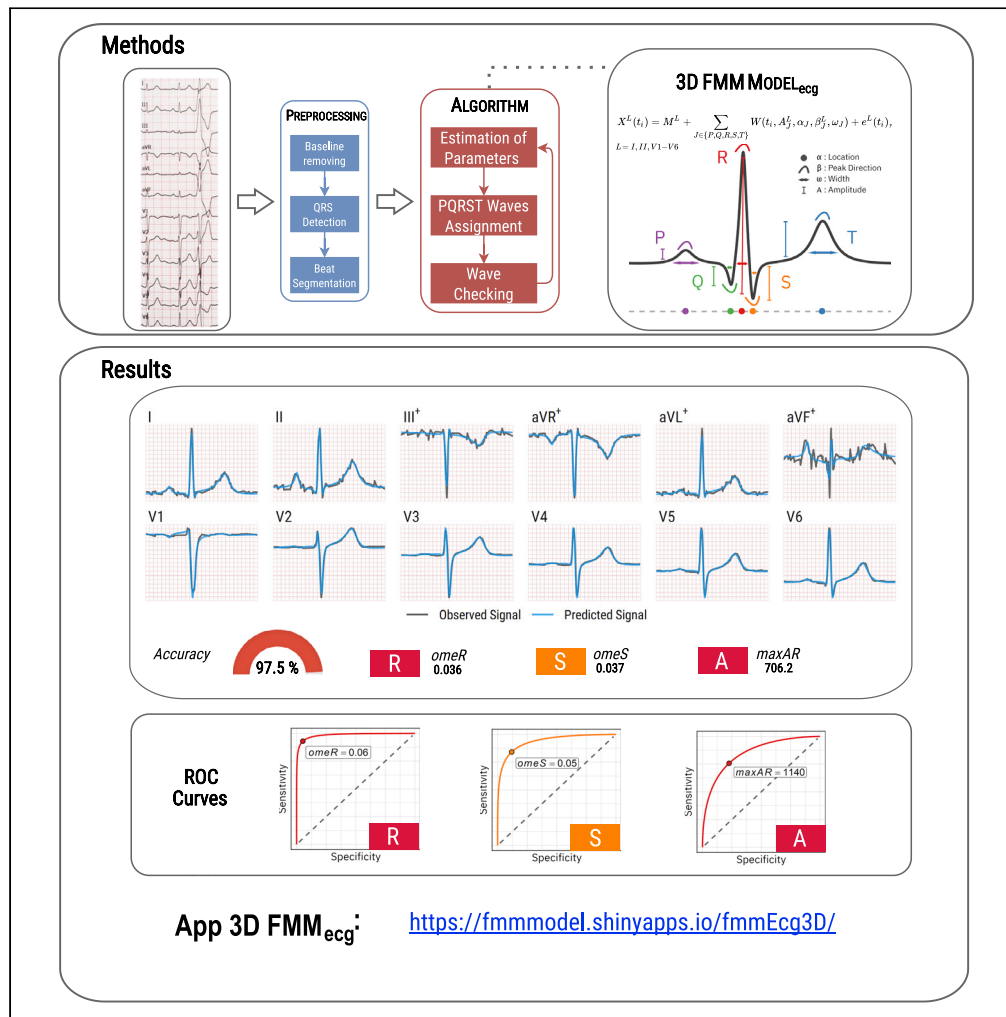## Article

A unique cardiac electrocardiographic 3D model. Toward interpretable AI diagnosis

Cristina Rueda,
Alejandro
Rodríguez-
Collado, Itziar
Fernández,
Christian Canedo,
María Dolores
Ugarte, Yolanda
Larriba

cristina.rueda@uva.es

Highlights

A solid mathematical modeling of the cardiac electric activity

The model accurately reconstructs, and generates, 12-lead ECG signals

Three markers for the diagnosis of LBBB, CBBB, and HYP pathologies

An APP analyses and diagnoses ECGs for any heart condition

Article

A unique cardiac electrocardiographic 3D model. Toward interpretable AI diagnosis

Cristina Rueda,[1,2,7,*] Alejandro Rodríguez-Collado,[1] Itziar Fernández,[1,3] Christian Canedo,[1] María Dolores Ugarte,[4,5,6] and Yolanda Larriba[1,2]

SUMMARY

Mathematical models of cardiac electrical activity are one of the most important tools for elucidating information about heart diagnostics. In this paper, we present an efficient mathematical formulation for this modeling simple enough to be easily parameterized and rich enough to provide realistic signals. It relies on a five dipole representation of the cardiac electric source, each one associated with the well-known waves of the electrocardiogram signal. Beyond the physical basis of the model, the parameters are physiologically interpretable as they characterize the wave shape, similar to what a physician would look for in signals, thus making them very useful in diagnosis. The model accurately reproduces the electrocardiogram signals of any diseased or healthy heart. This new discovery represents a significant advance in electrocardiography research. It is especially useful for diagnosis, patient follow-up or decision-making on new therapies; is also a promising tool for well-performing, transparent and interpretable AI approaches.

INTRODUCTION

The development of models and algorithms for the study of the cardiac electric system helps to better understand the physical processes governing the system and helps to guide therapeutic planning. In particular, the development of automatic methods with interpretable decisions is in high demand. The relevance of the topic has attracted the interest of scientists from different fields, such as mathematics, physics, bioengineering, and medicine.

The heart muscle is a composite tissue with a complex structure that consists of various cell types. The electric activation of the heart begins at the sinus node where pacemarker cells activate spontaneously. The corresponding current results in the excitation of the neighboring cells, and it then spreads, first along the atria, and then along the ventricles.[6,28]

Generalized assumptions on cardiac electrophysiological models that date back to the 1960s,[23] and remain part of the conventional approach are: the electric field is represented by a single or multiple dipoles, whereas the total electric activity is represented by a three-dimensional vector. This vector, denoted by $\vec{D}(t)$ at time t, is the sum of all the individual dipole vectors. Because the depolarization wavefront spreads through the heart, $\vec{D}(t)$ changes in magnitude and direction as a function of time. $\vec{D}(t)$ typically describes a trajectory with three loops corresponding to consecutive time segments: the P wave (atrial depolarization), the QRS complex (ventricular depolarization) and the T wave (ventricular repolarization), respectively. The loops described by the P and T waves are elliptical, whereas the QRS has an irregular shape.[25] Furthermore, the voltage measurements registered by the electrocardiogram (ECG) signal are the projections of $\vec{D}(t)$ in the directions of the axes of the recording electrodes located on the thoracic surface. In general, positive (negative) signals are produced when the depolarization front propagates toward (away from) a positive electrode. The opposite happens for the repolarization front. The standard ECG has signals from 12 projections or leads recorded using 10 electrodes.[6]

Despite the previous assumptions, mathematical formulations for ECG signals remain quite complex. Classical formulations include differential equation systems, representing the process with more or less biophysical detail. Aside from the mathematical complexity of the model formulation, some common criticisms of these models are that they hardly generate realistic 12-lead ECG signals, and that they depend on

[1]Department of Statistics and Operations Research, University of Valladolid, 47011 Valladolid, Spain

[2]Mathematics Research Institute of the University of Valladolid (IMUVA), 47011 Valladolid, Spain

[3]Biomedical Research Networking Center in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Carlos III National Institute of Health, 47011 Valladolid, Spain

[4]Department of Statistics, Computer Sciences and Mathematics, Public University of Navarre, 31006 Navarre, Spain

[5]Institute for Advanced Materials and Mathematics (INAMAT²), 31006 Navarre, Spain

[6]Healthcare Research Institute of Navarre (IdiSNA), 31008 Navarre, Spain

[7]Lead contact

*Correspondence: cristina.rueda@uva.es

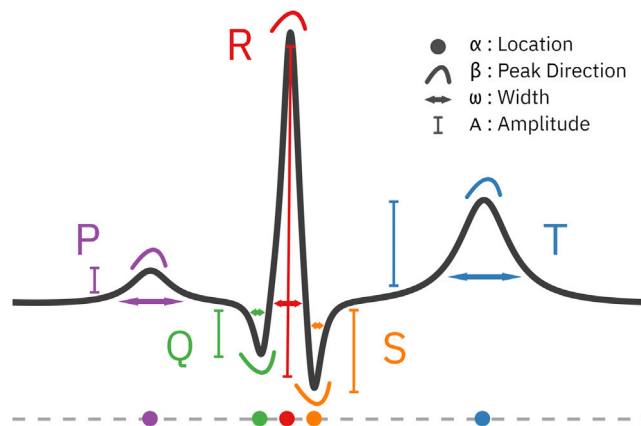https://doi.org/10.1016/j.isci.2022.105617

Figure 1. FMM parameter description
An ECG healthy heartbeat with waves decomposition and morphological parameter description

a large number of parameters which are hardly identifiable. In practice, a meaningful parameter identification is essential.

