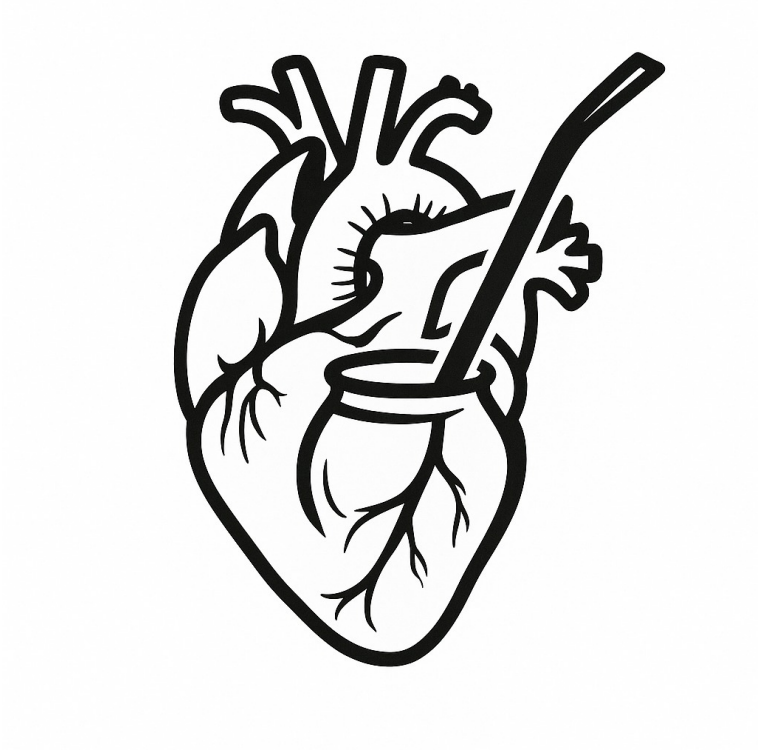
MATE

Machine Approach To ECG

Emirhan Kayar
Longato Alessandro
Cabrini Lodovico
Biagi Libero



June 9, 2025

Contents

1 Device Description and Specification **2**
- 1.1 General Description of the AI-based Device . . . 2
- 1.2 Technical Specifications . . . 2
- 1.3 Designation / Classification . . . 3

2 Design and manufacturing information **4**
- 2.1 Description of the design . . . 4
- 2.2 Description of the AI development . . . 5
  - 2.2.1 Development . . . 5
  - 2.2.2 General Logic . . . 5
  - 2.2.3 Architecture . . . 6
  - 2.2.4 Data Requirements . . . 7
  - 2.2.5 Human Oversight Measures . . . 7
  - 2.2.6 Validation and Testing . . . 7

3 General Safety and Performance Requirements **8**
- 3.1 AI Performance requirements . . . 9
  - 3.1.1 Gender Fairness assessment . . . 9
  - 3.1.2 Center Fairness assessment . . . 10
  - 3.1.3 ALTAI Requirements . . . 11

4 Benefit and risk-analysis **14**

5 AI Lifecycle **15**
- 5.1 Scoping . . . 15
- 5.2 Data . . . 15
- 5.3 Modelling . . . 15
- 5.4 Deployment . . . 16

6 Technical documentation on post-market surveillance **16**

1 Device Description and Specification

1.1 General Description of the AI-based Device

Intended Purpose

MATE (Machine Approach To ECG) is a predictive AI algorithm, whose aim is to aid clinicians in the identification of various types of arrhythmia, using preprocessed ECGs acquired from patients. The model is trained to identify four broad categories of rhythms:

- Atrial Fibrillation (AFIB = label 0)
- Grouped Supraventricular Tachycardia (GSVT = label 1)
- Sinus Bradycardia (SB = label 2)
- Sinus Rhythm (SR = label 3)

Where SR represents the healthy baseline.
The algorithm is supposed to process 12 lead ECG signals with sampling frequency of 500 Hz: using signals different from the one required may have unexpected consequences, of which the provider cannot be claimed guilty. The software is intended to be used as a support for qualified clinicians and cannot substitute the opinion of a professional.

1.2 Technical Specifications

The device is an AI classifier that, given the patient's ECG signal, produces one of four labels (AFIB, GSVT, SB, SR), corresponding to an healthy rhythm or to the family of cardiac diseases to which the patient's ECG belongs.
The AI system is supposed to be run in combination with a portable 12 lead ECG monitor, to provide immediate predictions of pathologies related to heart rhythm abnormalities.

