

Journal Pre-proof

Artificial Intelligence for Detection of Ventricular Oversensing Machine Learning Approaches for Noise Detection Within Non-Sustained Ventricular Tachycardia Episodes Remotely Transmitted by Pacemakers and Implantable Cardioverter Defibrillators

Marc Strik, MD PhD, Benjamin Sacristan, MD, Pierre Bordachar, MD PhD, Josselin Duchateau, MD PhD, Romain Eschalier, PhD, Pierre Mondoly, MD PhD, Julien Laborderie, MD, Narimane Gassa, Nejib Zemzemi, PhD, Maxime Laborde, Juan Garrido, Clara Matencio Perabla, Guillermo Jimenez-Perez, PhD, Oscar Camara, PhD, Michel Haïssaguerre, MD, Rémi Dubois, PhD, Sylvain Ploux, MD PhD



PII: S1547-5271(23)02404-9

DOI: <https://doi.org/10.1016/j.hrthm.2023.06.019>

Reference: HRTM 9928

To appear in: *Heart Rhythm*

Received Date: 24 February 2023

Revised Date: 13 June 2023

Accepted Date: 28 June 2023

Please cite this article as: Strik M, Sacristan B, Bordachar P, Duchateau J, Eschalier R, Mondoly P, Laborderie J, Gassa N, Zemzemi N, Laborde M, Garrido J, Perabla CM, Jimenez-Perez G, Camara O, Haïssaguerre M, Dubois R, Ploux S, Artificial Intelligence for Detection of Ventricular Oversensing Machine Learning Approaches for Noise Detection Within Non-Sustained Ventricular Tachycardia Episodes Remotely Transmitted by Pacemakers and Implantable Cardioverter Defibrillators, *Heart Rhythm* (2023), doi: <https://doi.org/10.1016/j.hrthm.2023.06.019>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Artificial Intelligence for Detection of Ventricular Oversensing

Machine Learning Approaches for Noise Detection Within Non-Sustained Ventricular Tachycardia Episodes Remotely Transmitted by Pacemakers and Implantable Cardioverter Defibrillators

Running title: Machine Learning for Detection of Lead Noise

Marc Strik MD PhD^{1,2}, Benjamin Sacristan MD^{1,2}, Pierre Bordachar MD PhD^{1,2}, Josselin Duchateau MD PhD^{1,2}, Romain Eschali r PhD³, Pierre Mondoly MD PhD⁴, Julien Laborderie MD⁵, Narimane Gassa², Nejib Zemzemi PhD², Maxime Laborde², Juan Garrido⁶, Clara Matencio Perabla⁶, Guillermo Jimenez-Perez PhD⁶, Oscar Camara PhD⁶, Michel Ha ssaguerre MD^{1,2}, R  mi Dubois PhD², Sylvain Ploux MD PhD^{1,2}

1. Bordeaux University Hospital (CHU), Cardio-Thoracic Unit, F-33600 Pessac, France
2. IHU Liryc, Electrophysiology and Heart Modeling Institute, Fondation Bordeaux Universit  , Universit   de Bordeaux, F-33600 Pessac- Bordeaux, France
3. Department of Cardiology, University Hospital Clermont-Ferrand, Clermont-Ferrand, France
4. Department of Cardiology, University Hospital Rangueil, Toulouse, France.
5. Department of Cardiology, Bayonne Hospital, Bayonne, France.
6. Universitat Pompeu Fabra, Barcelona, Spain

Word count: 4100 words

Address for correspondence: Dr Strik ; Service Pr Ha ssaguerre, H  pital cardiologique du Haut-L  v  que, Avenue de Magellan, 33600 Pessac, France.

e-mail:marcstrik@gmail.com

Conflict of Interest: None declared

Funding: This work received financial support from the French Government as part of the "Investments of the Future" program managed by the National Research Agency (ANR) [Grant number ANR-10-IAHU-04]

Abstract

Background

Pacemakers (PMs) and implantable cardioverter defibrillators (ICDs) increasingly automatically record and remotely transmit non-sustained ventricular tachycardia (NSVT) episodes which may reveal ventricular oversensing.

Objectives

We aimed to develop and validate a machine learning algorithm which accurately classifies NSVT episodes transmitted by PMs and ICDs in order to lighten healthcare workload burden and improve patient safety.

Methods

PMs or ICDs (Boston Scientific) from four French hospitals with ≥ 1 transmitted NSVT episode were split into three subgroups: training set, validation set, and test set. Each NSVT episode was labelled as either physiological or non-physiological. Four machine learning algorithms (2DTF-CNN, 2D-DenseNet, 2DTF-VGG, and 1D-AgResNet) were developed using a training and validation dataset. Accuracies of the classifiers were compared with an analysis of the remote monitoring team of the Bordeaux University Hospital using F2 scores (favoring sensitivity over predictive positive value) using an independent test set.

Results

807 devices transmitted 10,471 NSVT recordings (82% ICD, 18% PM), of which 87 devices (10.8%) transmitted 544 NSVT recordings with non-physiological signals. The classification by the remote monitoring team resulted in an F2 score of 0,932 (sensitivity of 95%, specificity of 99%) The four machine learning algorithms showed high and comparable F2 scores (2DTF-CNN: 0,914, 2D-DenseNet: 0,906, 2DTF-VGG: 0,863, 1D-AgResNet: 0,791) and only 1D-AgResNet had significantly different labeling as compared with the remote monitoring team.

Conclusion

Machine learning algorithms were accurate in detecting non-physiological signals within EGMs transmitted by pacemaker and ICDs. An artificial intelligence approach may render remote monitoring less resourceful and improve patient safety.

Key words:

Pacemaker, implantable cardioverter defibrillator, remote monitoring, machine learning, artificial intelligence, lead noise

Journal Pre-proof

Introduction

Remote monitoring of pacemakers (PM) and implantable cardioverter defibrillators (ICD) has been shown to improve patient outcomes by early detection of technical issues and enjoys a class I recommendation.^{1,2} Remote monitoring in its current form requires significant human resources trained to extract critical information from the transmitted data.^{3,4} This approach is associated with rising costs and major organizational challenges. As an increasing amount of implanted PMs and ICDs transmit increasing amounts of mostly abundant data, remote monitoring may become the victim of its own success, possibly becoming non-sustainable in the near future. A striking illustration of this dilemma is the analysis of non-sustained ventricular tachycardia (NSVT) electrograms (EGM). The occurrence of successive fast ventricular sensed events triggers classification and recording of NSVT EGMS which are automatically sent to the proprietary RM platform. Contemporary PMs and ICDs RM systems are design to allow RM notification for each of these NSVT episodes which may be numerous and often of low clinical importance. Careful analysis remains however critical since these episodes occasionally reveal ventricular oversensing which may be related to lead failure, physiological oversensing or electromagnetic interference.⁵ The early identification of ventricular oversensing is crucial for the prevention of clinical complications such as delivery of inappropriate shocks and syncope due to inappropriate inhibition of pacing.^{6,7} Artificial intelligence is increasingly used in cardiac signal analysis and may also be able to automatically classify transmitted NSVT episodes through the use of machine learning (ML). A ML algorithm could alert healthcare professionals in case of suspicion of ventricular oversensing. When implemented in remote monitoring, this new approach may lighten the burden of healthcare professionals but also improve patient safety by prioritizing episodes showing ventricular oversensing. Artificial intelligence may therefore aid and speed-up clinical decision-making, avoiding inappropriate therapies and symptomatic lead dysfunction. Our objective was to develop and validate a machine learning algorithm which classifies NSVT episodes remotely transmitted by PMs and ICDs.