The literature dealing with dipole models for the electric activity of the heart is very extensive.[5,12–16,32,37,38,54,58] Alternative approaches dealing with specific aspects of the heart's electric activity are also too many to be easily summarised.[21,22,30,50–52] Furthermore, many papers since the early 1950s have been devoted to vectorcardiography (VCG) models.[25] Instead of 12-lead standard ECG models, they only consider three leads scanned in quasi-orthogonal axes. However, the VCG system is not common in clinical practice as there are fewer experts trained in these signals. ECG and VCG models have contributed to an improved understanding of the functioning of the heart, but have so far had little success in convincing clinicians. In particular, they fail in an important prerequisite for clinical applications, the ability to faithfully replicate ECG recorded from any diseased or healthy heart.

Here we present a novel model, named $3DFMM_{ecg}$ that is formulated for the 12-lead signals. The term 3D is used because it is assumed that these signals are projections of a three-dimensional electrical signal. The model relies on the classical physiological premise: the electric source is represented by a multiple dipole model. The novelty, however, is that it assumes that $\vec{D}(t)$ combines the electric signals from exactly five different sources, which represent differentiated myocardium segments, further associated with the five fundamental waves in ECG signals. Namely, $\vec{D}(t) = \vec{d}_P(t) + \vec{d}_Q(t) + \vec{d}_R(t) + \vec{d}_S(t) + \vec{d}_T(t)$, where t represents the real time and varies in $(0, 2\pi]$ for each heartbeat. Furthermore, the new model is also based on a key idea and original assumption: $\vec{d}_J(t)$, with $J \in \{P, Q, R, S, T\}$, moves in the same plane as time progress, and its trajectory is described as a complex FMM signal, a parameterized mathematical equation that describes an elliptical trajectory. Accordingly, the magnitude of $\vec{d}_J(t)$ in a given direction is represented by a one-dimensional FMM wave, which reflects depolarization and repolarization voltage changes. An FMM wave, $W(t, A, \alpha, \beta, \omega)$, is an equation defined in terms of four parameters.[44] The parameter A, a positive real number, measures the wave amplitude; α is a location parameter with values in $(0, 2\pi]$; whereas β, with values in $(0, 2\pi]$, describes the asymmetry degree of the wave shape and also informs when the unimodal pattern corresponds to a crest or a trough. Finally, the parameter ω, with values in $0, 1$, measures the sharpness of the peak. The value $\omega = 1$ corresponds to an exact sinusoidal shape. Figure 1 illustrates each parameter in a simulated ECG signal. These four parameters are wave-specific, but, whereas A and β are also lead-specific, α and ω are the same across leads. The role of these common parameters is essential in the modeling process. They provide connectivity between the signals from different leads thus, simplifying the model substantially. For a given lead L and wave J, the quadruple $[A_J^L, \alpha_J, \beta_J^L, \omega_J]$ describes precisely the wave shape, in the same way a physician looks at the signal. From the model definition it also follows that projections of $\vec{D}(t)$, in directions different from those in the standard 12-lead system, are also formulated as a sum of five FMM waves where α and ω do not change. Thus, the approach unifies VCG and ECG systems within a single formulation.

The inclusion of an intercept and an error term that accounts for data noise results in a powerful statistical model. To make the model identifiable and even more biologically interpretable, we assume that $\alpha_P \leq \alpha_Q \leq \alpha_R \leq \alpha_S \leq \alpha_T$. These restrictions respond to the natural depolarization order of different myocardium segments. These restrictions and the fact that the number of parameters is small, some of them common to different leads, guarantee an accurate identification of the model parameters, even in the cases of highly anomalous patterns or noisy signals. A quadratic optimization problem must be solved to obtain the model parameters. Here, we design a backfitting iterative approach to solve the problem. Specifically, the common parameters, α and ω, are identified using a grid search. Then, the lead-specific parameter estimates A and β are easily derived using standard linear regression models. An algorithm is designed to analyze 12-lead ECG fragments of any length. In the preprocessing stage, the signal is divided beat by beat. Then, for each beat, parameter estimates are derived. The output of the algorithm provides, for each wave, the series of parameter values, corresponding to consecutive beats, which can be summarized to get average patterns as well as the changes in the patterns over time. These quantities can be used directly in the clinic. For example, a specific measure of the QRS complex duration, a very well-known feature in cardiology, can be defined with the distance between α_Q and α_S. In the literature, there is no universal way to calculate this measurement. First, because the definition is vague: the QRS interval starts from the beginning of the Q wave to the end of the S wave. Second, because the observed signal is noisy, and finally, because different values may be derived from different leads. Compared to the commonly used measures of QRS duration, that based on FMM parameters have a more precise meaning and, more importantly, are not dependent on the lead, the measurement device, or the investigator eye.

The evaluation of the 3DFMM$_{ecg}$ model is an integral question but the evaluation technique and the model to compare with, depend on the specific tasks we want to solve. The 3DFMM$_{ecg}$ model is a multi-purpose approach and there is no an alike model to compare with. Indeed, there is no other multiple lead model, useful for automatic diagnosis, beyond black-box models. The better performance of the FMM single lead model against alternatives has been already shown in Rueda et al.,[43] in terms of global goodness of fit, consistency, robustness, and also for fiducial marks detection and pattern identification. In this paper, to evaluate the performance of the 3DFMM$_{ecg}$ model, we proceed as follows. First, we analyze the model's ability to reconstruct realistic ECG signals using 1D, 2D, and 3D representations generated from different parameter configurations. Second, all the ECG fragments from patients in the Database:PTB-XL[56] were analyzed with the new model. PTB-XL is a benchmark database, that contains a great variety of healthy, pathological, and noisy cases. To illustrate how the model works see Figure 2 and Figures S5–S12 for healthy and different heart conditions, respectively. We emphasize that the input and output of the 3DFMM$_{ecg}$ model are the 12-lead signals although is based on the assumption that these signals are projections of a 3D signal. Specifically, the model generates a predicted multivariate signal that mirrors the original 12-lead ECG signal with high reliability, as we quantify in the paper. This is something that other models are very far from reaching except for methods specifically designed for data compression or denoising, which are useless for solving other issues, particularly in automatic diagnostics. Third, the applicability of the model in clinical practice is illustrated with three markers, defined in terms of the model parameters, which are proved to be very useful for the detection of three well-known cardiac pathologies. Specifically, ROC curves from the Database:PTB-XL are given. These markers have at least the following advantages compared to black-box models: (1) The definition is very simple, it is applied to each patient individually and the result does not depend on the rest of the patients or their pathologies. In particular, no training set is used. (2) Its universal character allows the analysis of signals obtained with any recording device or preprocessing. (3) The cardiographic interpretation is clear and in accordance with standard medical diagnostic criteria. (4) The markers can be easily monitored and their evolution can be studied, providing valuable diagnostic information over time.

Furthermore, the model validation can be extended to any other dataset as an app is provided: https://fmmmodel.shinyapps.io/fmmEcg3D/, which allows the researchers to check the model with their own data. It includes a variety of examples that clearly illustrate how well our model reconstructs multivariate ECG signals, and help in diagnosis providing the markers values.