Software Requirements

In order to be able to run the software, Python 3.12 is required, along with the following Python libraries:

- signal-grad-cam
- tensorflow version 2.18
- openpyxl
- pandas
- PyQt5
- pyqtgraph
- pyqt-svg-button
- absresgetter
- scikit-learn
- matplotlib

User Interface

The user interface is composed of four main components, which are:

- ECG list (top left)
- Control buttons (bottom left)
- ECG plot (top right)
- Information panel (bottom right)

1) The user will have to load the folder containing the saved ECG using the first button on the left: once the folder has been loaded and the correct signal selected, the plot will be available and the information panel will display any metadata available (i.e. age, sex, etc.).
2) To get the model's prediction, the user will have to press the last button: the output will be displayed on the information panel.
3) After the prediction, the output may be saved using the third button, which will allow the user to generate a plot showing which sections of the ECG are considered most relevant by the model for the prediction: this plot may be saved as a PDF file using the fifth button.
4) To unmount the current ECG folder, one may use the second button.
Note: if more than one model will be made available for performance reasons, the preferred option may be selected using the dropdown menu at the top of the information panel.

1.3 Designation / Classification

- Being a software intended by the manufacturer to be used by human beings for the diagnosis and prediction of diseases, MATE falls into the definition of Medical Device provided by the article 2(1) of MDR, which states that " 'medical device' means any instrument, apparatus, appliance, software, implant, reagent, material or other article intended by the manufacturer to be used, alone or in combination, for human beings for one or more of the following specific medical purposes: diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease ", more precisely a Medical Device Software (MDSW), not involving the analysis of biological samples (In Vitro Diagnostics).

- According to the MDR classification described in the Annex VIII 6.3 rule 11, which states that "Software intended to provide information which is used to take decisions with diagnosis or therapeutic purposes is classified as class IIa, except if such decisions have an impact that may cause: a serious deterioration of a person's state of health or a surgical intervention, in which case it is classified as class IIb.", our device belongs to class IIb, as the information it provides is used to make decisions with diagnostic purposes, which may cause serious deterioration or a surgical intervention.

- Coherently with the AI Act Recital 52, stating that "As regards stand-alone AI systems, namely high-risk AI systems other than those that are safety components

of products, or that are themselves products, it is appropriate to classify them as high-risk if, in light of their intended purpose, they pose a high risk of harm to the health and safety or the fundamental rights of persons, taking into account both the severity of the possible harm and its probability of occurrence", and with the EU AI Act Article 6.1, stating that "Irrespective of whether an AI system is placed on the market or put into service independently of the products referred to in points (a) and (b), that AI system shall be considered to be high-risk where both of the following conditions are fulfilled:

- (a) the AI system is intended to be used as a safety component of a product, or the AI system is itself a product, covered by the Union harmonisation legislation listed in Annex I;
- (b) the product whose safety component pursuant to point (a) is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment, with a view to the placing on the market or the putting into service of that product pursuant to the Union harmonisation legislation listed in Annex I."

with our system falling under the products present in Annex I as "11. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (OJ L 117, 5.5.2017, p. 1)." we have our system classified as high risk.

2 Design and manufacturing information

2.1 Description of the design

Phases

- Problem specification
  - Dataset research, focus on data quality and relevance of the chosen topic.
  - Possible approaches to the task.
- Data engineering
  - Choice of the denoised ECGs for the signal-based architectures, including min-maxing of the signals.
  - For the tabular version of the dataset: cleaning, features selection, encoding and normalization.
- Prototype drafts
  - Assessment of the capabilities of various architectures with respect to the task, ending with the selection of the most robust.
  - Key architecture choice: CNN.
- Best Model Selection

  - Different configurations of the model were tried, ending with the selection of the one displaying better performances.

- Fairness Analysis
  - Investigation of possible source of bias coming from the "gender" variable.
  - Slight better performances for women.

- External Validation
  - Assessment of the capabilities of the best found model on an external dataset, namely, [1].

- User Interface Development
  - Data loading support, explainability integration and pdf print option.