Methods

Non-sustained ventricular arrhythmia episodes

A wireless transmitter device is first connected to the PM/ICD and then given to the patient to be plugged to a power socket in the vicinity of where he or she sleeps. In case of registration of an EGM due to suspicion of NSVT, the associated episode is transmitted the following night to the proprietary remote monitoring platform as an alert. All PM/ICDs (Boston Scientific) from four French Hospitals (Bordeaux University Hospital, Clermont Ferrand University Hospital, Toulouse University Hospital, Bayonne Hospital), which transmitted at least one NSVT episode to the Latitude remote monitoring

platform were included. No devices were excluded. All NSVT episodes were extracted from the platform and associated EGMs and markers were saved into a single datastructure. Each episode was labelled as either *physiological* (all ventricular markers related to ventricular depolarization) or *non-physiological* (presence of ventricular markers caused by anything other than ventricular depolarization) by an electrophysiologist. Discrimination between ventricular or supraventricular arrhythmia was not performed; both were considered as *physiological*. The patients (using device serial numbers) were split into three subgroups: training set, validation set, and test set. The training dataset was used to train the model. The validation dataset was used to provide an evaluation of a model fit on the training dataset while tuning model hyperparameters. The test set comprised the sample of data used to provide an unbiased evaluation of the final model. These three data sets were randomly generated from the complete data set. Finally, in order to compare the machine learning algorithms with healthcare professionals, the test set was also analyzed by the remote monitoring nurses of the Bordeaux University Hospital (n=5, each analyzed 20%). The research reported in this paper adhered to the Helsinki Declaration, data were de-identified prior to analysis and the ethics review board of Bordeaux University approved the study under reference CER-BDX 2023 – 46.

Machine Learning algorithms

Four deep learning (DL) classifiers were tested in order to test three key aspects of the classifier (**Figure 1**): 1) the input space for handling 1D signals; 2) the imbalanced data-set with missing signals; and 3) the deep network architecture itself. The EGMs were not of equal length and did not always have the same number of channels. A first classifier was designed to input the 1D raw signals as a matrix composed of constant duration, and 3 rows corresponding to the 3 channels right ventricle (RV), right atrial (RA), and Far Field (coil-can). When a channel was not available in a given record, the corresponding row was replaced with zeros. The architecture of the DL classifier was based on the augmented residual network architecture.⁸ This network is subsequently referred to as the 1D-AgResNet. Since deep learning architectures with convolutional layers have shown superiority over other algorithms when dealing with 2D images, a second DL classifier uses as input 2D images representing the trace from the RV channel. The DL architecture used to handle these images is a Dense Net architecture.⁹ This network is referred to as 2D-DenseNet. Fourier transform is known to efficiently analyze noise in temporal recordings. Hence, time/frequency map were computed using a sliding Fourier transform of the RV channel of each recording over a fixed duration window and used as inputs for two last DL classifiers. Then two architectures were designed: a standard Convolutional Neural Network, referred to as 2DTF-CNN and a VGG architecture referred to as 2DTF-Vgg.¹⁰ For each of the 4 proposed classifiers, the hyperparameters (including the optimization algorithm, the class weights to deal with unbalanced classes, and the final number of parameters (i.e., the number of layers

for a given architecture) were optimized using a 5-fold cross-validation scheme on the training data set. More details on the ML algorithms are found in the Supplemental Material. Once these parameters were set, several training corresponding to several random initializations of weights were performed on the entire training set for each of the 4 DL classifiers. The networks showing the best results on the validation set were selected as the final networks for each DL classifier. The performances of the DL classifiers were compared on the training set.

Statistical Analysis

Sensitivity, specificity, positive predictive value and negative predictive value for detecting “non-physiological” recordings were calculated for each algorithm and the remote monitoring team on the basis of the “test” data set. An F2 score (the weighted harmonic mean of the precision (positive predictive value) and recall (sensitivity)) was computed for test comparison. The F2 score gives more weight to recall than to precision which means a higher score for a lower rate of false negative (misclassified “non-physiological” recordings). As each machine learning algorithm and the remote monitoring team labeled the exact same episodes, the McNemar test was performed.¹¹ The McNemar test evaluates whether the change in labeling proportions in the same set of episodes across two classifiers are significant. A p-value <0.05 was considered statistically significant. In order to compare the remote monitoring team capability of detecting non-physiological signals with 95% sensitivity and 99% specificity to a novel machine learning algorithm using a non-inferiority approach, with a non-inferiority margin of 5%, and the “disease” prevalence is 5%, the required sample size of the test dataset should be at least 1112 episodes.

Data Availability

The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Results

A total of 807 devices transmitted 10,471 NSVT recordings. 82% of the recordings were transmitted from ICDs, 18% from pacemakers. The serial numbers were split into three datasets comprising the training set with 417 devices and 4,998 recordings, the validation set with 142 devices and 2,621 recordings and finally the test set with 328 devices and 2,843 recordings. A flow chart of the devices and their recordings are shown in **Figure 2**. 87 devices (10.7%) transmitted 544 (5.2%) non-physiological recordings with a mean number of non-physiological recordings of 6.17 (range 1 - 86).

The final diagnosis for these recordings were: external noise oversensing (379, 69.7%), lead related oversensing (69, 12.7%), diaphragm myopotential oversensing (89, 16.4%) and P-wave oversensing (7, 1.2%). No T-wave oversensing recording was found in the dataset. Examples of physiological and non-physiological recordings are shown in **Figure 3**.

The classification by the remote monitoring team resulted in a sensitivity of 95%, a specificity of 99.3%, and an F2 score of 0,932. The four machine learning algorithms (2DTF-CNN, 2D-DenseNet, 2DTF- VGG, and 1D-AgResNet) showed similar performance as shown by the comparable F2-Scores on the test database for each algorithm in **Figure 4**. The CNN-based network (2DTF-CNN), using time-frequency maps of the ventricular bipolar signal as an input resulted in the highest F2-Score of 0,914. The other algorithms returned a F2-Score of 0,906 for 2D-DenseNet (based on the analysis of 2D images of the ventricular signal), 0,863 for 2DTF-VGG (based on time/frequency maps, pre-trained) and 0,791 for 1D-AgResNet (based on the analysis of 1D signal).

The results of each approach (remote monitoring team and machine learning approaches) in terms of Sensitivity, Specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) are shown in Figure 5. The remote monitoring team had a sensitivity of 95% (90%-98% 95% CI) and a specificity of 99% (98.96%-99.61%, 95% CI). When compared to the remote monitoring team, only for 1D-AgResNet was the agreement in labeling significantly different ($p < 0.001$, based mainly on the higher amount of false negatives). The contingency tables for each architecture across the three datasets and labeling agreements as compared with the remote monitoring team are shown in the supplemental data.

Discussion

Machine learning can be used to accurately discriminate between physiological and non-physiological signals in NSVT episodes transmitted by PMs/ICDs through remote monitoring. These automatic algorithms show diagnostic performance comparable with dedicated remote monitoring healthcare professionals.

