Deep Learning System for Left Ventricular Assist Device Candidate Assessment from Electrocardiograms

Antonio Mendoza[1], Mehdi Razavi[2], Joseph R. Cavallaro[1]

[1] Rice University, Houston, TX, United States
[2] Texas Heart Institute, Houston, TX, United States

Abstract

Left Ventricular Assist Devices (LVADs) are increasingly used as long-term implantation therapy for advanced heart failure patients, where candidacy assessment is crucial for successful treatment and recovery. A Deep Learning system based on Electrocardiogram (ECG) diagnoses criteria to stratify candidacy is proposed, implementing multi-model processing, interpretability, and uncertainty estimation. The approach includes beat segmentation for single-lead classification, 12-lead analysis, and semantic segmentation, achieving state-of-the-art results on the classification evaluation of each model, with multilabel average AUC results of 0.9924, 0.9468, and 0.9956, respectively, presenting a novel approach for LVAD candidacy assessment, serving as an aid for decision-making.

1. Introduction

A Left Ventricular Assist Device (LVAD) is an implantable device, which is increasingly being used as destination therapy for advanced heart failure management.

Determining the suitability of a patient and the proper timing for intervention is important for successful LVAD implantation and patient recovery. Deep Learning tools that can help assess the severity of the heart's condition in a patient can aid the specialist in making the decision. Electrocardiogram (ECG) signals, fundamental tools in cardiology, are proposed as input data for the system.

Deep Learning for ECG classification has been used in recent years, with Convolutional related architectures still being the most widely used for this end, and the potential and need for Interpretability and Uncertainty awareness for real-world application is evidenced [1].

This work focuses on the candidacy assessment problem and proposes ECG criteria that are then used to train the models. It combines three models to get a report that highlights information that is useful for the physician. It also implements interpretability and uncertainty estimation to increase the trust and applicability of the system.

2. Criteria, Interpretability and Uncertainty

For identification of the ECG diagnoses and the validation of the proposed criteria, expert input from professionals at the Texas Heart Institute was sought. The selected diagnoses would be most helpful in selecting LVAD candidates, by discriminating problems - often related to the left ventricle - while at the same time looking for potential right heart problems (up to 53% of LVAD patients have right heart failure after implantation [2]).

The proposed ECG diagnosis criteria are divided in three groups. Major criteria proposed are Left Bundle Branch Block (LBBB), Premature Ventricular Contraction (PVC), Left Ventricular Hypertrophy (LVH), Anterior Myocardial Infarction (AMI) and Congestive Heart Failure (CHF), as well as QRS duration, which has a risk/relevance for candidacy that increases linearly starting at a duration of 110 miliseconds. Minor criteria are Inferior Myocardial Infarction (IMI) and Atrioventricular Block (AVB); and potential contraindications relate mostly to problems with the right chambers of the heart, and comprise Right Ventricular Hypertrophy (RVH), Right Bundle Branch Block (RBBB) and Right Atrial Enlargement/Overload (RAE). The output is the estimated probability of having each one of these diagnoses. No single numerical score is proposed as a final measure of candidacy, rather a more informative report is proposed.

Interpretability in healthcare applications is important for increasing trust and model adoption. Uncertainty estimation is also important [3]. If an observation is classified as having high uncertainty, it is noted in the output report to be considered further by the physician.

3. System for LVAD Candidate Assessment From ECG

3.1. System overview

The proposed system consists of 3 main parts: i) Single-lead classifier, ii) 12 lead classifier, and iii) Semantic seg-

mentation classifier. All datasets used were publicly available datasets found on Physionet [4]. More details about the datasets and the processing of the data are in Section 4. The single lead and 12-lead classifiers output predicted probabilities, shown in the report as probability bins: 0-30%: Not detected, 30-45%: Cannot rule out, 45-60%:Consider, 60-75%:Possible and 75-100%:Consistent with. This choice was taken to provide a more flexible approach to decision making, by having the outputs of the system in a more nuanced manner, similar to real-world annotations that are often not binary. Interpretability results from Grad-CAM are available for the 12-lead classifier, and uncertainty awareness results from Monte Carlo Dropout are shown in the output report.