2.2 Description of the AI development

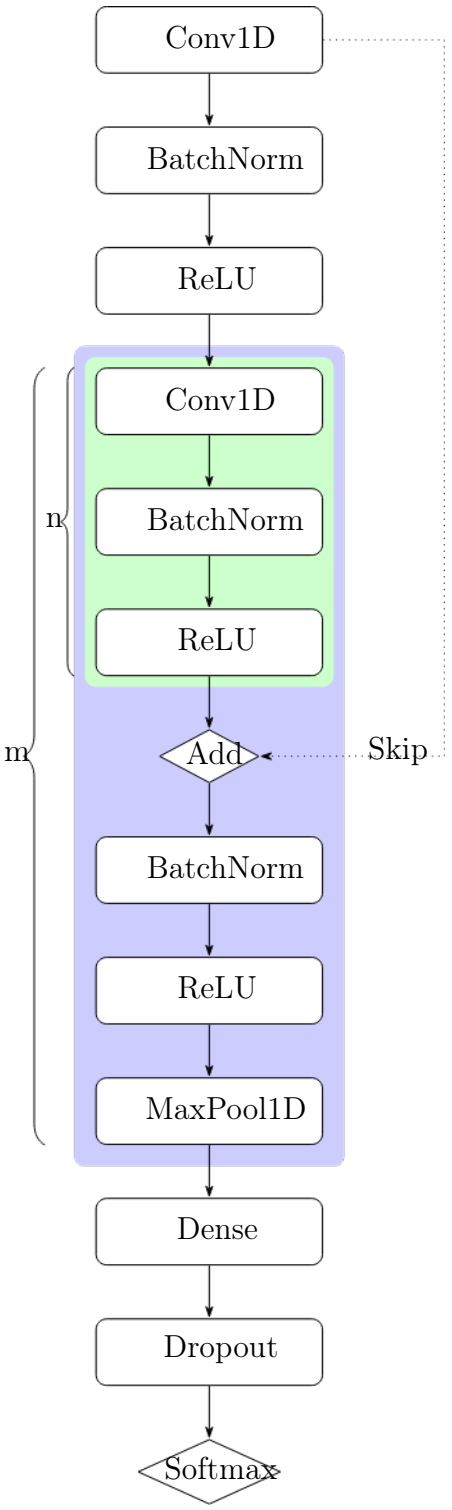
2.2.1 Development

In order to select the best possible architecture, a number of approaches were tried, including Extreme Gradient Boosting, TabNet, Random Forest Classifier, Support Vector Machine and Multilayer Perceptron for the tabular configuration of the data and Convolutional Neural Networks/Recurrent Neural Networks for the denoised ECG signals. The dataset [3] was split in train, validation and test set, and the models were assessed using holdout validation. Eventually, the tabular approach was discarded for not achieving the necessary performance. Finally, the best convolutional model was chosen among the ones showing better performances.

2.2.2 General Logic

The AI system we provide is a 1-dimensional Convolutional Neural Network (CNN) with skip-connections, which allows for the analysis of time series data such as ECGs. The output of the system is one of the four labels mentioned at point 1.1.

2.2.3 Architecture



The network begins with a Conv1D layer followed by BatchNorm and ReLU activation for initial feature extraction from input sequences.

The core consists of n residual blocks, each containing Conv1D, BatchNorm, and ReLU layers. A skip connection bypasses these blocks, enabling gradient flow and preventing vanishing gradients in deeper networks.

The model includes m pooling stages with BatchNorm, ReLU, and MaxPool1D operations. This hierarchical structure progressively reduces sequence length while preserving

important features.

Final layers comprise a Dense layer, Dropout for regularization, and Softmax for probability distribution over output classes.

- **Residual Learning**: Skip connections facilitate training of deeper architectures
- **Hierarchical Processing**: Multi-scale feature learning at different temporal resolutions
- **Regularization**: BatchNorm throughout and Dropout prevent overfitting