The growing role of artificial intelligence in ECG interpretation

Aiming to facilitate health care decisions and reduce costs, computerized interpretation of the electrocardiogram (ECG) exists since several decades. Limited accuracy of these algorithms based on signal processing require most ECGs to be over read by healthcare professionals. In parallel, there is an exponential use of wearable ECG acquisition technologies, in particular for detection of AF.¹²⁻¹⁴ This adds to the ever increasing amount of ECG data which needs to be interpreted. While most devices

are capable of automatic recognition of AF, repeat analysis by a healthcare professional with expertise in ECG rhythm interpretation is still necessary to establish a definitive diagnosis. Accompanied by a growing need in automated ECG interpretation, the arrival of artificial intelligence is on its way to revolutionize the field of automatic interpretation of ECGs.^{15,16} The ability of artificial intelligence and particularly ML to recognize patterns and learn useful features from recorded signals renders it particularly well suited to interpret ECGs and has already been proven to be efficient in diagnosing several ECG conditions including recognition of noise artifacts.¹⁷

Artificial intelligence in the context of CIED remote monitoring

Remote monitoring of ICDs and but also pacemakers allows for earlier detection of clinical and technical complications and reduces the risk of all cause hospitalization and death.⁴ While these benefits are significant, so are the associated costs. New generation PMs/ICDs automatically transmit large amounts of data and comprehensive interpretation requires a significant amount of human resource and expertise.¹⁸ Despite embedded algorithms designed to alert for specific conditions, the transmitted information has poor positive predictive value, meaning many false positives. At the remote monitoring center of the Bordeaux University Hospital, we follow over 7000 implanted patients which send over a hundred thousand transmissions each year. Less than 4% of these transmissions requires examination by a cardiologist, mean that 96% is filtered by the five remote monitoring healthcare professionals. The conception of our study was therefore driven by a clinical need: to decrease the work burden of the remote monitoring staff. All transmitted NSVT episodes are currently analyzed by healthcare professionals for the purpose of detecting lead oversensing (in ICDs in particular), often manifested by short runs of spurious RR intervals. Although rare, these events are important to identify as they are the cornerstone of inappropriate therapy prevention.¹⁹ NSVT may reveal lead conductor fracture, lead insulation defect, EMI, oversensing of P- R- or T-waves. Per definition, these episodes terminate before binning of the VT/VF counter but may be warning signs of imminent inappropriate shocks. In pacemakers, these episodes often show inappropriate inhibition and may be a warning sign of symptomatic lead dysfunction with possible catastrophic outcome. Two (out of five) device manufacturers do not provide dedicated NSVT alerts on the premise not to overwhelm the healthcare professionals with mostly non-significant events. We hope that ML algorithms as proposed in our study will prioritize transmission of NSVTs in order to improve RM benefits and patient safety.

Different ML approaches

Four deep network models were tested in this study, with various architectures and input features. In the end, whatever the model, the results are very close, suggesting that deep neural networks are well suited for this classification task. Time frequency maps were first computed from the raw EGMs for the 2DTF-CNN and 2DTF-VGG networks. These maps seemed to be particularly suitable for a noise classification problem. However, this preprocessing step brings little benefit in terms of results compared to a raw processing of the 2D images of the plots. The 1D-agResNet was the only one to use the information of all the available channels. However, this study deals with various types of devices, thus some channels are missing in some configurations (e.g. Coil-Can channel for pacemakers). This network is the one with the lowest results, indicating that the information gained from the additional channels when available is counterbalanced with the complexity introduced by the management of the missing data when the RV is the unique channel. Probably, a training on ICDs only, with the coil-can channel always present, would give better results.

Future applications/improvements

A major goal of remote monitoring is to decrease inappropriate therapies from sustained oversensing. From clinical practice and literature, we know that they are most often preceded by non-sustained events.⁵ From clinical practice and literature, we know that they are most often preceded by non-sustained events. In a multicenter evaluation of the SecureSense algorithm (Abbott) we found that 95% of the lead failure were detected by non-sustained oversensing events versus 5% for sustained events of oversensing. Our AI approach aims to identify at-risk patients by targeting nonsustained events as those are under analyzed today because of the associated work burden on remote monitoring staff and low yield. Today and in the near future, sustained events are less a target for AI approaches as they are prioritized by the devices and platforms and analyzed by professionals to ensure proper device functioning. This approach could also be applied to other types of transmitted EGMs such as other type of arrhythmia episodes or real-time EGMs. Machine learning could not only be used to detect lead noise but also to improve discrimination between supraventricular and ventricular arrhythmia or other decisions now made by algorithms based on limited rules.²⁰⁻²⁴ The ultimate goal is to entrust the analysis of transmitted EGMs to ML-algorithms in order to maximize the number of patients followed per remote monitoring center.^{20,24} This would mean that more patients would profit from the benefit of remote monitoring, with a direct effect on the health care system. Alternatively, such algorithms may be used to help healthcare professionals to make a decision. In contemporary devices, the available room for additional algorithms is extremely limited. We do not expect for AI algorithms to become embedded within devices anytime soon. We do expect them to be rapidly integrated within remote monitoring platforms. With the availability of a continuously

improving AI classifier after transmission of each episode, existing onboard algorithms may be changed to work at increased sensitivity.

Limitations

The non-physiological to physiological episode ratio is highly unbalanced which added to the challenge of the ML algorithms. However, the accuracy of all algorithms was highly satisfactory. The ML algorithms were developed with Boston Scientific tracings of NSVT and may not be directly applicable for the analysis of tracings from a different manufacturer. Because of heterogeneity in the clinical and demographic data, confounding factors may have been missed, such as the higher pre-test probability that ICD wearers transmit true NVST due to underlying cardiac disease.

Conclusion

Machine learning algorithms were accurate in detecting non-physiological signals within EGMs transmitted by pacemakers and ICDs. An artificial intelligence approach may render remote monitoring more effective and improve patient safety.

References

1. Wilkoff BL, Fauchier L, Stiles MK, et al.: 2015 HRS/EHRA/APHRS/SOLAECE expert consensus statement on optimal implantable cardioverter-defibrillator programming and testing. *Heart Rhythm Elsevier*, 2016; 13:e50–e86.
2. Glikson M, Nielsen JC, Kronborg MB, et al.: 2021 ESC Guidelines on cardiac pacing and cardiac resynchronization therapy: Developed by the Task Force on cardiac pacing and cardiac resynchronization therapy of the European Society of Cardiology (ESC) With the special contribution of the European Heart Rhythm Association (EHRA). *Eur Heart J* [Internet] 2021 [cited 2021 Sep 10]; . Available from: <https://doi.org/10.1093/eurheartj/ehab364>
3. C K, L F, S A, et al.: Remote monitoring for cardiac implantable electronic devices: A practical guide. *Arch Cardiovasc Dis* [Internet] *Arch Cardiovasc Dis*, 2022 [cited 2022 Nov 3]; 115. Available from: <https://pubmed-ncbi-nlm-nih-gov.proxy.insermbiblio.inist.fr/35688787/>
4. Ploux S, Strik M, Varma N, Eschalier R, Bordachar P: Remote monitoring of pacemakers. *Arch Cardiovasc Dis* 2021; 114:588–597.
5. Welte N, Strik M, Eschalier R, et al.: Multicenter investigation of an implantable cardioverter-defibrillator algorithm to detect oversensing. *Heart Rhythm* 2017; 14:1008–1015.
6. Ploux S, Varma N, Strik M, Lazarus A, Bordachar P: Optimizing Implantable Cardioverter-Defibrillator Remote Monitoring: A Practical Guide. *JACC Clin Electrophysiol* 2017; 3:315–328.
7. Ploux S, Swerdlow CD, Strik M, et al.: Towards eradication of inappropriate therapies for ICD lead failure by combining comprehensive remote monitoring and lead noise alerts. *J Cardiovasc Electrophysiol* 2018; 29:1125–1134.
8. He K, Zhang X, Ren S, Sun J: Deep Residual Learning for Image Recognition. 2016 IEEE Conf Comput Vis Pattern Recognit CVPR 2016, pp. 770–778.
9. Huang G, Liu Z, van der Maaten L, Weinberger KQ: Densely Connected Convolutional Networks [Internet]. *arXiv*, 2018 [cited 2022 Nov 3];. Available from: <http://arxiv.org/abs/1608.06993>
10. Simonyan K, Zisserman A: Very Deep Convolutional Networks for Large-Scale Image Recognition [Internet]. *arXiv*, 2015 [cited 2022 Nov 3];. Available from: <http://arxiv.org/abs/1409.1556>
11. McNemar Q: Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 1947; 12:153–157.
12. Abu-Alrub S, Strik M, Ramirez FD, et al.: Smartwatch Electrocardiograms for Automated and Manual Diagnosis of Atrial Fibrillation: A Comparative Analysis of Three Models. *Front Cardiovasc Med* 2022; 9:836375.
13. Strik M, Ploux S, Ramirez FD, et al.: Smartwatch-based detection of cardiac arrhythmias Beyond the differentiation between sinus rhythm and atrial fibrillation. *Heart Rhythm* 2021; :S1547-5271(21)01731-8.
14. Racine H-P, Strik M, van der Zande J, et al.: Role of Coexisting ECG Anomalies in the Accuracy of Smartwatch ECG Detection of Atrial Fibrillation. *Can J Cardiol* 2022; 38:1709–1712.

