market evaluation of an already approved product, this

study aligns with the evolving expectations from regulatory bodies, which increasingly emphasize the importance of ongoing evidence generation.

Lastly, Adedinsewo, et al. evaluated the performance of FDA-cleared AI algorithms embedded in both 12-lead ECG and digital stethoscope platforms in detecting low ejection fraction [5]. The tested AI algorithms had shown effectiveness before in different retrospective and pilot prospective studies in the US. Their purpose was to conduct a pragmatic RCT among pregnant and postpartum women to evaluate whether AI-guided screening improves the diagnosis of pregnancy-related left ventricular systolic dysfunction in an obstetric population in Nigeria, compared to usual care. The study demonstrated excellent diagnostic performance for the detection of a left ventricular ejection fraction <50%, with areas under the curve of 0.928 and 0.976 for the ECG and stethoscope-based models, respectively. AI-guided screening was associated with an increase in the diagnosis of cardiomyopathy compared to usual care. With this type of design, investigators may not only evaluate model performance but also clinical consequences and cost-effectiveness data into real-world workflows.

Conclusion

All together, these studies highlight that external validation is not a regulatory formality but a scientific and ethical imperative, ensuring that AI tools perform as expected when applied in the diverse and dynamic conditions of clinical care. External validation studies should be numerous, transparent, and adapted to different designs according to their objectives. Future directions to embrace this mindset may include further guideline development, inclusion of reporting standards within scientific journal guidelines and regulatory updates related to AI medical devices.

References

1. Martínez-Sellés M, Marina-Breyse M. Current and future use of artificial intelligence in electrocardiography. *J Cardiovasc Dev Dis.* 10(4):175 (2023).
2. Cai Y, Cai YQ, Tang LY, et al. Artificial intelligence in the risk prediction models of cardiovascular disease and development of an independent validation screening tool: A systematic review. *BMC medicine.* 22(1):56 (2024).
3. Attia IZ, Tseng AS, Benavente ED, et al. External validation of a deep learning electrocardiogram algorithm to detect ventricular dysfunction. *Int J Cardiol.* 329:130-135 (2021).
4. Guio FD, Rienstra M, Castellano JM, et al. Enhanced detection of atrial fibrillation in single-lead electrocardiograms using a cloud-based artificial intelligence platform. *Heart Rhythm.* 22(7):1667-1674 (2025).
5. Adedinsewo DA, et al. Artificial intelligence guided screening for cardiomyopathies in an obstetric population: A pragmatic randomized clinical trial. *Nat Med.* 30(10):2897-2906 (2024).