RESULTS

Next sections show how the 3DFMM$_{ecg}$ successfully solves the main challenges of the forward and inverse problems in electrocardiography, and give an overview of the contributions to interpretable AI. Furthermore, a simulation study that shows how the identification algorithm provides sensible and robust parameter estimates across different configuration scenarios was conducted (see STAR Methods). The analysis of
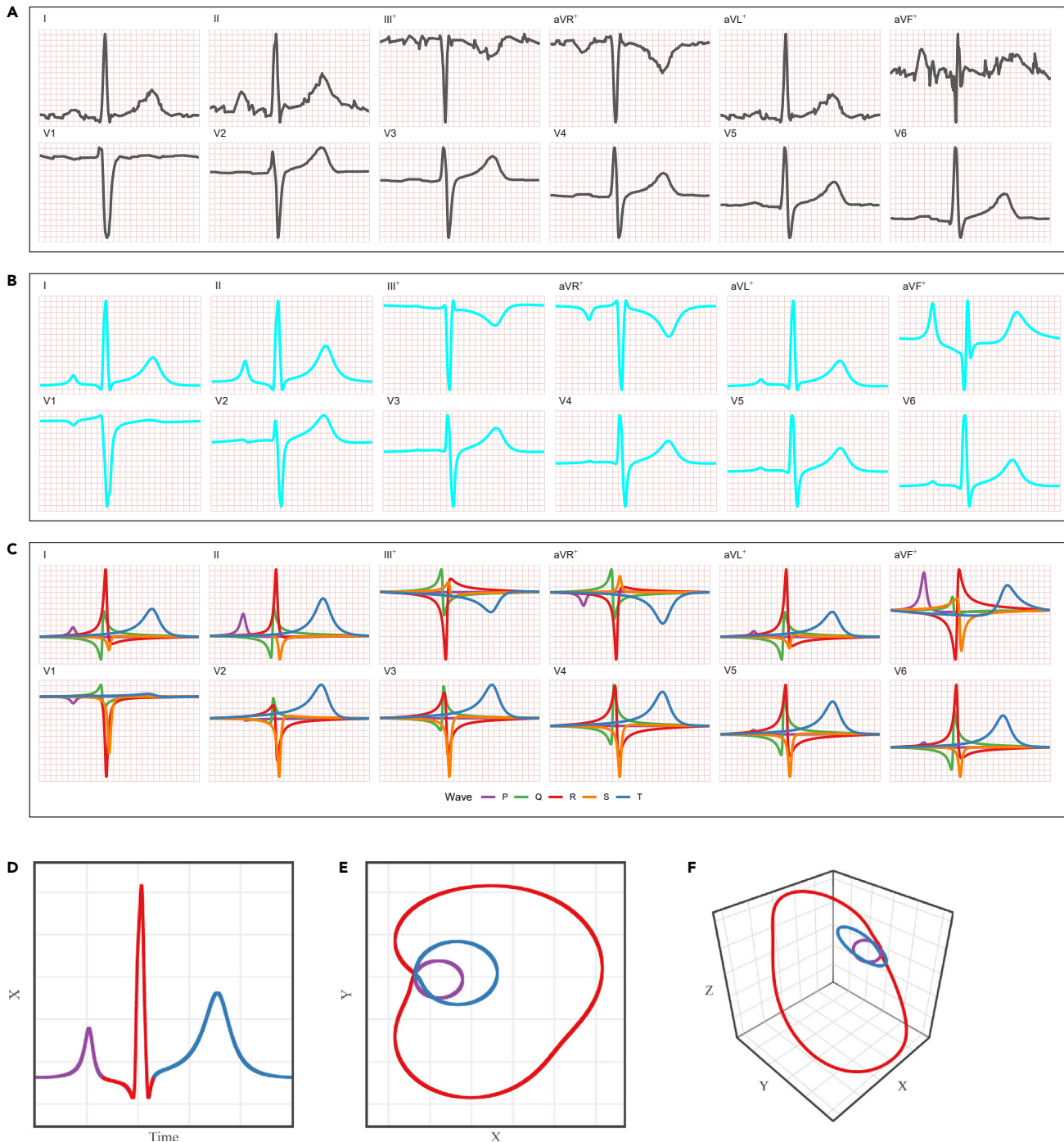
Figure 2. ECG from patient id 8 (beat n°9) in the Database:PTB-XL

(A) 12-lead observed signals (scaled).

(B) 12-lead FMM predicted signals.

(C) 12-lead FMM wave patterns. The symbol + indicates linear combinations of other signals.

(D), (E) and (F): 1D, 2D and 3D representations of the predicted signals, respectively, where $X = LeadII$; $Y = LeadII^{(i)}$; $Z = LeadV2 - 2 * Y$. The colors differentiate the P (purple), QRS (red) and T loops (blue)

Table1. Percentiles of $\overline{R}$, *omeR*, *omeS* and *maxAR* across NORM, CLBBB, CRBBB, HYP and ALL patients in Database:PTB-XL

		$\overline{R}$			o_{meR}			*omeS*			*maxAR*		
Diagnostic	N°	5*th*	Median	95*th*	5*th*	Median	95*th*	5*th*	Median	95*th*	5*th*	Median	95*th*
NORM	8,578	0.91	0.95	0.97	0.03	0.03	0.04	0.02	0.03	0.05	525	837	1,379
CLBBB	505	0.91	0.96	0.98	**0.07**	**0.14**	**0.21**	0.04	0.06	0.10	806	1,585	2,673
CRBBB	523	0.87	0.94	0.97	0.03	0.04	0.11	**0.04**	**0.08**	**0.14**	396	811	1,627
HYP	599	0.92	0.96	0.98	0.03	0.04	0.09	0.02	0.03	0.06	**852**	**1,465**	**2,403**
ALL	21,331	0.92	0.95	0.97	0.03	0.04	0.09	0.02	0.03	0.07	520	888	1,677

real ECG data requires a preprocessing stage to remove baseline and other noise artifacts. Data preprocessing is detailed in the STAR Methods.

Reconstruction of realistic ECG signals

A configuration of parameters for a healthy heart, labeled NORM, has been identified using the median values for all patients, labeled NORM in the Database:PTB-XL. This configuration is used as a reference. Alternative configurations describing different pathological conditions of the heart have been generated by changing the reference values of the most relevant parameters, summarized in Table S1, see the STAR Methods. The standard 12-lead unidimensional signals, and selected 2D and 3D representations, are provided for each configuration scenario. The axes we have selected for the 2D and 3D representations are defined as follows;

$$X = LeadII; Y = LeadII^{(i)}; Z = LeadV2 - 2 * Y$$

X is the most commonly used lead in studies, and (X, Y) is a natural bidimensional representation described by the Analytic Signal, defined in the complex plane[45,47] (additional details are given in the Methods Section). Finally, Z has been selected after checking other alternatives, as it offers an interesting visualization of the three loops: P, QRS and T across different patterns.

To conserve space, the figures for these configurations are shown in the Supplemental Information (see Figures S1–S4). In particular, the figure for the NORM configuration looks similar to Figure 2. The other configurations generate representations with characteristic and differentiated patterns.

Accuracy of 3DFMM$_{ecg}$ predictions and markers

To validate the performance of the 3DFMM$_{ecg}$ model in real practice, and in particular the three new markers, the ECG signals from the Database:PTB-XL have been analyzed. Using the preprocessing step, 210 patients (less than 1%) out of 21,837, were discarded because of very noisy patterns. In addition, 295 patients with pacemakers were discarded because of their highly atypical but predictable ECG patterns. A diagnostic label has been assigned to each patient in the database using SNOMED CT ontology,[34] (bioportal.bioontology.org/ontologies/SNOMEDCT). Labels are assigned only to patients with a diagnostic likelihood of 80 or more. Examples of observed and predicted signals from diverse heart conditions and noisy cases are given from Figures S5–S9 in Supplemental Information, illustrating the ability of the model to successfully handle pathological patterns. Some of the more recognized SNOMED labels used here are Complete Bundle Branch Blocks (CLBBB, CRBBB) left and right, respectively, and ventricular Hypertrophy (HYP).

ECG data at the patient level are a series of consecutive beats. The new markers are defined for a given patient, in a given center and device, and at a given moment of time; and these are obtained using median values of FMM parameters from individual beat values. Specifically, we define three new markers, namely: and *omeS* as the median of the ω_R andω_S individual beat values, respectively; and *maxAR* as the maximum, across leads, of the median of the A_R individual beat values.

Table 1 gives percentiles for different measures obtained for CLBBB, CRBBB, HYP and NORM patients. On the one hand, a global model quality measure, named $\overline{R}$, is given. It is defined as the mean across leads of anR^2 measure, which is, in turn, the proportion of variation in the signal that is explained by the model (see Methods section for details). The figures in Table 1show that the $\overline{R}$ values are quite high across diagnostics, being higher than 90% for more than 95% of patients. The only candidate models with which to compare the values of R^2 in Table 1 are the models designed for data compression or denoising. These models are calibrated in the literature using measures similar to R^2, such as $PRD (PR D^2 = 1 - R^2)$.
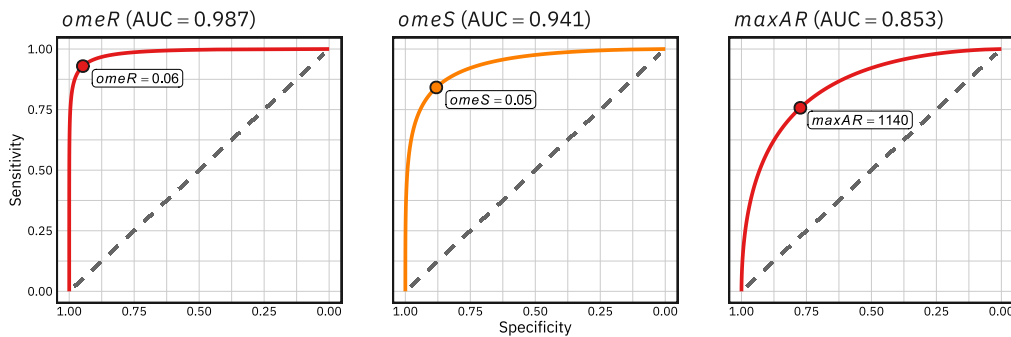
Figure 3. ROC curves and AUC values for the markers *omeR*, *omeS* and *maxAR* for CLBBB, CRBBB and HYP, respectively

PRD values of the order of 10% or less are considered sufficient to maintain the specific characteristics of an ECG signal.[31] The average *PRD* values were obtained using the 3DFMM models with 5–12 components, ranging from 20% to 9%, for the Database:PTB-XL. These values show that the FMM approach is also a very interesting method for data compression tasks while retaining diagnostics markers. On the other hand, Table 1 also shows that *omeR* and *omeS* are directly associated with CLBBB and CRBBB, respectively, whereas *maxAR* does with HYP, as long as *omeR* is within the percentile range of NORM. The interest of the *omega* FMM parameters in the diagnosis had already been evidenced in Rueda et al.,[42] where simple diagnostic rules for CLBBB and CBBB (Complete Bundle Branch Block) have been defined, and can be used even when only one of the lead signals is available. To better illustrate the diagnostic ability of binary classifier systems based on these markers, ROC curves have been plotted in Figure 3. There are many proposed rules in the literature to classify these diagnoses with surprisingly high AUC values in some cases.[48] The AUC values obtained with the FMM markers are not far from them whereas the rules based on the new markers have several advantages compared to black-box models, as commented in the Introduction.