15. Hannun AY, Rajpurkar P, Haghpanahi M, et al.: Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 2019; 25:65–69.
16. Mittal S, Oliveros S, Li J, Barroyer T, Henry C, Gardella C: AI Filter Improves Positive Predictive Value of Atrial Fibrillation Detection by an Implantable Loop Recorder. *JACC Clin Electrophysiol* 2021; 7:965–975.
17. Herraiz ÁH, Martínez-Rodrigo A, Bertomeu-González V, Quesada A, Rieta JJ, Alcaraz R: A Deep Learning Approach for Featureless Robust Quality Assessment of Intermittent Atrial Fibrillation Recordings from Portable and Wearable Devices. *Entropy Basel Switz* 2020; 22:733.
18. Maines M, Tomasi G, Moggio P, et al.: Implementation of remote follow-up of cardiac implantable electronic devices in clinical practice: organizational implications and resource consumption. *J Cardiovasc Med Hagerstown Md* 2020; 21:648–653.
19. Swerdlow CD, Ploux S, Poole JE, Nair SG, Himes A, Ellenbogen KA: Interpreting device diagnostics for lead failure. *Heart Rhythm* 2022; 19:154–164.
20. Strik M, Defaye P, Eschaliér R, et al.: Performance of a specific algorithm to minimize right ventricular pacing: A multicenter study. *Heart Rhythm* 2016; 13:1266–1273.
21. Frontera A, Strik M, Eschaliér R, et al.: Electrogram morphology discriminators in implantable cardioverter defibrillators: A comparative evaluation. *J Cardiovasc Electrophysiol* 2020; 31:1493–1506.
22. Strik M, Frontera A, Eschaliér R, et al.: Accuracy of the pacemaker-mediated tachycardia algorithm in Boston Scientific devices. *J Electrocardiol* 2016; 49:522–529.
23. Jankelson L, Bordachar P, Strik M, Ploux S, Chinitz L: Reducing right ventricular pacing burden: algorithms, benefits, and risks. *Eur Eur Pacing Arrhythm Card Electrophysiol J Work Groups Card Pacing Arrhythm Card Cell Electrophysiol Eur Soc Cardiol* 2019; 21:539–547.
24. Strik M, Socié P, Ploux S, Bordachar P: Unexpected and undesired side-effects of pacing algorithms during exercise. *J Electrocardiol* 2018; 51:1023–1028.

Authors' contributions

Marc Strik, writing - review & editing

Benjamin Sacristan, writing - original draft

Pierre Bordachar, supervision

Josselin Duchateau, data acquisition

Romain Eschalier, data acquisition

Pierre Mondoly, data acquisition

Julien Laborderie, data acquisition

Narimane Gassa, data analysis

Nejib Zemzemi, data analysis

Maxime Laborde, data analysis

Juan Garrido, data analysis

Clara Matencio Perabla, data analysis

Guillermo Jimenez-Perez, data analysis

Oscar Camara, data analysis

Rémi Dubois, data analysis, review & editing

Sylvain Ploux, supervision, review & editing

role of funding source

This work received financial support from the French Government as part of the “Investments of the Future” program managed by the National Research Agency (ANR) [Grant number ANR-10-IAHU-04] which was used to pay part of the salaries.

Conflict of interest

The authors report no conflicts of interest.

Figures

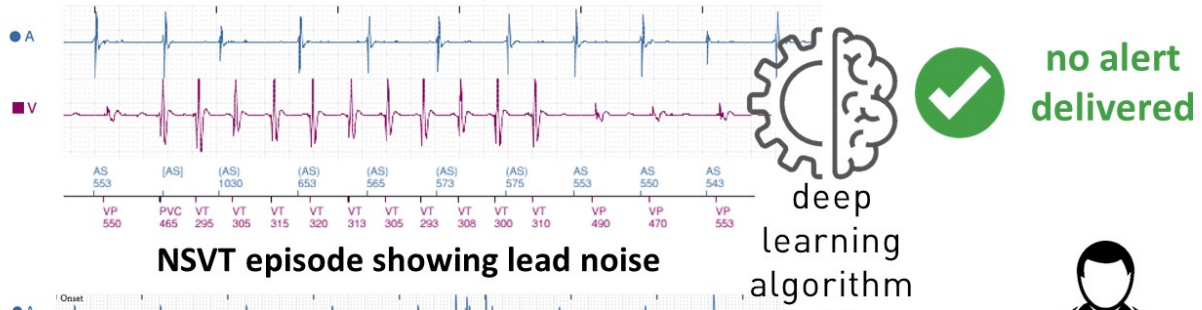
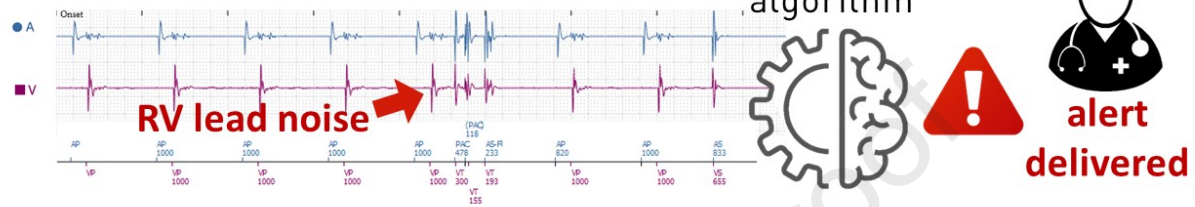
Figure 1. An overview of the four deep learning classifiers with 1 to 3 EGM channels as input and diagnosis Physiological or Non-Physiological as output.