2.2.4 Data Requirements

As already stated, it is important that the input data resembles the data on which the model has been trained, that is, 12-lead ECG signals with a sampling rate of 500 Hz. The data-set used for training was provided by [3], whose study aimed to create a high-quality electrocardiogram database for research on arrhythmias. The data is publicly available at `https://doi.org/10.6084/m9.figshare.c.4560497.v2`, and is composed of 10,646 patient ECGs including 5,956 males and 4,690 females, with 17% of those having normal sinus rhythm and 83% having at least one abnormality. We followed the labeling choices that the paper's authors made, that is, mapping AFIB (Atrial Fibrillation), AF (Atrial Flutter) to the AFIB category, SVT (Supraventricular Tachycardia), AT (Atrial Tachycardia), SAAWR (Sinus Atrium to Atrial Wandering Rhythm), ST (Sinus Tachycardia), AVNRT (Atrioventricular Node Reentrant Tachycardia), AVRT (Atrioventricular Reentrant Tachycardia) to the GSVT category, SB (Synus Bradycardia) to the SB category and SR (Sinus Rhythm), SI (Sinus Irregularity) to the SR category.

2.2.5 Human Oversight Measures

In accordance with article 14 of the EU AI Act, measures to enable human oversight were implemented. The explainability of the system has been made possible thanks to the work of [2], publicly available at `https://github.com/bmi-labmedinfo/signal_grad_cam`. Once the model has produced the prediction for a given ECG, the clinician can press the graph button to visualize the relevance of the various parts of the signal for the model's output.

2.2.6 Validation and Testing

Table 1: Confusion Matrix (External - Model 7)

True \ Pred	**0**	**1**	**2**	**3**
0	0.85	0.12	0.00	0.02
1	0.14	0.86	0.00	0.00
2	0.01	0.00	0.96	0.04
3	0.05	0.04	0.13	0.78

From the confusion matrix is possible to see that the performance on the classes ranges from 0.78 to 0.96 implying a good accuracy regarding the true labels. The performance shows that MATE is robust model also on external data.

3 General Safety and Performance Requirements

Evidence of compliance with the General Safety and Performance Requirements:

- 15.1 “Diagnostic devices and devices with a measuring function, shall be designed and manufactured in such a way as to provide sufficient accuracy, precision and stability for their intended purpose, based on appropriate scientific and technical methods. The limits of accuracy shall be indicated by the manufacturer.”
  - The requirement is applicable due to the diagnostic nature of the device
  - For further details regarding the satisfaction of the aforementioned requirement, please refer to AI Performance Requirements and AI Development sections.
- 17.2 “For devices that incorporate software or for software that are devices in themselves, the software shall be developed and manufactured in accordance with the state of the art taking into account the principles of development life cycle, risk management, including information security, verification and validation.”
  - The requirement is applicable due to our device being a software
  - For further details regarding the satisfaction of the aforementioned requirement, please refer to sections Description of AI Development and AI Lifecycle.
- 23.4 "The instructions for use shall contain all of the following particulars:"
  (b) "The device's intended purpose with a clear specification of indications, contra-indications, the patient target group or groups, and of the intended users, as appropriate.",
  (c) "Where applicable, a specification of the clinical benefits to be expected.",
  (e) "The performance characteristics of the device.",
  (f) "where applicable, information allowing the healthcare professional to verify if the device is suitable and select the corresponding software and accessories.",
  (g) "Any residual risks, contra-indications and any undesirable side- effects, including information to be conveyed to the patient in this regard.",
  (q) "For devices intended for use together with other devices and/or general purpose equipment: — information to identify such devices or equipment, in order to obtain a safe combination, and/or — information on any known restrictions to combinations of devices and equipment." and
  (ab) "For devices that incorporate electronic programmable systems, including software, or software that are devices in themselves, minimum requirements concerning hardware, IT networks characteristics and IT security measures, including protection against unauthorised access, necessary to run the software as intended.".
  - Refer to section 1.1
  - Refer to section 2.2.6
  - Refer to section 1.2
  - Refer to section 1.1
  - Refer to section 3.1
  - Refer to section 1.2
  - Refer to section 1.2

3.1 AI Performance requirements

As seen in the previous section the majority of misclassifications happened on class 3, where the patient is healthy. A misdiagnosis can give a wrong insight, but the effect should be minimized, as the counsel is to be taken as a suggestion. The explainability was ensured via the use of Signal GradCam [2]. We remind that, in accordance with the article 14.5 of the EU AI Act, no action or decision should be taken by the deployer on the basis of the identification resulting from the system unless that identification has been separately verified and confirmed by at least two natural persons with the necessary competence, training and authority.