Figures S10–S12 in the Supplemental Information, illustrate how the model works under other heart conditions.

Contributions to interpretable AI

Black-box Machine Learning (ML) algorithms rapidly growing in cardiology, lack interpretability and can cause problems in high-stakes decision making.[2,17,21,22,36,41,52,55,61] The ways the 3DFMM$_{ecg}$ approach helps to explain how ML models work are diverse but mainly rely on two powerful properties. On the one hand, the model provides a set of robust indexes that characterize in detail the morphology of ECG signals. These indexes can be incorporated as features in a classification or prediction task generating inherently interpretable ML models. Besides, they can be used to explain the decision of an ML algorithm with statistic summaries and visualization. In particular, the relevance of different waves and/or leads can be easily illustrated, identifying the heart regions related to the algorithm decision as well. On the other hand, the 3DFMM$_{ecg}$ model is able to reconstruct any real 12 lead ECG signal. This property allows us to elucidate how the ML model operates through well designed simulations, detecting possible shortcomings problems and identifying the causes of its decisions. Furthermore, a human-like interpretable and trustworthy AI FMM-based system could be developed by taking advantage of the ability of the FMM indexes to mimic almost any electrocardiographic feature, in such a way that the system is able to reproduce approximately the steps that a physician follows to read and interpret an ECG. Beyond the transparency and interpretability, other property that makes the 3DFMM$_{ecg}$ model a very promising tool is its universal character. It works regardless of the preprocessing and denoising, the recording device, the number of leads, the length of data, or the differences between datasets label distributions. It provides consensus indexes for relevant features such as the QRS complex duration that would also contribute to reduce the differences in the interpretation of electrocardiograms.

DISCUSSION

Here we present a unique cardiac electrophysiological model, the 3DFMM$_{ecg}$. The model is formulated under the assumption that the 12-lead signals are projections of a three-dimensional electric signal. Furthermore, the equations in the 3D model and in the univariate model in Rueda et al.[43] are identical. This is a great advantage of the 3D model as the generalization of the univariate model. The key of the 3D model is that the α's and ω's

parameters are common to the 12 leads. These parameter are estimated using information from all the leads, avoiding problems if, for example, one of the signals had large disturbances. In addition, a set of parameters, two for each wave and lead, complete the description of the signals. The estimation approach using the 3D model is much more robust, the model parameters are accurately estimated avoiding identifiability problems.

The 3DFMM$_{ecg}$ model also succeeds in providing reliable quantitative results to enhance the understanding of biophysical processes, starting with the α's and ω's parameters that represent the underlying electric signal and synchronize the waves across leads. Actually, all the parameters have a morphological interpretation and are reliable for extrapolation and prediction, thus simplifying the diagnostic task. Another crucial contribution of the 3D Model is that it permits to generate realist 12-lead ECG signals whereas those signals generated using univariate models do not necessarily correspond to projections of the same 3D underlying signal.

Furthermore, the results in the paper validate the premise that the electrical source can be represented with a five dipole model, on which the new model is built. Of interest, dipole models have been widely used in different forms in the literature. An additional motivation of this study is the capacity of the FMM approach for data compression, as minimal hard disk space is needed to fully reconstruct the signal. The monitoring of patients and, in particular, telemedicine requires the storage of ECG signals and a fast on-line transmission of information,[7,8] and therefore efficient data compression methods are in high demand. No other model had ever come close to achieving these goals.

Many challenges arise from this study. On the one hand, there is much work to do from a research perspective, such as the design of new diagnostic rules based on the FMM parameters; the identification of wave patterns specific to exceptional conditions such as atrial flutter, which do not adapt so well to the model; the integration of the 3DFMM$_{ecg}$ model in AI systems; or the inclusion of new terms accounting for different sources of variability. Specifically, we will start by studying a rule for HYP, using a combination of markers including the *maxAR* defined in this paper. On the other hand, there is the committed objective of getting the new markers to be used in clinical practice. To that end, devices that record the signals should provide the new markers, and more importantly, physicians need to be trained. Moreover, the current paper opens up great research opportunities for studying biological electric systems beyond the cardiac one, which have an impact on the knowledge of the effect and causes of multiple diseases, as well as the development of drugs and therapies. In particular, electric signals from other organs, such as the brain or the eyes, can be modeled using adapted 3DFMM models. Thus, the important question of analyzing the relationship between signals recorded in different regions, could also be addressed with these models. Some work in this line is in progress.

Limitations of the study

Regarding the limitations, the computational time is still too high to give interactive outputs. This is because of the exhaustive search and the backfitting loop, which are time-consuming processes. However, by implementing the current method in C and using general-purpose computing on graphics processing units, the computational cost could be greatly reduced. Furthermore, the algorithm, as described here, does not identify the *P* wave when it is not located before the *QRS* complex. New research into detecting and adapting the algorithm in these cases could bring about the correct identification of the *P* wave.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Data preprocessing
  - The 3DFMM$_{ecg}$ model
  - Evaluation metrics
  - Identification algorithm
  - Model validation
  - 3DFMM$_{ecg}$*app*

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.105617.

AUTHOR CONTRIBUTIONS

C.R. conceived the aims, theoretical proposal, mathematical model developments, and wrote the manuscript. C.R. and A.R-C. developed the design and implementation of the identification algorithm. C.R., Y.L., and I.F. designed the preprocessing stage and the simulations. A.R.-C. and C.C. developed the computational code and the app. M.D.U. provided critical feedback on the manuscript. All the authors revised and approved the manuscript.

DECLARATION OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

1. Agostinelli, C. and Lund, U. (2017).R Package Circular: Circular Statistics (version 0.4-93).

2. Al-Zaiti, S., Besomi, L., Bouzid, Z., Faramand, Z., Frisch, S., Martin-Gill, C., Gregg, R., Saba, S., Callaway, C., and Sejdić, E. (2020). Machine learning-based prediction of acute coronary syndrome using only the pre-hospital 12-lead electrocardiogram. Nat. Commun. *11*, 3966.

3. Aspuru, J., Ochoa-Brust, A., Félix, R.A., Mata-López, W., Mena, L.J., Ostos, R., and Martínez-Peláez, R. (2019). Segmentation of the ecg signal by means of a linear regression algorithm. Sensors *19*, E775.

4. Attali, D. (2020).Shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds. R Package Version 2.0.0.

5. Balakrishnan, M., Chakravarthy, V.S., and Guhathakurta, S. (2015). Simulation of cardiac arrhythmias using a 2d heterogeneous whole heart model. Front. Physiol. *6*, 374.

6. Bayes de Luna, A. (2007). Basic Electrocardiography (John Wiley & Sons).

7. Bhalerao, S., Ansari, I.A., Kumar, A., and Jain, D.K. (2019). A reversible and multipurpose ecg data hiding technique for telemedicine applications. Pattern Recognit.Lett. *125*, 463–473.

8. Bhaskar, S., Bradley, S., Sakhamuri, S., Moguilner, S., Chattu, V.K., Pandya, S., Schroeder, S., Ray, D., and Banach, M. (2020). Designing futuristic telemedicine using artificial intelligence and robotics in the covid-19 era. Front. Public Health *8*, 556789.

9. Canento, F., Lourenço, A., Silva, H., and Fred, A. (2013). Review and comparison of real time electrocardiogram segmentation algorithms for biometric applications. In Proceedings 6th Internatonal Conference Health Information(HEALTHINF).

10. Chang, W. and Borges, B. (2018).Shinydashboard: Create Dashboards with 'Shiny'. R Packageversion 0.7.1.

11. Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., and Borges, B. (2021).Shiny: Web Application Framework for R. R Packageversion 1.6.0.