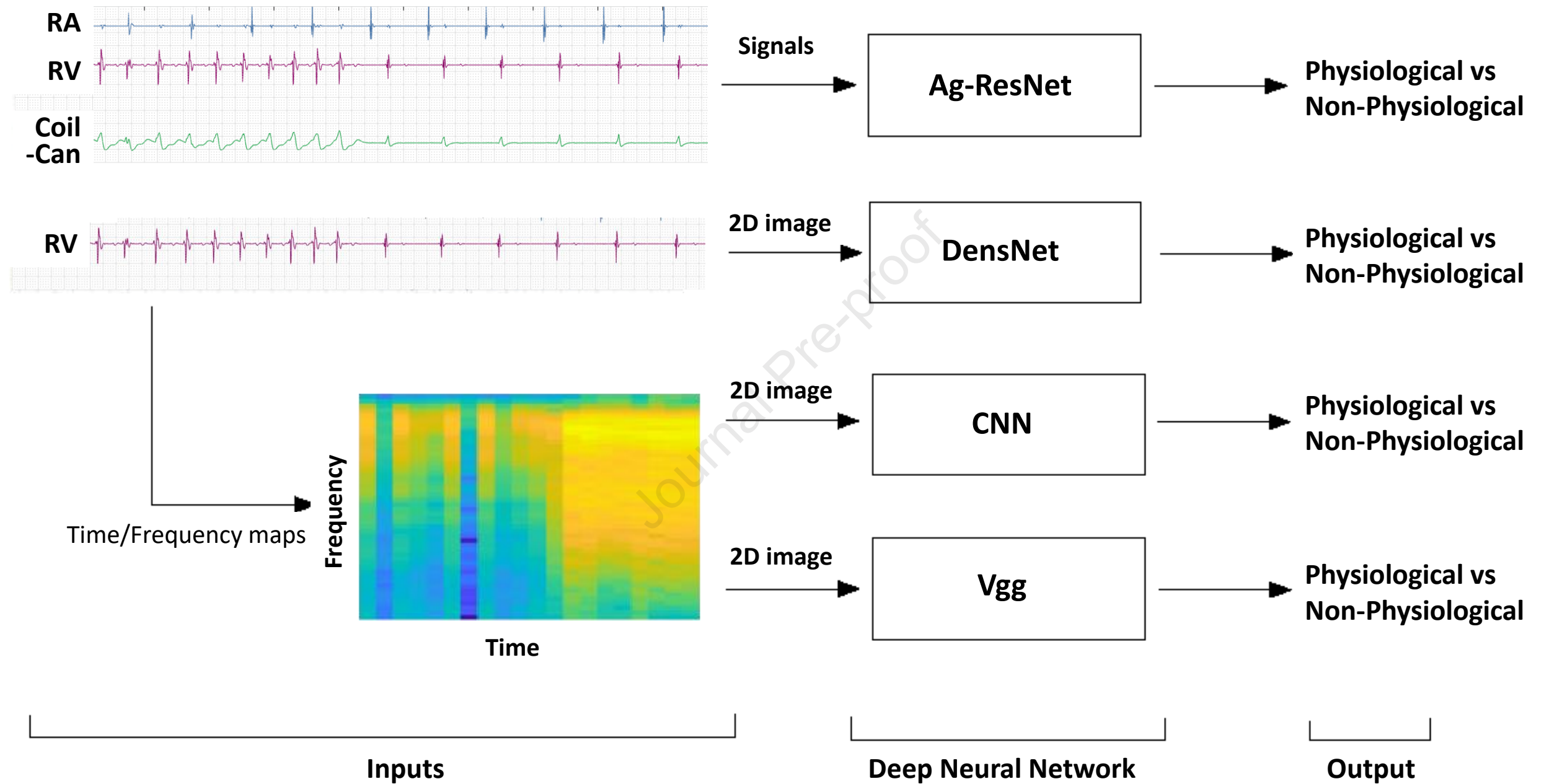
Figure 2. Flowchart of the patients and NSVT episodes included into the study.

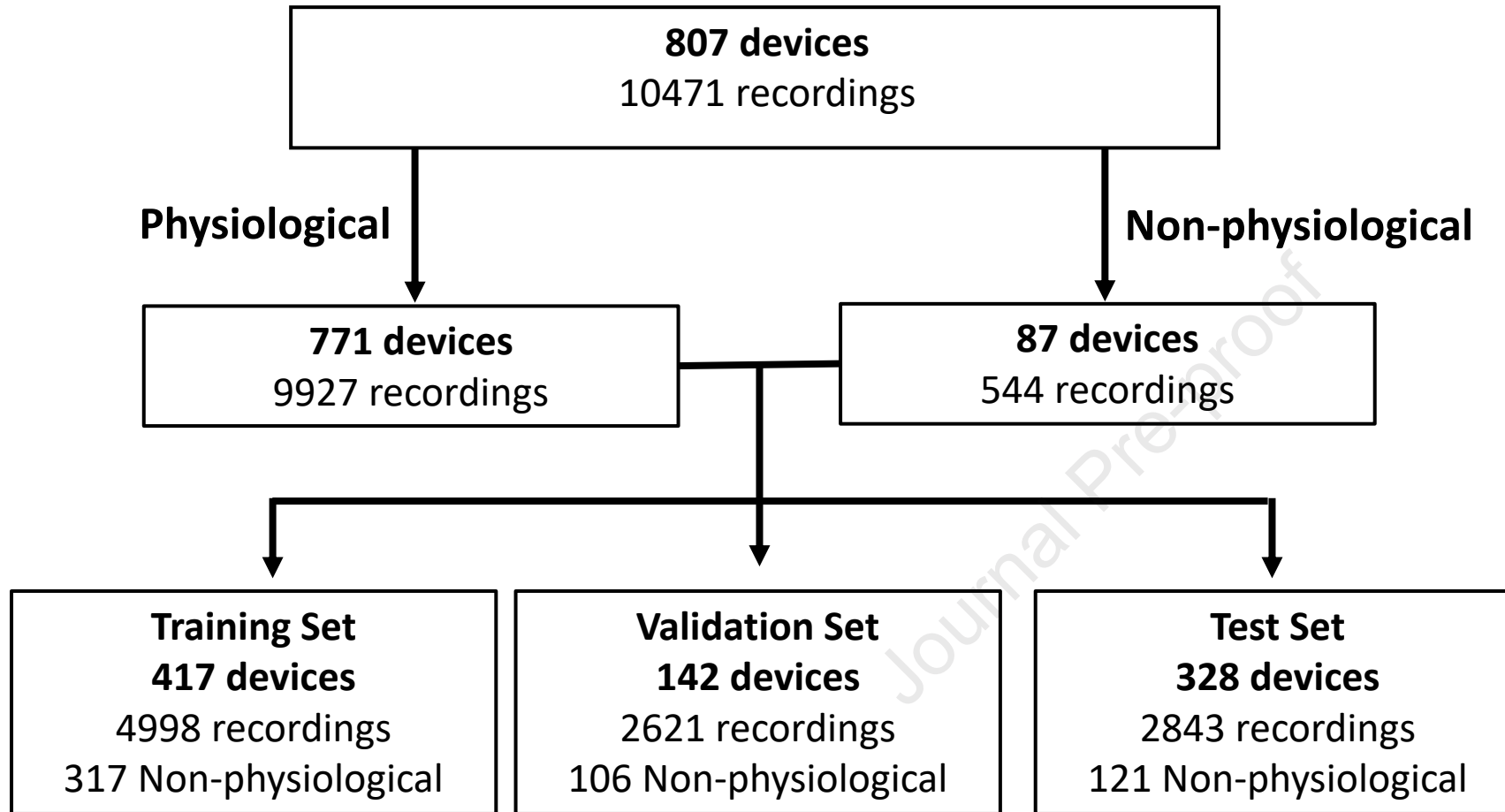
Figure 3. Examples of physiologic and non-physiologic recordings.

Figure 4 – F2-Scores for each algorithm and for the remote monitoring team (test set).

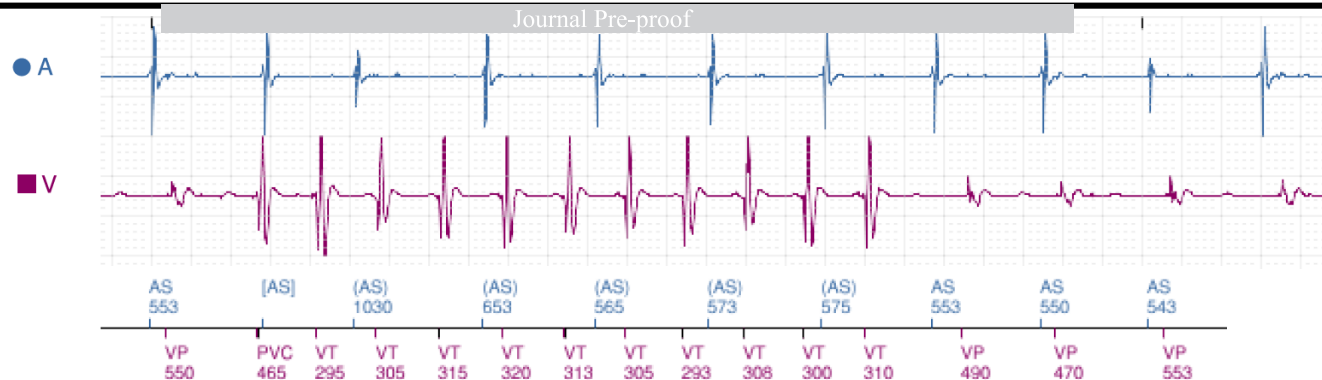
Figure 5 – Diagnostic performance of each machine learning algorithm and remote monitoring team