3.1.1 Gender Fairness assessment

Here there are the reports on female and male patients

Table 2: Classification Report (Test Data) - Women

Class	**Precision**	**Recall**	**F1-score**	**Support**
0	0.59	0.88	0.70	432
1	0.26	0.87	0.40	116
2	0.26	0.93	0.41	193
3	1.00	0.80	0.89	5085
Accuracy		0.81		5826
Macro avg	0.53	0.87	0.60	5826
Weighted avg	0.93	0.81	0.85	5826
AUC		0.9673		

Table 3: Classification Report (Test Data) - Men

Class	**Precision**	**Recall**	**F1-score**	**Support**
0	0.72	0.83	0.77	576
1	0.29	0.86	0.44	113
2	0.28	0.97	0.43	263
3	1.00	0.76	0.86	4056
Accuracy		0.78		5008
Macro avg	0.57	0.86	0.62	5008
Weighted avg	0.91	0.78	0.82	5008
AUC		0.9553		

As it's possible to see from the data the model has better results on women than men.

3.1.2 Center Fairness assessment

Table 4: Classification Report (Test Data) - Germany

Class	Precision	Recall	F1-score	Support
0	0.18	0.98	0.31	46
1	0.00	0.00	0.00	0
2	0.00	0.00	0.00	0
3	1.00	0.78	0.87	6432
Accuracy		0.78		6478
Macro avg	0.30	0.44	0.30	6478
Weighted avg	0.99	0.78	0.87	6478

Table 5: Classification Report (Test Data) - China

Class	Precision	Recall	F1-score	Support
0	0.83	0.84	0.84	880
1	0.01	1.00	0.01	2
2	0.00	0.00	0.00	0
3	0.97	0.50	0.66	830
Accuracy		0.67		1712
Macro avg	0.45	0.58	0.38	1712
Weighted avg	0.90	0.67	0.75	1712

Table 6: Classification Report (Test Data) - USA

Class	Precision	Recall	F1-score	Support
0	0.41	0.91	0.57	82
1	0.82	0.86	0.84	227
2	0.96	0.96	0.96	456
3	0.99	0.93	0.96	1879
Accuracy		0.93		2644
Macro avg	0.79	0.92	0.83	2644
Weighted avg	0.95	0.93	0.94	2644

The classification reports across centers (Germany, China, USA) show significant variation in performance. While the model performs very well in the USA, with high precision and recall across all classes, the performance in Germany and China is notably lower, especially for minority classes. This suggests potential center-related bias, possibly due to differences in data distribution, acquisition protocols, or sample imbalance. Further calibration or domain adaptation techniques may be necessary to ensure equitable performance across all centers.

The appropriateness of the metrics is due to the broad view of the performances they offer, allowing to assess the trade-offs of the model.

3.1.3 ALTAI Requirements

1. **Human Agency and Oversight**

   - **Is the AI system designed to interact, guide or take decisions by human end-users that affect humans or society" Could the AI system generate confusion for some or all end-users or subjects on whether a decision, content, advice or outcome is the result of an algorithmic decision? Are end-users or other subjects adequately made aware that a decision, content, advice or outcome is the result of an algorithmic decision?**

     MATE is based on deep learning techniques and professionals are made aware of that in the paragraph "Intended Purpose".

   - **Could the AI system affect human autonomy by generating over-reliance by end-users? Did you put in place procedures to avoid end-users over-rely on the AI system?**

     To avoid over-reliance, a disclaimer will be put on the instruction manual.

   - **Please determine whether the AI system (choose as many as appropriate): Is a self-learning or autonomous system; Is overseen by a Human-in-the-Loop; Is overseen by a Human-on-the-Loop; Is overseen by a Human-in-Command.**

     Is overseen by a Human-in-Command.

   - **Have the humans (human-in-the-loop, human-on-the-loop, human-in-command) been given specific training on how to exercise oversight?**

     The current technical documentation and the to-be-written instruction manual will provide enough information to the human user.

2. **Technical Robustness and Safety**

   - **Did you identify the possible threats to the AI system (design faults, technical faults, environmental threats) and the possible consequences?**

     Due to the design of our software, it is extremely reliant on the ECG monitor on which it has been mounted, which will need extensive controls to assure the correct acquirement of the signals. Unexpected results when dealing with new data are addressed by the post-market surveillance.

   - **Did you assess the risk of possible malicious use, misuse or inappropriate use of the AI system?**

     The risk of misclassification deriving from misuse of the software has been repeatedly addressed in the current documentation.