12. Cheffer, A., Savi, M.A., Pereira, T.L., and de Paula, A.S. (2021). Heart rhythm analysis using a nonlinear dynamics perspective. Appl. Math. Model. *96*, 152–176.

13. Clayton, R.H., Bernus, O., Cherry, E.M., Dierckx, H., Fenton, F.H., Mirabella, L., Panfilov, A.V., Sachse, F.B., Seemann, G., and Zhang, H. (2011). Models of cardiac tissue electrophysiology: progress, challenges and open questions. Prog.Biophys. Mol. Biol. *104*, 22–48.

14. Corrado, C., Avezzù, A., Lee, A.W.C., Mendoca Costa, C., Roney, C.H., Strocchi, M., Bishop, M., and Niederer, S.A. (2021). Using cardiac ionic cell models to interpret clinical data. WIREs Mech. Dis. *13*, e1508.

15. Dasgupta, S., Das, S., and Bhattacharya, U. (2021). Cardiogan: an attention-based generative adversarial network for generation of electrocardiograms. In 2020 25th International Conference on Pattern Recognition (ICPR) (IEEE), pp. 3193–3200.

16. El Houari, K., Kachenoura, A., Albera, L., Bensaid, S., Karfoul, A., Boichon-Grivot, C., Rochette, M., and Hernández, A. (2017). A fast model for solving the ecg forward problem based on an evolutionary algorithm. In 2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) (IEEE), pp. 1–5.

17. Elul, Y., Rosenberg, A.A., Schuster, A., Bronstein, A.M., and Yaniv, Y. (2021). Meeting the unmet needs of clinicians from ai systems showcased for cardiology with deep-learning-based ecg analysis. Proc. Natl. Acad. Sci. USA *118*, e2020620118.

18. Fernández, I., Rodríguez-Collado, A., Larriba, Y., Lamela, A., Canedo, C., and Rueda, C. (2021). FMM: rhythmic patterns modeling by FMM models. R package version 0.3.0.

19. Fernández, I., Rodríguez-Collado, A., Larriba, Y., Lamela, A., Canedo, C., and Rueda, C. (2022). FMM: an R package for modeling rhythmic patterns in oscillatory systems. R J. *14*, 361–380.

20. Goldberger, A.L., Amaral, L.A., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., and Stanley, H.E. (2000). Physiobank, physiotoolkit, and physionet. Circulation *101*, E215–E220.

21. Han, X., Hu, Y., Foschini, L., Chinitz, L., Jankelson, L., and Ranganath, R. (2020). Deep learning models for electrocardiograms are susceptible to adversarial attack. Nat. Med. *26*, 360–363.

22. Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P., and Ng, A.Y. (2019). Computer-interpreted electrocardiograms: benefits and limitations. Nat. Med. *25*, 65–69.

23. Holt, J.H., Jr., Barnard, A.C., Lynn, M.S., and Svendsen, P. (1969). A study of the human heart as a multiple dipole electrical source: I. normal adult male subjects. Circulation *40*, 687–696.

24. Iannone, R., Allaire, J., and Borges, B. (2020). Flexdashboard: R Markdown Format for Flexible Dashboards. R package version 0.5.2.

25. Jaros, R., Martinek, R., and Danys, L. (2019). Comparison of different electrocardiography with vectorcardiography transformations. Sensors *19*, 3072.

26. Kulkarni, S., and Rajankar, S. (2016). Preprocessing techniques of electrocardiogram. Int. J. Eng. Comp. Sci. *5*.

27. Madeiro, J.P.V., Cortez, P.C., Marques, J.A.L., Seisdedos, C.R.V., and Sobrinho, C.R.M.R. (2012). An innovative approach of qrs segmentation based on first-derivative, hilbert and wavelet transforms. Med. Eng. Phys. *34*, 1236–1246.

28. Malmivuo, J., and Plonsey, R. (1995). Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetic Fields (USA: Oxford University Press).

29. Mondelo, V., Lado, M.J., Méndez, A.J., Vila, X.A., and Rodríguez-Liñares, L. (2017). Combining 12-lead ecg information for a beat detection algorithm. J. Adv. Theor. Appl. Inform. *3*, 5.

30. Nayak, S.K., Bit, A., Dey, A., Mohapatra, B., and Pal, K. (2018). A review on the nonlinear dynamical system analysis of electrocardiogram signal. J. Healthcare Eng.

31. Nêmcová, A., Smisek, R., Maršánová, L., Smital, L., and Vitek, M. (2018). A comparative analysis of methods for evaluation of ecg signal quality after compression. BioMed Res. Int.

32. Niederer, S.A., Lumens, J., and Trayanova, N.A. (2019). Computational models in cardiology. Nat. Rev. Cardiol. *16*, 100–111.

33. Pan, J., and Tompkins, W.J. (1985). A real-time qrs detection algorithm. IEEE Trans. Biomed. Eng. *32*, 230–236.

34. Perez Alday, E.A., Gu, A., J Shah, A., Robichaux, C., Ian Wong, A.K., Liu, C., Liu, F., Bahrami Rad, A., Elola, A., Seyedi, S., et al. (2021). Classification of 12-lead ecgs: the physionet/computing in cardiology challenge 2020. Physiol. Meas. *41*, 124003.

35. Perrier, V., Meyer, F., and Granjon, D. (2021). shinyWidgets: Custom Inputs Widgets for Shiny. R Packageversion 0.6.0.

36. Petch, J., Di, S., and Nelson, W. (2022). Opening the black box: the promise and limitations of explainable machine learning in cardiology. Can. J. Cardiol. *38*, 204–213.

37. Quarteroni, A.L.F., Manzoni, A., and Vergara, C. (2017). The cardiovascular system: mathematical modelling, numerical algorithms and clinical applications. Acta Numer. *26*, 365–590.

38. Quiroz-Juárez, M.A., Jiménez-Ramírez, O., Aragón, J.L., Del Río-Correa, J.L., and Vázquez-Medina, R. (2019). Periodically kicked network of rlc oscillators to produce ecg signals. Comput. Biol. Med. *104*, 87–96.

39. Rahul, J., Sora, M., and Sharma, L. (2019). Baseline correction of ecg using regression estimation method. In 2019 4th International Conference on Internet of Things (Smart Innovation and Usages (IoT-SIU)), pp. 1–5.

40. Rezk, S., Join, C., and El Asmi, S. (2011). An algebraic derivative-based method for R wave detection. In 19th European Signal Processing Conference (Barcelone, Spain: EUSIPCO 2011).

41. Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., and Zhong, C. (2022). Interpretable machine learning: fundamental principles and 10 grand challenges. Stat. Surv. *16*, 1–85.

42. Rueda, C., Fernández, I., Larriba, Y., Rodríguez-Collado, A., and Canedo, C. (2022). Compelling new electrocardiographic markers for automatic diagnosis. Comput. Methods Programs Biomed. *221*, 106807.

43. Rueda, C., Larriba, Y., and Lamela, A. (2021a). The hidden waves in the ecg uncovered revealing a sound automated interpretation method. Sci. Rep. *11*, 3724.

44. Rueda, C., Larriba, Y., and Peddada, S.D. (2019). Frequency modulated möbius model accurately predicts rhythmic signals in biological and physical sciences. Sci. Rep. *9*, 18701–18710.

45. RuedaSabater, C., Rodríguez-Collado, A., and Larriba, Y. (2021b). A novel wave decomposition for oscillatory signals. IEEE Trans. Signal Process. *69*, 960–972.

46. Sadhukhan, D., and Mitra, M. (2012). R-peak detection algorithm for ecg using double difference and rr interval processing. Procedia Technology *4*, 873–877.

47. Sandoval, S., and De Leon, P.L. (2018). The instantaneous spectrum: a general framework for time-frequency analysis. IEEE Trans. Signal Process. *66*, 5679–5693.

48. Sangha, V., Mortazavi, B.J., Haimovich, A.D., Ribeiro, A.H., Brandt, C.A., Jacoby, D.L., Schulz, W.L., Krumholz, H.M., Ribeiro, A.L.P., and Khera, R. (2022). Automated multilabel diagnosis on electrocardiographic images and signals. Nat. Commun. *13*, 1583.

49. Sannino, G., and De Pietro, G. (2018). A deep learning approach for ecg-based heartbeat classification for arrhythmia detection. Future Generat.Comput. Syst. *86*, 446–455.

50. Satija, U., Ramkumar, B., and Manikandan, M.S. (2018). A review of signal processing techniques for electrocardiogram signal quality assessment. IEEE Rev. Biomed. Eng. *11*, 36–52.

51. Schläpfer, J., and Wellens, H.J. (2017). Computer-interpreted electrocardiograms: benefits and limitations. J. Am. Coll. Cardiol. *70*, 1183–1192.