NSVT episode showing arrhythmia**NSVT episode showing lead noise**

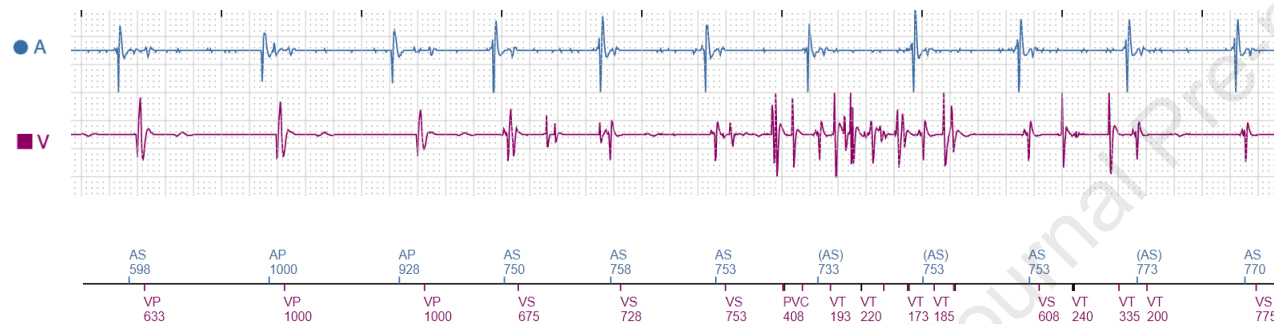


Non Sustained Ventricular Tachycardia

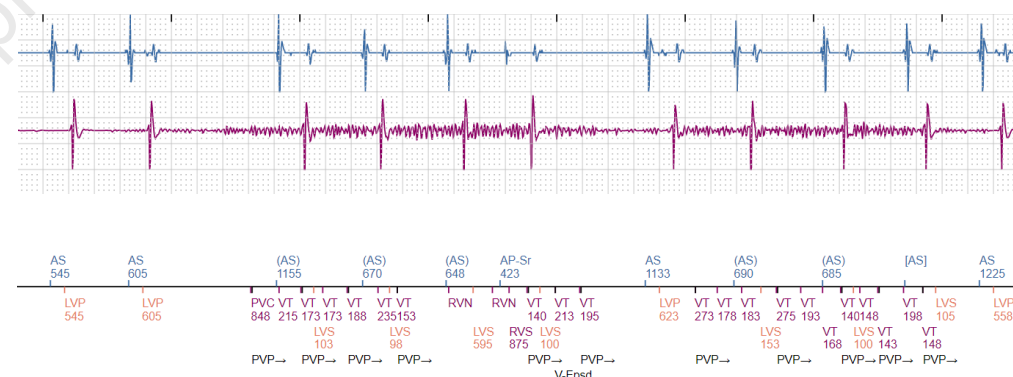
N = 9927



lead related oversensing N = 69



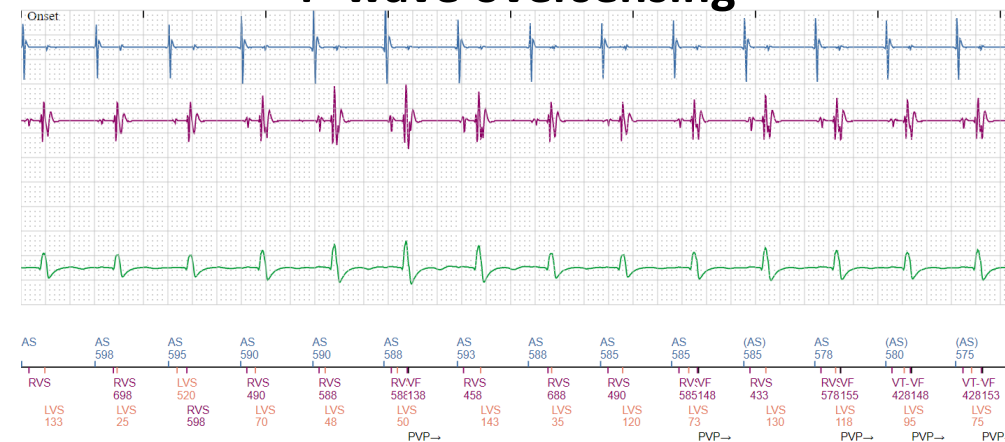
diaphragmatic oversensing N = 89

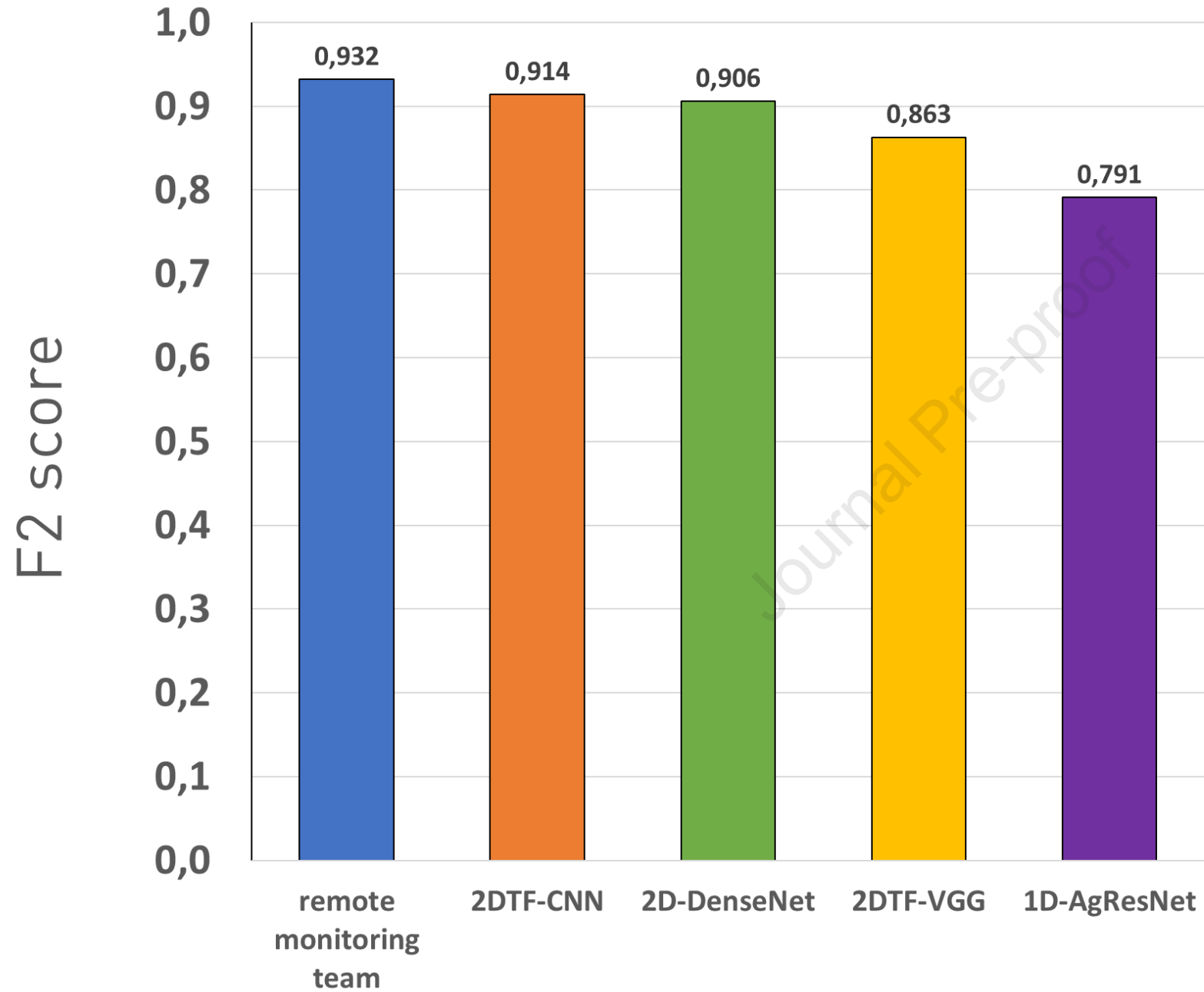


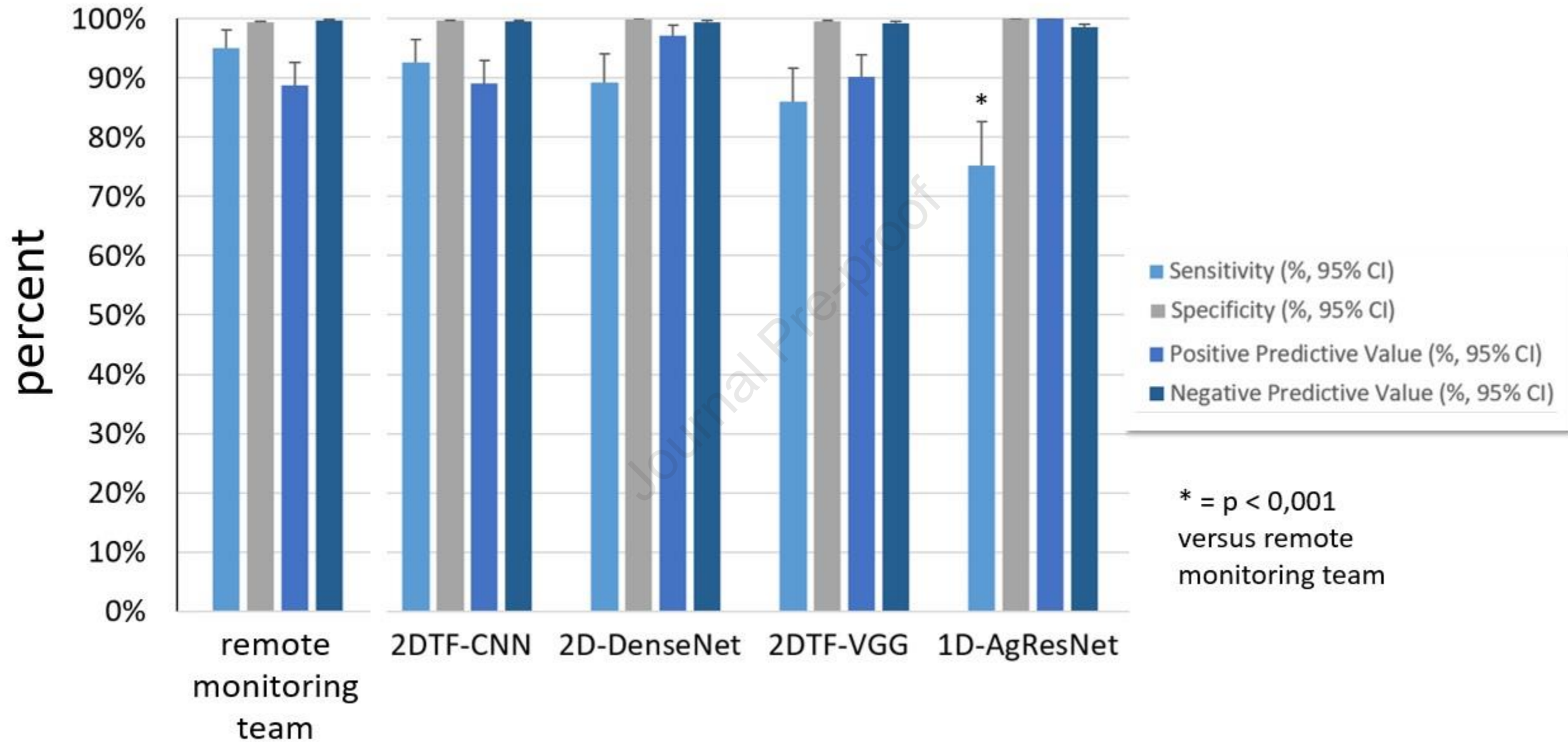
external source oversensing N = 379



P-wave oversensing N = 7







Supplemental data – Neural Network Architectures

1. 1D-AgResNet

All Recordings were unevenly sampled at the maximum rate 200 Hz for efficient storage in the device. We applied data interpolation so that all timesteps were 5 ms. Considering that the recordings were not of the same length, window sliding was applied. It consisted of taking the first W points from the recording of L points, and then moving forward by S points and taking the following W points, at the end getting $(L-W)/S + 1$ samples of each recording. This way we achieved two things: (i) samples of equal length; and (ii) a larger sample size. Moreover, not all recordings have the same number of channels. As not all recordings had the same number of channels, we artificially added signal lines at zero to have 3 channels in all recordings. We finally obtained samples with the same number of channels, time-step, and number of points.