3.2. Single –lead model

Some of the diagnoses of interest can be correctly classified from a single lead, and high quality datasets are available with beat-level annotation for them. The model used for the single lead classifier is based on a 1D Convolutional Neural Network with residual blocks.

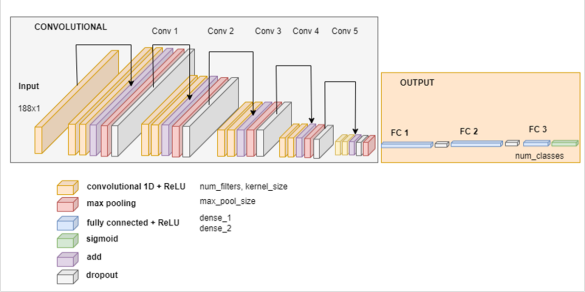

Figure 1. Single lead ECG classification model. Consists of five levels with 1D Convolutional layers, max-pooling, dropout and skip connections, followed by fully connected layers and sigmoid activation.

Fig. 1 shows the model, which consists of 5 sections, each with 1D-convolutional, max-pooling, dropout layers and residual connections. After the last section, three fully connected layers are added for classification followed by a final sigmoid activation unit. After using Keras Tuner to find the best combination of hyperparameters, a model with 64 filters, Kernel size of 6, MaxPool size of 4, and 48/80 units in the last two dense layers was used.

Training was performed using the Adam optimizer, with a learning rate scheduled with an exponential decay rate (0.0001 initial learning rate, 0.75 rate), with callback monitoring validation loss, for 30 epochs. The loss function employed was Binary Cross-entropy.

3.3. 12 lead model

Some ECG diagnoses are better classified using more than a single lead. Furthermore, some datasets had ECGs belonging to the diagnoses of interest with data for all 12 leads, which made it a good addition to the system. Regarding its architecture, the model takes as its basis the single-lead classifier, adding a sixth level of depth, still with 1D-convolutional layers as the main operation, implementing one such group of layers for every input channel, i.e. every lead. After concatenation of the extracted features of the 12 channels, Dense layers are added along with a final sigmoid activation for multilabel classification.

Training was performed using the Adam optimizer, with a learning rate scheduled with an exponential decay rate (0.0001 initial learning rate, 0.75 rate), with callback monitoring validation loss, for 15 epochs. The loss function employed was Binary Cross-entropy.

3.4. Semantic segmentation model

For semantic segmentation of the ECG signals, used in this system to determine the QRS average duration in milliseconds, the U-net model first proposed in [5] was adapted to process the 1-D signals of each ECG lead. It comprises 5 levels of grouped max-pooling, 1D-convolutional and dropout layers and performs upsampling with 1D-transposed convolutional layers. It receives as input a single lead waveform of 1000 samples and outputs a segmentation mask of background, P segment, QRS segment and T segment for each of the 1000 samples from the last Softmax activation layer.

Training was performed using the Adam optimizer, with a learning rate scheduled with an exponential decay rate (0.0001 initial learning rate, 0.75 rate), with callback monitoring validation loss for 85 epochs with a batch size of 64. Categorical Cross-entropy was used as the loss function.

4. Design And Implementation

4.1. Datasets

Public ECG datasets were selected from Physionet [4], looking for the diagnoses of interest. To increase the robustness of the model each diagnosis of interest had observations from at least two datasets included in the training of the models. The datasets used in this work are the following:

- MIT-BIH Arrhythmia Database: Normal, Other beat, LBBB, RBBB, PVC
- MIT-BIH Supraventricular Arrhythmia Database: Normal, Other beat, PVC
- PTB Diagnostic ECG Database: Normal, AMI, IMI, CHF

• PTB-XL: Normal, LBBB, RBBB, LVH, RVH, AVB, RAE/RAO, AMI, IMI
• Lobachevsky University Electrocardiography Database (LUDB): Normal, LBBB, RBBB, LVH, RVH, AVB, RAE, PVC, AMI, QRS duration (semantic segmentation of P, QRS, T segments)
• BIDMC Congestive Heart Failure Database: CHF

Training, validation, and test split sets were performed, in an approximately 0.7/0.2/0.1 split for the single-lead and semantic segmentation models, and a 0.8/0.1/0.1 split for the 12-lead model. In all cases, the split was stratified, keeping the same ratio among all classes in each subset of the data. In all cases, an inter-patient paradigm has been used, i.e., the ECG data (the heartbeats) of a patient can only be used in one of the training, validation, or test sets to avoid data leakage. For two minority classes (RVH and RAE) augmentation was performed doing a slight stretch that expands the signal in the time domain (with a random factor 1.05 to 1.3); and scaling (by a random factor between -0.875 and 1.125). The augmentations have been applied only to the 12-lead classifier.