   - **Did you define safety criticality levels (e.g. related to human integrity) of the possible consequences of faults or misuse of the AI system?**

Refer to sections 1.2 and 3.1.

- **Did you align the reliability/testing requirements to the appropriate levels of stability and reliability?**

  Refer to section 2.2.

- **Could a low level of accuracy of the AI system result in critical, adversarial or damaging consequences?**

  Yes; however, this low level of accuracy should be prevented by the human supervision and the validation put in place for our model.

- **Did you put in place a series of steps to monitor, and document the AI system's accuracy?**

  Refer to section 2.2.

- **Did you put processes in place to ensure that the level of accuracy of the AI system to be expected by end-users and/or subjects is properly communicated?**

  Refer to section 3.1.

- **Could the AI system cause critical, adversarial, or damaging consequences (e.g. pertaining to human safety) in case of low reliability and/or reproducibility? Did you put in place a well-defined process to monitor if the AI system is meeting the intended goals? Did you test whether specific contexts or conditions need to be taken into account to ensure reproducibility?**

  Refer to section 2.2.

- **Did you put in place verification and validation methods and documentation (e.g. logging) to evaluate and ensure different aspects of the AI system's reliability and reproducibility?**

  Refer to section 2.2.

3. **Privacy and Data Governance**

   - **Did you consider the impact of the AI system on the right to privacy, the right to physical, mental and/or moral integrity and the right to data protection?**

     Input data isn't saved by the software and its management is responsibility of the healthcare center.

   - **Is your AI system being trained, or was it developed, by using or processing personal data (including special categories of personal data)? Did you consider the privacy and data protection implications of data collected, generated or processed over the course of the AI system's life cycle?**

     Training data was acquired from [3], who already provided to process the data, removing sensible information. Input data isn't saved by the software

and its management is responsibility of the healthcare center. Output data isn't stored, and its management is responsibility of the clinician.

4. **Transparency**

   - **Did you put in place measures to continuously assess the quality of the input data to the AI system? Can you trace back which data was used by the AI system to make a certain decision(s) or recommendation(s)? Can you trace back which AI model or rules led to the decision(s) or recommendation(s) of the AI system?**

     Guidelines for the correct acquisition of valid input data are provided in the current documentation. For information regarding model explainability, please refer to section 2.2.5.

   - **Did you explain the decision(s) of the AI system to the users?**

     For information regarding model explainability, please refer to section 2.2.5.

   - **Do you continuously survey the users if they understand the decision(s) of the AI system?**

     Refer to section 6.

   - **Did you establish mechanisms to inform users about the purpose, criteria and limitations of the decision(s) generated by the AI system? Did you communicate the benefits of the AI system to users? Did you communicate the technical limitations and potential risks of the AI system to users, such as its level of accuracy and/or error rates?**

     Refer to sections 1.1 and 3.1.

5. **Diversity, Non-discrimination and Fairness**

   - **Did you consider diversity and representativeness of end-users and/or subjects in the data? Did you identify the subjects that could potentially be (in)directly affected by the AI system, in addition to the (end-)users and/or subjects?**

     Please refer to previous points of section 3.1.

6. **Societal and Environmental Well-being**

   - **Could the AI system create the risk of de-skilling of the workforce? Did you take measures to counteract de-skilling risks?**

     As our software requires medical competences in order to be used effectively and to assess the meaningfulness of its predictions, the risk of deskilling is mitigated.

7. **Accountability**

- **Did you ensure that the AI system can be audited by independent third parties?**

  Our code is publicly available at
  - `https://github.com/Emirhankayar/ECGclassification`,
  - `https://github.com/Emirhankayar/MATECG-UI`,
  - `https://github.com/a-longato/ecg_mdlab/tree/main`

  whereas the data for training and validation is available at [3] and [1].

4 Benefit and risk-analysis

Establish and document a risk management plan for each device

Every six months (or under exceptional circumstances such as data drift or extraordinary updates) the devices will be recalled to be updated and maintained to ensure robustness and improve the model with new data.

Identify and analyze the known and foreseeable hazards associated with each device

The main hazards are misclassification, deskilling, poor generalization and data drift.