52. Siontis, K.C., Noseworthy, P.A., Attia, Z.I., and Friedman, P.A. (2021). Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nat. Rev. Cardiol. *18*, 465–478.

53. Sörnmo, L., and Laguna, P. (2006). Electrocardiogram (ECG) Signal Processing (John Wiley & Sons, Ltd).

54. Versaci, M., Angiulli, G., and La Foresta, F. (2020). A modified heart dipole model for the generation of pathological ecg signals. Computation *8*, 92.

55. Vijayakumar, D.T., Vinothkanna, M.R., and Duraipandian, D.M. (2021). Fusion based feature extraction analysis of ecg signal interpretation - a systematic approach. AICN J *3*, 1–16.

56. Wagner, P., Strodthoff, N., Bousseljot, R.-D., Kreiseler, D., Lunze, F.I., Samek, W., and Schaeffter, T. (2020a). Ptb-xl, a large publicly available electrocardiography dataset. Sci. Data *7*, 154.

57. Wagner, P., Strodthoff, N., Bousseljot, R.-D., Samek, W., and Schaeffter, T. (2020b). Ptb-xl, a Large Publicly Available Electrocardiography Dataset (Physionet). version 1.0.1.

58. Whittaker, D.G., Clerx, M., Lei, C.L., Christini, D.J., and Mirams, G.R. (2020). Calibration of ionic and cellular cardiac electrophysiology models. Wiley Interdiscip. Rev. Syst. Biol. Med. *12*, e1482.

59. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (New York: Springer-Verlag).

60. Wu, H.T., and Soliman, E.Z. (2018). A new approach for analysis of heart rate variability and qt variability in long-term ecg recording. Biomed. Eng. Online *17*, 54.

61. Xue, J., and Yu, L. (2021). Applications of machine learning in ambulatory ecg. Hearts *2*, 472–494.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
PTB-XL Database. Version: 1.0.2	https://doi.org/10.1038/s41597-020-0495-6	https://doi.org/10.13026/zx4k-te85
Software and algorithms		
R statistical language, version 4.4.1	The Comprehensive R Archive Network	https://www.r-project.org/
FMM package version 3.0.0	https://cran.r-project.org/web/packages/FMM/index.html	https://doi.org/10.32614/rj-2022-015
Data generation, simulations, analysis and figure production scripts	This paper	https://github.com/FMMGroupVa
Online app	This paper	https://fmmmodel.shinyapps.io/fmmEcg3D/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Cristina Rueda (cristina.rueda@uva.es).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- This work makes use of the Database:PTB-XL[57] that is publicly available at Physionet.[20]

- All the presented results are reproducible through the 3DFMM$_{ecg}$ app, publicly available at https://fmmmodel.shinyapps.io/fmmEcg3D/. A detailed description of the 3DFMM$_{ecg}$ app, including implementation and usage examples, is provided in the STAR Methods.

- The code is available at https://github.com/FMMGroupVa.

METHOD DETAILS

Data preprocessing

Electrocardiogram (ECG) signals are usually interfered with baseline drifts, noises, or conduction artifices which severely affect ECG signal analysis quality.[50] Data preprocessing is required both to achieve reliable ECG fragments and to divide them beat by beat.[9,26,53] However, the preprocessed signal may still be affected by other sources of variability, see Figure S8 in the Supplemental Information. The preprocessing is designed in four steps, as illustrated in Figure S13 of the Supplemental Information. The inputs of the algorithm are the raw ECG signals from the 12 leads. The four steps are: *1) Baseline removing*; *2) Single-leadQRSdetections and checking*; *3) Multi-leadQRSdetections*; and *4) ECG segmentation and checking*. Unlike most of the preprocessing algorithms,[3,9,49] the denoising step is dismissed from our proposal, due to the robustness of the 3DFMM$_{ecg}$ model. The details of the algorithm are given below.

Step 1: Baseline removing

For each ECG fragment, a low-frequency filter based on the fitting of a local polynomial regression model is used. Following the ideas in Rahul et al.,[39] these estimates subtracted to the signals which corrects the baseline wander.

Step 2: Single-lead QRS detections and checking

Signals are independently analyzed with the Pan Tompkins algorithm[33] to obtain lead-specific *QRS* detections.

- *QRS* Revision. The median *RR* length of the fragment is used as benchmark to remove too short/long detections[40,42,46] preventing false positive detections. *RR* is the time elapsed between two consecutive *QRS* detections.

Step 3: Multi-lead QRS detection

Single-lead *QRS* detections provided in *Step 2* are combined to obtain a common reference annotation set. Following the ideas given in Mondelo et al.,[29] the single detections are increasingly sorted to build groups of at least four *nearby* detections, in a way that the groups are *far enough* from each others. The median of the *nearby* detections within groups, called t^{QRS}, are computed. Hence, t^{QRS} are the annotations on the reference set, that is common across leads. This procedure ensures false positives detections and overcomes the problems caused by undetected *QRS* in some leads. One may refer to Mondelo et al.[29] for details. From now on, only the leads on the set $Lred = \{I, II, V1, V2, V3, V4, V5, V6\}$ will be considered.

Step 4: ECG segmentation and checking

As it was done in Rueda et al.,[42] the heartbeat segmentation across leads is given by $[t^{QRS} - 40\%RR, t^{QRS} + 60\%RR]$.

- ECG Segmentation Revision. To avoid meaningless heartbeats,[27,60] remove from the ECG fragment:

    1. The first (*resp. last*) beats which are notably shorter than the median *RR* length; or with a high percentage of initial (*resp. ending*) observations being constant.

    2. The beats for which t^{QRS} locates out of the $35\% - 45\%$ of the beat length.

    3. The beats with a marked difference between the signal endpoints regarding the beat amplitude.

    4. The beats that exhibit conduction issues with amplitudes exceeding the standard limits; or anomalous relative beat amplitudes regarding the signal.

ECG signals pass through the analysis *iff* there are at least i) three beats in the reference annotation set; ii) three leads with more than three beats; and iii) a 20% of expected beats according to sampling frequency. The latter conditions allow discarding unacceptable levels of noise or distorted signals.

The outputs of the preprocessing are the t^{QRS} annotations and ECG beat segmentation.

The 3DFMM$_{ecg}$ model

Without loss of generality, it is assumed that the time points are in $(0, 2\pi]$. In any other case, transform the time points $t' \in [t_0, T + t_0]$ to $t = \frac{(t' - t_0)2\pi}{T}$.

Let us define an FMM wave as follows: $W(t, A, \alpha, \beta, \omega) = A\cos(\varphi(t, \alpha, \beta, \omega))$, where A is the wave amplitude and,

$$\varphi\left(t, \alpha, \beta, \omega\right) = \beta + 2\arctan\left(\omega\tan\left(\frac{t - \alpha}{2}\right)\right), t \in \left(0, 2\pi\right] \qquad \text{(Equation 1)}$$

is the wave phase.

$W(t, A, \alpha, \beta, \omega)$ is suitable for describing rhythmic up-down-up (or down-up-down) patterns, as is well justified in Rueda et al.[44] The parameters characterise various morphological aspects of the wave, as detailed in the introduction.

The Analytic Signal (AS) of a unit FMM wave, $A = 1$, is defined using the Hilbert Transform (HT) as follows: $S(t) = \mu(t) + i\nu(t)$; where, $\mu(t) = W(t, 1, \alpha, \beta, \omega)$, and $\nu(t) = HT(\mu(t))$. The AS signal has interesting properties and researchers often assume that the underlying complex signal associated with an oscillatory process is an AS.[47] Now, the FMM complex signal is defined generalizing the AS of an FMM unit wave as: $S(t) = A_{Re}\mu(t) + iA_{Im}\nu(t)$, where $A_{Re}, A_{Im} \in \Re^+$. A complex FMM wave has elliptical trajectories. Furthermore, in this paper, the AS is also used to generate 2D and 3D representations of ECG

signals, specifically, we consider that when $\mu(t) = P(\vec{D}(t)|LeadII) = \sum_{J=1}^{m} A_J \cos(\varphi_J(t))$ then $\nu(t) = \sum_{J=1}^{m} A_J \sin(\varphi_J(t)) = P(\vec{D}(t)|LeadII^{(i)})$, where $P()$ denotes the projection operator.

The $3DFMM_{ecg}$ model relies on two premises. Namely; 1.- the heart electric field originates from a multi dipole; and 2.- each dipole is represented by a vector, which moves in a plane as time progress and is mathematically described as a complex FMM signal.

Below, we prove that the projection of such a vector is an FMM wave with the property that the values of α and ω do not depend on the direction of the projection.