As preprocessing, for the single lead classifier the R-peaks are detected and the beats segmented. For both single and 12-lead classifiers, denoising with a 4th order Butterworth high pass filter and wavelet filtering is used. For the single lead and semantic segmentation models, min-max normalization was applied.

Table 1. Precision, Recall, F-1 and AUC results for the Single-lead, 12-lead and Semantic Segmentation models.

Class	Precision	Recall	F1-score	AUC
Single Lead Classifier				
Normal	0.91	0.99	0.95	0.994
Other	0.97	0.93	0.95	0.997
LBBB	1	1	1	1
RBBB	1	1	1	1
PVC	0.97	0.96	0.97	0.998
CHF	0.99	0.96	0.97	0.989
MI	0.96	0.96	0.96	0.997
12 Lead Classifier				
Normal	0.87	0.92	0.90	0.963
LVH	0.74	0.73	0.73	0.959
RBBB	0.90	0.82	0.86	0.988
IMI	0.75	0.78	0.77	0.950
AMI	0.79	0.82	0.80	0.969
AVB	0.52	0.57	0.55	0.956
RVH	0.13	0.83	0.22	0.945
RAO/RAE	0.43	0.50	0.46	0.945
LBBB	0.95	0.86	0.90	0.989
Semantic Segmentation				
Other	0.97	0.97	0.97	0.9918
P	0.88	0.90	0.89	0.9972
QRS	0.93	0.94	0.94	0.9990
T	0.90	0.90	0.90	0.9946

4.2. Model Evaluation Results

The three models were evaluated following the recommended metrics for each case. Precision, Recall, F-1 Score, and AUC per class are evaluated. The obtained results for the three models are shown in Table 1. Additionally, the Intersection-Over-Union (IoU) score is calculated for the Semantic Segmentation model, taking the ratio of lengths instead of areas, measuring the overlap between the predicted and ground truth regions. The obtained IoU scores from the test set with the semantic segmentation model are 0.902, 0.741, 0.867, and 0.786 for Other (Background), P, QRS, and T segments respectively. The weighted IoU score is 0.871.

4.3. Output Report

Interpretability was implemented with Grad-CAM [6], applied to the last 1D-convolutional layer of each of the 12 heads of the model, one per input lead. Then the results are plotted over the 1D signal of the lead being examined with a heatmap. Fig. 2 shows an example.

As shown in [7], the Monte Carlo (MC) dropout technique can yield a good approximation of the posterior probability distribution of a model. This was implemented in the system and plotted with notched box-plots. The observation is forward-passed through the model 100 times. At inference time, the system outputs a report highlighting the major, minor, and potential contraindication criteria, along with the saliency maps obtained. The physician can also inspect the boxplots of the reported criteria of interest, which also show the estimated uncertainty of the results. Fig. 3 shows four examples, one of a high candidacy result (a), a second one of a heart predicted as normal (b), a third one of a candidate with high potential contraindication (c), and the last one is an example of a high uncertainty observation (d). The rest of the report of a) reads: Major - Consistent with LBBB, Possible AMI, QRS duration 167ms. Minor - Cannot rule out AVB. Report of b) reads: Major - Consistent with NORMAL, QRS duration 112.5 ms (the next predicted probability is shown for comparison, AVB with low value). Report of c) reads: Major - Cannot rule out AMI, QRS duration 128.6 ms. Minor - Possile IMI, Consider AVB. Potential Contraindications - Consistent with RBBB. Report of d) reads: Major - Consider NORMAL (high uncertainty), Consider LVH.