The architecture of our model was based on the Residual neural network with 3 channels in the input and 2000 points per sample.

2. 2D-DenseNet

To take advantage of Deep Convolutional Neural Networks, we converted the time series data into images.

In this part, we only considered the recordings of the right ventricular channel, so we obtained two-dimensional images of the temporal traces of the ventricular signal over 2000ms. Then these images were fed to a Dense neural Network (DenseNet) for image classification. The input image dimension of the network was (248,372,1).

DenseNet is an improved version of convolutional neural networks that minimizes the vanishing gradient problem and can be trained with a few parameters which take less time and is found to achieve better results. [5]

The figure depicts the architecture of a typical DenseNet model, we have dense blocks composed of interconnected layers: each layer is receiving information from all preceding layers and in between, there are non-linear transformations (Batch Normalization, Rectified Linear Unit (ReLU), Convolutions). A transitional block is introduced between two dense blocks as a way of downsampling by applying a batch normalization, a convolution and a global average pooling.

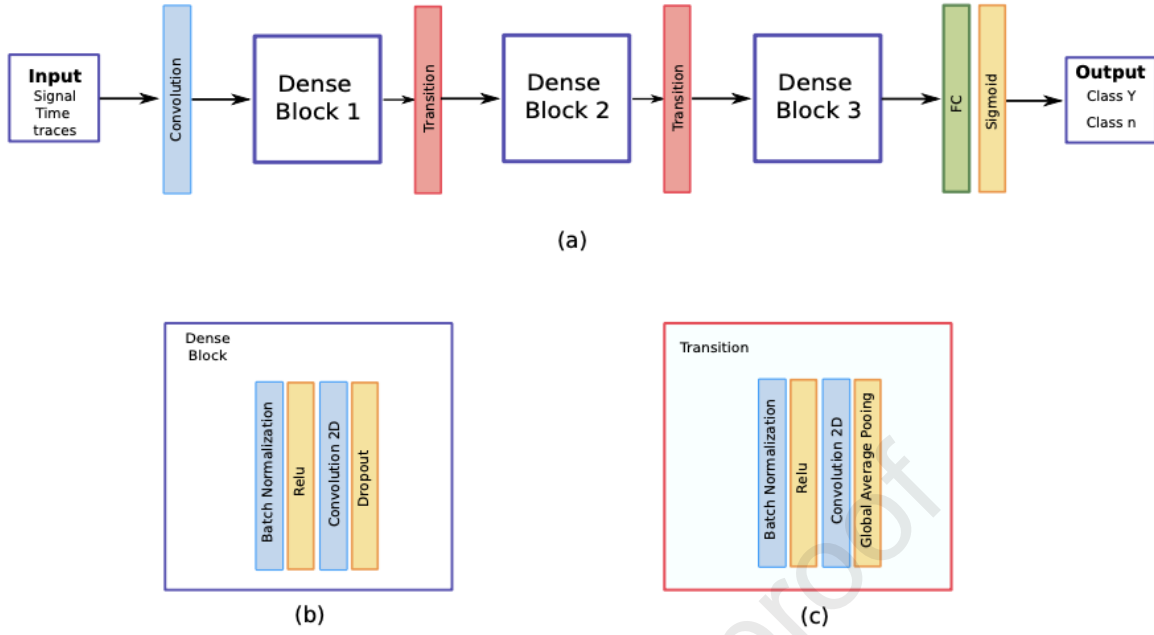


Figure 1 - (a) Representation of the neural network architecture (2D-DenseNet), (b) Dense block, (c) transition layer.