Theorem 1

Let $\{\vec{d}(t), t \in (0, 2\pi]\}$ be a set of vectors in the same plane, $\wp$, describing a complex FMM signal, let the projection of $\vec{d}(t)$ in the principal axe of the ellipse, L, is an FMM wave, as follows;

$$P\left(\vec{d}\left(t\right)\middle|L\right) = M^L + W\left(t, A^L, \alpha^L, \beta^L, \omega^L\right), t \in \left(0, 2\pi\right]$$

Then, the projection of $\vec{d}(t)$ in any other direction, L', is also an FMM wave, verifying that $\alpha^{L'} = \alpha^L$ and $\omega^{L'} = \omega^L$.

Proof:

First, assume that $(0,0,0)' \in \wp$. Consider the orthogonal system $(X, Y, Z) = (L, L^\perp, Z)$, where $L, L^\perp \in \wp$ are the axes of the ellipse describing the trajectory of $\vec{d}(t), t \in (0, 2\pi]$; also define the center of the ellipse as $\vec{O} = (O_L, O_{L^\perp}, 0)$. Then, the coordinates of $\vec{d}_*(t) = (\vec{d}(t) - \vec{O})$ are $(x, y, z) = (A^L \cos(\theta(t)), A^{L^\perp} \sin(\theta(t)), 0)$. From the theorem assumptions consequently: $\theta(t) = \varphi(t, \alpha^L, \beta^L, \omega^L)$

Now, for any other direction L', there exist $\lambda^{L'}, \mu^{L'}, \kappa^{L'}$ such that the projection of a vector $\vec{v} = (x, y, z)'$ in L' is:

$$P(\vec{v}|L') = \lambda^{L'}x + \mu^{L'}y + \kappa^{L'}z. \qquad \text{(Equation 2)}$$

Then, consequently,

$$P\left(\vec{d}_*(t)\middle|L'\right) = \lambda^{L'}A^L \cos\left(\varphi\left(t, \alpha^L, \beta^L, \omega^L\right)\right) + \mu^{L'}A^{L^\perp} \sin\left(\varphi\left(t, \alpha^L, \beta^L, \omega^L\right)\right). \qquad \text{(Equation 3)}$$

and,

$$P\left(\vec{d}(t)\middle|L'\right) = P\left(\vec{O}\middle|L'\right) + \lambda^{L'}A^L \cos\left(\varphi\left(t, \alpha^L, \beta^L, \omega^L\right)\right) + \mu^{L'}A^{L^\perp} \sin\left(\varphi\left(t, \alpha^L, \beta^L, \omega^L\right)\right). \qquad \text{(Equation 4)}$$

Now, considering the trigonometric relation:

$$\lambda \cos(\theta) + \mu \sin(\theta) = \sqrt{\lambda^2 + \mu^2}\cos\left(\theta + atan\left(\frac{\lambda}{\mu}\right) + \frac{3\pi}{2}\right)$$

and applying this relation to equation *(4)*, where $\theta = \varphi(t, \alpha^L, \beta^L, \omega^L)$, $\lambda = \lambda^{L'}A^L$, and $\mu = \mu^{L'}A^{L^\perp}$, it is immediate to derive that,

$$P\left(\vec{d}(t)\middle|L'\right) = P\left(\vec{O}\middle|L'\right) + A^{L'} \cos\left(\varphi\left(t, \alpha^L, \beta^{L'}, \omega^L\right)\right)$$

where,

$$A^{L'} = \sqrt{\left(\lambda^{L'}A^L\right)^2 + \left(\mu^{L'}A^{L^\perp}\right)^2}; \beta^{L'} = \beta^L + atan\left(\frac{\lambda^{L'}A^L}{\mu^{L'}A^{L^\perp}}\right) + \frac{3\pi}{2}$$

and the result of the theorem follows when $M^L = 0$.

Now, when $(0,0,0)' \notin \wp$, the projections in any direction are FMM waves, but they also include an intercept term that accounts for the location of $\wp$ relative to the origin. Specifically, let $z_0 = P((0,0,0)|\wp)$, for each $\vec{d}(t) \in \wp$, then $\vec{v}(t) = \vec{d}(t) - z_0 \in \wp_{d_0}$, where $\wp_{d_0}$ is a plane passing through the origin.

Therefore, $P(z_0|L) = M_0^L$ and $P(\vec{v}(t)|L) = O_L + A^L \cos(\varphi(t, \alpha^L, \beta^L, \omega^L))$.

Finally,

$$P\left(\vec{d}(t)\big|L'\right) = M^{L'} + A^{L'} \cos\left(\varphi\left(t, \alpha^L, \beta^{L'}, \omega^L\right)\right)$$

where, $M^{L'} = M_0^{L'} + P(\vec{O}|L')$, and the theorem follows. □

Now, consider the tridimensional space with its origin of the central point in the chest and the voltage recorder with the ten electrodes generating the 12-lead ECG signals: $Lset = \{I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5, V6\}$. Each of these signals is the projection of $\vec{D}(t)$ in a given direction and, in turn, the projection of $\vec{D}(t)$ is the sum of the projections of the five dipole vectors, for each of which the result of the theorem can be applied individually.

Finally, the 3DFMM$_{ecg}$ model is derived, taking into account all the above considerations. Specifically, the ECG unidimensional signals are formulated as a signal plus error model in Definition 1, where the error term accounts for artifacts in the data and the intercept accounts for location changes as follows.

Definition 1. 3DFMM$_{ecg}$ model

Let $X^L(t_i)$, $t_1 < \ldots < t_n$, be an observation from lead L, $L \in Lset$. Then,

$$X^L(t_i) = M^L + \sum_{J \in \{P, Q, R, S, T\}} W\left(t_i, A_J^L, \alpha_J, \beta_J^L, \omega_J\right) + e^L(t_i); \qquad \text{(Equation 5)}$$

where, for $L \in Lset$, and, $J \in \{P, Q, R, S, T\}$:

- $M^L \in \Re$,

- $\beta_J^L \in (0, 2\pi]$,

- $\alpha_P \leq \alpha_Q \leq \alpha_R \leq \alpha_S \leq \alpha_T \leq \alpha_P$,

- $\omega_J \in [0, 1]$,

- $A_J^L \in \Re^+$,

- $(e^L(t_1), \ldots, e^L(t_n))' \sim N_n(\mathbf{0}, \sigma^L \mathbf{I})$.

It is relevant to note that restrictions imposed on the α parameters, response to the assumption that the atrial depolarization is previous to the ventricles depolarization. Occasionally, the atria may repolarize later. In such a case, the P wave would not go before the QRS. To prevent these cases, the model can also be defined by numbering the waves and making a subsequent identification of the numbers with the letters, maintaining the relationship between the QRS and the T.

Other important parameters of practical use are the peak and trough times. They are defined as functions of the basic parameters.[43] In addition, other indices, such as the distances between waves, are easily derived from the set of basic 3DFMM$_{ecg}$ model parameters. Obtaining good estimators of these indices is crucial for clinicians because they are useful tools in the diagnosis.

Evaluation metrics

In this study, a global model quality measure, named $\overline{R}$, is defined for a given patient. Namely,

$$\overline{R} = \frac{1}{12} \sum_{L \in Lset} median\left(R_{L_b}^2\right),$$

where, for a fixed beat b and lead L,

$$R_{L_b}^2 = 1 - \frac{\sum_{i=1}^n \left(X_b^L(t_i) - \widehat{X_b^L}(t_i)\right)^2}{\sum_{i=1}^n \left(X_b^L(t_i) - \overline{X_b^L}(t_i)\right)^2}$$

is the proportion of variation in the signal that is explained by the predicted values. The higher the $\overline{R}$, the better the model.

The *PRD*, a related measure that is widely used in the data compression literature is defined as follows:

$$PRD_{L_b} = \sqrt{\frac{\sum_{i=1}^{n}\left(X_b^L(t_i) - \widehat{X_b^L(t_i)}\right)^2}{\sum_{i=1}^{n}\left(X_b^L(t_i) - \overline{X_b^L}(t_i)\right)^2}} = \sqrt{1 - R_{L_b}^2},$$

Identification algorithm

The identification problem reduces to solving the following optimization problem:

$$Min_{\theta \in \Theta} \sum_{L \in Lset} \frac{1}{\sigma^L} \sum_{i=1}^{n}\left[X^L(t_i) - \mu^L(t_i, \theta)\right]^2,$$

where θ is the vector of the model parameters, Θ is the parametric space, and $\mu^L(t_i, \theta) = M^L + \sum_{J \in \{P,Q,R,S,T\}} W(t_i, A_J^L, \alpha_J, \beta_J^L, \omega_J)$. For a typical ECG pattern, Θ is initially defined as in Definition 1. However, Θ is reduced when atypical or very noisy patterns are analyzed to achieve a correct physiological identification of waves. The values σ^L, $L \in Lset$ are identified as part of the optimization process.