5. Conclusions

This work presents the first step in a novel approach to assessing candidacy for LVAD implantation directly from

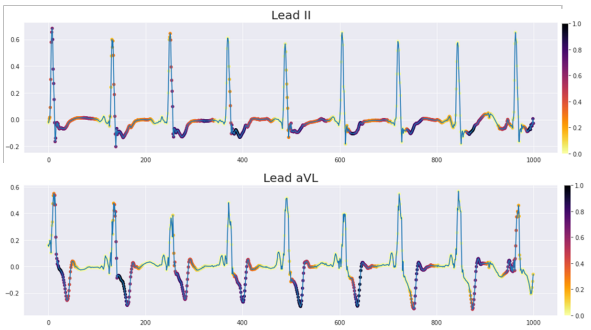

Figure 2. Saliency map showing the weights on the last convolutional layer of the top two leads (Lead II and Lead aVL) for an observation with LBBB and AMI.

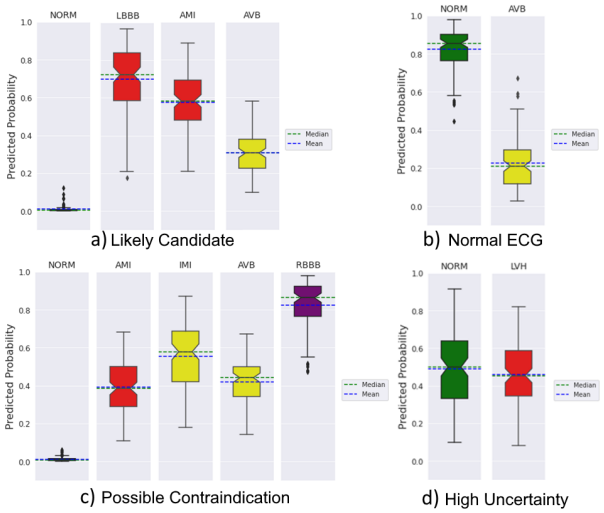

Figure 3. Example boxplots, showing the reported probability of relevant diagnoses for each case.

physiological signals, in this case from the ECG.

A multi-model Deep Learning system was built, achieving state-of-the-art results on each model and combining the predictions into a report.

The importance of having Interpretability was confirmed, and a way of showing per lead and segment importance has been implemented.

To increase trust in the model, uncertainty calculation was implemented, reporting graphically the predicted probability and uncertainty.

5.1. Limitations and Future Work

ECG alone does not convey the full information that would be desirable to make a complete assessment of a patient for the suitability of LVAD implantation. The use of clinical data to complement the models will be explored in this sense. The current work used a single ECG snapshot for each patient; having ECG data for patients at different points in time would help establish progression markers for the diagnoses of interest and improve the system.

In future work, models to estimate Left Ventricular Ejection Fraction will be developed, and the system will be validated with the ECGs of LVAD patients.

Measures to improve the classification results for rare diagnoses will be explored, such as the addition of handcrafted features that can be calculated from Semantic Segmentation results.

Acknowledgments: This work was supported by the United States Department of Defense grant HT9425-23-1-0663 and National Institutes of Health grant 1R01HL166724-01A1.

References

[1] Hong S, Zhang W, Sun C, Zhou Y, Li H. Practical lessons on 12-lead ECG classification: Meta-analysis of methods from PhysioNet/computing in cardiology challenge 2020. Frontiers in Physiology Jan. 2022;12:811661.

[2] Chaudhry S, DeVore AD, Vidula H, Nassif M, Mudy K, Birati EY, Gong T, Atluri P, Pham D, Sun B, Bansal A, Najjar. Left ventricular assist devices: A primer for the general cardiologist. Journal of the American Heart Association Dec. 2022;e027251.

[3] Kompa B, Snoek J, Beam AL. Second opinion needed: communicating uncertainty in medical machine learning. npj Digital Medicine 2021;4(1):1–6.

[4] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 2000;101(23):E215–220. ISSN 1524-4539.

[5] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In Navab N, Hornegger J, Wells WM, Frangi AF (eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Cham: Springer International Publishing. ISBN 978-3-319-24574-4, 2015; 234–241.

[6] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. International Journal of Computer Vision Feb. 2020;128(2):336–359.

[7] Gal Y, Ghahramani Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48, ICML'16. JMLR.org, 2016; 1050–1059.

Address for correspondence:

Antonio Mendoza
Rice University, 6100 Main Street, Duncan Hall 3041, Houston, TX 77005, antonio@rice.edu