Estimate and evaluate the risks associated with, and occurring during, the intended use and during reasonably foreseeable misuse

- From misclassification we may have as risks false negatives and false positives.
- Deskilling may cause overdependence on the software.
- Consequences of poor generalization may be misdiagnosis on underrepresented groups.
- From data drift we may observe a decline in the model's performances.

Eliminate or control the risks referred to in point (c) in accordance with the requirements of Section 4

Given the fact that the software ought to be used by trained professionals, as humans-in-command, in order to be fully effective, the risks stemming from the first two hazards are minimized.
Risks originating from the last two hazards have been preventively addressed in our post-market surveillance plan. Moreover, we remind that our technical documentation already provides the requirements for the correct use of the software, and it is responsibility of the user to ensure that they are satisfied.

Evaluate the impact of information from the production phase and, in particular, from the post-market surveillance system, on hazards and the frequency of occurrence thereof, on estimates of their associated risks, as well as on the overall risk, benefit-risk ratio and risk acceptability

Our post-market surveillance has been crafted in order to account for what is being required: for more details on the impact evaluation, please refer to 6.

Based on the evaluation of the impact of the information referred to in point (e), if necessary amend control measures in line with the requirements of Section 4

As stated above, the problem has been considered in the post-market surveillance plan.

Eliminate or reduce risks as far as possible through safe design and manufacture

The device is programmed and verified via a deep consideration of the data and the possible errors, and explainability is applied to ensure safety. The hardware manufacture responsibility is on the producer.

Where appropriate, take adequate protection measures, including alarms if necessary, in relation to risks that cannot be eliminated

The users have been warned through this technical documentation on the limitations of the software performances.

Provide information for safety (warnings/precautions/contra-indications) and, where appropriate, training to users

The required information is already present in the current technical documentation.

All known and foreseeable risks, and any undesirable side-effects, shall be minimised and be acceptable when weighed against the evaluated benefits to the patient and/or user arising from the achieved performance of the device during normal conditions of use.

The device is not considered a one hundred percent accurate classifier, but should be used to speed up the process of diagnosis for medical professionals. Wrong diagnosis are responsibility of the user. The users have been warned through this technical documentation on the limitations of the software performances.

5 AI Lifecycle

5.1 Scoping

The aim of MATE is to aid clinicians in identifying the cardiac macro pathologic group to which a certain patient belongs to, given its ECG signal, provided directly to the software through a 12 lead ECG monitor with a sample rate of 500 Hz.

5.2 Data

The data, as we reported, is curated by licensed clinicians and consists of 12 lead ECG signals with 500 Hz sampling rate, having as label the heart condition to which they correspond. Data preprocessing includes the normalization of the signals.

5.3 Modelling

The architecture we chose is a 1D Convolutional Neural Network, preferred over other signal-processing architectures due to its lower computational burden and superior speed. The best model was selected among the ones showing better performances.

5.4 Deployment

A deep interplay between physicians and the software will be necessary in order to maintain the effectivity of the device. Professionals using the device may be asked to give a feedback on the model output in order to assess the ongoing usefulness of the predictions and perform a re-training of the model, when a decline in performance is manifest.

6 Technical documentation on post-market surveillance

- The first month of software deployment will be a shadow deployment, where the prediction of the model will be evaluated against the diagnoses of the professionals and the true label of the software coming from the follow-ups of the patients, in order to validate the real-world performance of the model. Its robustness will be assessed through confusion matrices and key metrics, such as weighted average F1-score.
- If the software performance meets pre-defined thresholds, full deployment will be actuated, otherwise retraining will be scheduled.
- Once in service, clinicians will play a fundamental role in assessing the quality of the predictions by reporting, after the first follow-up, the outcomes of the model support. These logs will be reviewed periodically.
- In case of evident drops in weighted average F1-score (<0.7) or reports of increased high-risk misclassifications, retraining will be scheduled. A shadow deployment period will precede the release of the new model.

References

[1] Erick A Perez Alday et al. *Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020*, volume 41 of *Physiol. Meas.* IPEM, 2020.

[2] Samuele Pe, Tommaso Mario Buonocore, Nicora Giovanna, and Parimbelli Enea. SignalGrad-CAM. `https://github.com/bmi-labmedinfo/signal_grad_cam`, 2025.

[3] Zhang J. Danioko S. et al. Zheng, J. *A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients*, volume 7. Sci Data, 2020.