For computational efficiency, the *Lset* is reduced to *Lred* = {*I,II,V1,V2,V3,V4,V5,V6*}, as the rest of the lead signals are linear combinations of *I* and *II*. However, the weight of *I* and *II* increase by a factor of 3 in order to maintain the global weight of the derivations from the frontal plane.[6]

From a statistical point of view, the optimization problem defined above is that of finding the Maximun Likelihood Estimates of the parameters. We adapt an iterative algorithm[45] to the multivariate setting. The algorithm alternates M and I steps. It is assumed that the preprocessing stage provides a *QRS* annotation point, denoted as t^{QRS}, for each beat. Moreover, exceptionally, one or more than one, of the eight leads are discarded in the preprocessing stage due to noise artifacts; only information on at least one of *I, II, V2,* or *V5* lead is required. In those cases, the identification stage is conducted with the selected leads, providing estimates for the common parameters; while the estimates for the lead-specific parameters of missing leads are derived by solving a standard multiple linear regression problem at the end of each M step.

The M step obtains $K \geq 5$ FMM waves using a backfitting algorithm and the I step assigns $K \leq 5$ letters to, at most, five of these waves. Typically, $K = 5$; however, in the presence of significant noise or when the morphology is pathological, it is possible that the interesting waves may be hidden between the sixth and seventh waves (exceptionally up to the tenth). The values σ^L are updated in each iteration of the algorithm, averaging the squared difference between the expected values and the predicted values by the model. They are initially set to 1.

M step

A standard backfitting algorithm is designed by fitting a single FMM wave simultaneously to the leads in *Lred*. The fitting of a single FMM is repeated successively to the residuals. The numbers of backfitting passes programmed in each step M is five. The final estimates of M^L, A_J^L and β_J^L; $L \in Lred$; $J \in \{P, Q, R, S, T\}$, are derived by solving a standard multiple linear regression problem.

I step

The R wave is assigned in the first I step. It corresponds to the one with the highest explained variability among those closest to t^{QRS}, which also has a positive peak on leads *I* or *II* and a negative peak in lead *V2*. In the few cases where these assumptions do not hold, additional conditions are used. Next, the preassignation of *P, Q, S* and *T* to the free components among the first five is done using $\alpha_P \leq \alpha_Q \leq \alpha_R \leq \alpha_S \leq \alpha_T$. This preassigment corresponds, in most cases, to the final assignment. However, in the presence of significant noise, or when the morphology is pathological, the interesting waves may sometimes be hidden between the sixth and seventh waves, which are checked for reassignments using thresholds on the main model parameters. In particular, noisy components are detected with too small or too high ω values.

The algorithm finishes when there is no significant increase in the objective function of the optimization problem.

In addition, the predictions from other leads in the frontal plane, *III, aVL, aVR* and *aVF* are derived using the predictions of the *I* and *II* leads, using the known linear relations between them.[6] In the case that any of the *I* and *II* leads is tentatively eliminated in the preprocessing step, the estimates for leads in the frontal plane can be derived by solving a standard multiple linear regression problem, as is done with the missing leads in step M. A flow chart of the algorithm is given in Figure S14 of the Supplemental Information.

Model validation

A simulation study is performed to evaluate the properties of the parameter estimators provided by the identification algorithm across different ECG signal patterns. Four different parameter configurations defining the $3DFMM_{ecg}$ signal are considered, representing healthy (NORM), complete left bundle-branch block (CLBBB), complete right bundle-branch block (CRBBB) and hypertrophy (HYP) hearts. Figures S1–S4 show the simulated 3D signals and Table S1 gives the values of the main parameters of the different configurations. The parameter for the NORM configuration corresponds to the median values of the patients labeledNORM in PTB-XL database. As it is shown in Table S1, the main changes in CLBBB, CRBBB and HYP in relation to NORMare ω_R, ω_S and *maxAR*, respectively.

For each configuration, 100 synthetic beats are simulated at 300 equally time spaced points using the $3DFMM_{ecg}$ model (see Definition 1 of the main text). The error term is generated using a normal distribution with mean zero and lead-dependent SD. The choice of SD values is based on the residual SD from NORMsubjects of PTB-XL database. Depending upon the lead, three values of SD are considered: $\sigma^L = 25$ for $L \in \{I, II\}$, $\sigma^L = 20$ for $L \in \{V1, V2, V3\}$ and $\sigma^L = 12$ for $L \in \{V4, V5, V6\}$.

Three measures of the estimation performance are shown in Table S2: *bias*, defined as the average difference between the true parameter value and its estimate across the 100 simulated beats in each ECG pattern; empirical SE(*empSE*), computed as the SD of the estimates over the 100 synthetic beats; and a coefficient of variation (*CV*) that, for the euclidean parameters $\theta \in \{\omega_J, maxAR\}$, $J \in \{P, Q, R, S, T\}$ is defined as,

$$CV_\theta = \sqrt{\frac{\frac{1}{N}\sum_{j=1}^{N}\left(\theta_0 - \widehat{\theta}_j\right)^2}{\theta_0^2}} \qquad \text{(Equation 6)}$$

whereθ_0 is the true parameter value and N the number of simulated beats. For the angular parameters, α_J, this *CV* is defined as

$$CV_\theta = \sqrt{\frac{\frac{1}{N}\sum_{j=1}^{N}\left(\left(\widehat{\theta}_j - \theta_1\right) - \pi\right)^2}{\pi^2}} \qquad \text{(Equation 7)}$$

where $\theta_1 = \theta_0 - \pi$. For the angular parameters, circular mean and SD are computed using the `circular` R package.[1] The median and percentile ranges of global model quality measure, $\overline{R}$, are shown in Table S3.

Estimates for α and ω's parameters exhibit a bias near zero and a very small empirical SE ECG patterns, pointing that the proposed algorithm provides high accurate and precise parameter estimates. As expected, the *CV*s are low in these cases, below the 10% for almost all of the estimates. The higher *CV* values are observed for ω_Q in the HYP pattern and for ω_S in the CLBBB pattern, both near an acceptable 30%. Although, in absolute terms, the bias of the *maxAR* estimates is higher, this bias is small with respect to the true parameter value, ranging from 600 to 1700. It is worth noting that this parameter is the maximum, across leads, of the A_R values, so bias is to be expected. Additionally, this estimate has a very high precision. Furthermore, by observing the distribution of this estimate in PTB-XL (see Table 1), it is clear that such biases have no effect on the ability of *maxAR* as a diagnostic marker. Regarding the values for the $\overline{R}$, in all patterns are above the 97%, showing an excellent match between the predicted and simulated ECG signals.

The performance of the $3DFMM_{ecg}$ in real cases for healthy and pathological heart conditions is illustrated in Figures S5–S12 of the Supplemental Information.

3DFMM$_{ecg}$app

The 3DFMM$_{ecg}$app has been developed using `shiny`[11] from the programming language `R`. Shiny applications are mainly composed of two modules: the frontend or user interface, designed with `HTML` widgets and `CSS` elements, and the backend or server, responsible for the computational tasks implemented in `R`. User interface flexibility is achieved through several `R` packages.[4,10,24,35,59] While the backend has been mainly implemented using the `R` package `FMM`.[18,19]

ECG data can be analyzed in the *app* by uploading the data in tabular format (extensions.csv,.xls, or.xlsx), and indicating the sampling rate in Hertz (Hz). Data must contain multiple beats recorded in the standard twelve leads, and include at least one of *I, II, V2,* or *V5* leads. The more of these derivations provided, the better the estimates obtained will be. Each column of the file should correspond to a lead and the column header names should indicate the lead name (labeled *I, II, III, aVL, aVR, aVF, V1, V2, V3, V4, V5,* or *V6*).

Once the file is uploaded, data is preprocessed and fragmented in beats using the preprocessing described in this document. Then, the user must select the beat to be analyzed, which is displayed in the app interface, and the number of backfittings desired in the estimation algorithm. In particular, a higher number of backfittings can improve the predictions at the cost of a higher computational time. Finally, the analysis starts by pressing the prediction button. The following outputs are given:

- The predicted FMM signal plotted along with the input data for each lead.
- Individual waves plot for each lead. Plot of estimated $W(t, A_J^L, \alpha_J, \beta_J^L, \omega_J)$, $L \in Lset$, $Lset = \{I, II, III, aVL, aVR, aVF, V1, V2, V3, V4, V5, V6\}$, $J \in \{P, Q, R, S, T\}$, $t \in (0, 2\pi]$.
- Accuracy measure, in particular $100 \times \overline{R}$.
- The estimates for the relevant parameters given in Tables S1 and S2.

To illustrate the *app* functionality, four ECG fragments from PTB-XL database are provided, labeled NORM, CLBBB, CBBB and HYP, respectively. The recordings sampling rate is 500 Hz, default value in the *app*. In particular, signals are from patients 976 (NORM), 11,174 (CLBBB), 10,938 (CBBB), and 15,260 (HYP).