3. 2DTF-CNN

All recordings were sampled at 200Hz. In this section, we have only considered the first 15000ms for each sample and we carried out an interpolation in order to have a fixed time-step equal to 5ms and a number of recordings equal to 3000.

A spectrogram is the pointwise magnitude of the Fourier transform of a segment of an audio signal. We compute, for each time signal, the corresponding spectrogram. We obtain two-dimensional graphs and a third dimension represented by colors; time is running along the x – axis. The vertical axis represents frequency. The amplitude of a particular frequency at a particular time is represented by the third dimension.

The 2D time/frequency maps of the ventricular bipolar signal are then fed to convolutional neural for classification purpose.

We propose here a classical CNN architecture constituted of stacking of two convolutional layers alternating with two pooling layers and terminating with two fully connected (FC) layers and a dropout layer ($p = 0.5$) in-between. The output is a probability distribution obtained utilizing a sigmoid function on the output of the final FC layer. The network architecture is represented in Figure 2.

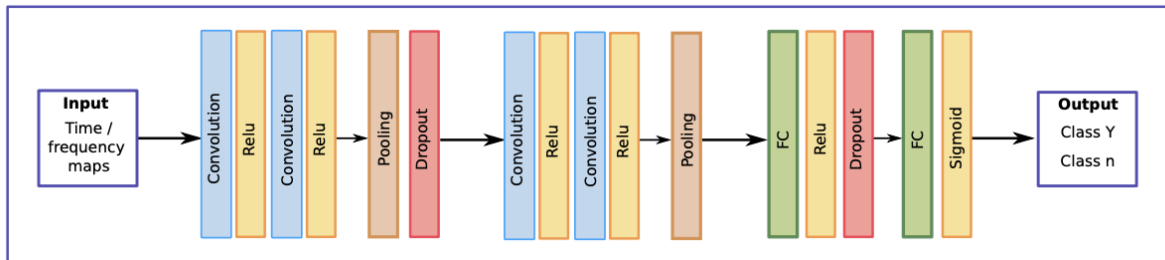


Figure 2 - Representation of the neural network architecture (2DTF-CNN).

We used a stochastic gradient descent method (Adam) for training the proposed model with 32 samples for batch size. The hyper-parameters are set as follows: learning rate of 10^{-3} , and number of epochs set to 100.

4. 2DTF-VGG

In this part, RV-time/frequency maps are also used as inputs of the network. Each RV recording is first divided into regular time segments each 10 seconds wide. We then compute for each part the spectrogram in the range [0-100Hz] with a frequency resolution of 2Hz, and a 50% overlapping sliding windows of 1s leading to one 50x20 image for each 10-second segment. These 2D representations feed a VGG-like neural network. The final architecture was selected using a 5 fold cross-validation process over the training dataset.

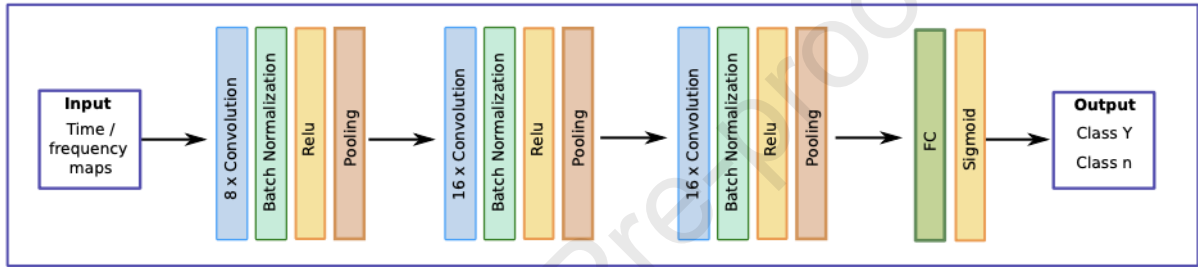


Figure 3 - Representation of the neural network architecture (2DTF-VGG)

In order to manage the unbalanced data set, we train a cost-sensitive classifier and penalize misclassifications from the minority class y more than the majority class n by introducing weights to our loss function as follows:

$$Loss = \frac{1}{n} \sum_{i=1}^n W y_i \log(\hat{y}_i) + (1 - W)(1 - y_i) \log(1 - \hat{y}_i)$$

Where y_i is the probability of input i to be a noise event; and n the total number of examples in the training database. W was set to 0.66 using the cross-validation process. A record being composed of several parts, we classify the recording as noise if at least one of its parts is classified as noise.

Supplemental tables

Table 1 - Contingency tables for each deep learning architecture

	Train	Valid	Test
TN	4677	2508	2705
FN	5	11	9
FP	4	7	17
TP	312	95	112

(a) 2DTF-CNN

	Train	Valid	Test
TN	4680	2513	2718
FN	1	3	13
FP	1	2	4
TP	316	103	108

(b) 2D-DenseNet

	Train	Valid	Test
TN	4679	2502	2708
FN	0	8	17
FP	2	13	14
TP	317	98	104

(c) 2DTF-VGG

	Train	Valid	Test
TN	4681	2511	2722
FN	0	10	30
FP	0	4	0
TP	317	96	91

(d) 1D-AgResNet

Table 2 - Diagnostic performance for each deep-learning algorithm

	2DTF-CNN	2D-DenseNet	2DTF-VGG	1D-AgResNet
Sensitivity (% , 95% CI)	92,6% (86,35%- 96,54%)	89,3% (82,33-94.15%)	85.95% (78.46%- 91.60%)	75.21% (66.54%- 82.60%)
Specificity (% , 95% CI)	99,7% (99.00- 99,64%)	99,9% (99.62%-99.96%)	99.49% (99.14%-99.72%)	100% (99.86%-100.00%)
Positive Predictive Value (% , 95% CI)	89,05% (83,46%- 92,90%)	97.09% (92.59%-98.89%)	90.16% (84.40%-93.95%)	100%
Negative Predictive Value (% , 95% CI)	99,59% (99,24- 99,78%)	99.41% (99.02%-99.65%)	99.23% (98.81%- 99.50%)	98.66% (98.18%-99.01%)

Table 3 - The McNemar test framework for each deep-learning algorithm as compared with the remote monitoring team interpretation and the corresponding p -values

		RM team		
		Physiol	Non-Physiol	
2DTF-CNN	Physiol	2690	16	$p = 0.61$
	Non-Physiol	20	109	
2D-DenseNet	Physiol	2706	19	$p = 0.09$
	Non-Physiol	9	105	
2DTF-VGG	Physiol	2693	26	$p = 0.22$
	Non-Physiol	17	99	
1D-AgResNet	Physiol	2708	36	$p < 0.0001$
	Non-Physiol	2	89	