

Screening for Chagas disease from the electrocardiogram using a deep neural network

Carl Jidling¹; Daniel Gedon^{*1}; Thomas B. Schön^{*1}; Claudia Di Lorenzo Oliveira²; Clareci Silva Cardoso²; Ariela Mota Ferreira³; Luana Giatti⁴; Sandhi Maria Barreto⁴; Ester C. Sabino⁵; Antonio L. P. Ribeiro^{†, 4}, and Antônio H. Ribeiro^{*1}

¹Uppsala University, Sweden

²Universidade Federal de São João del-Rei, Brazil

³Universidade Estadual de Montes Claros, Brazil

⁴Universidade Federal de Minas Gerais, Brazil

⁵Universidade de São Paulo, Brazil

*{daniel.gedon, thomas.schon, antonio.horta.ribeiro}@it.uu.se

†antonio.ribeiro@ebserh.gov.br

Abstract

Background Worldwide it is estimated that more than 6 million people are infected with Chagas disease (ChD). It is considered one of the most important neglected diseases and, when it reaches its chronic phase, the infected person often develops serious heart conditions. While early treatment can avoid complications, the condition is often not detected during its early stages. We investigate whether a deep neural network can detect ChD from electrocardiogram (ECG) tracings. The ECG is inexpensive and it is often performed during routine visits. Being able to evaluate ChD from this exam can help detect potentially hidden cases in an early stage.

Methods We use a convolutional neural network model, which takes the 12-lead ECG as input and outputs a scalar number associated with the probability of a Chagas diagnosis. To develop the model, we use two data sets, which jointly consist of over two million entries from Brazilian patients, compiled by the Telehealth Network of Minas Gerais within the SaMi-Trop (São Paulo-Minas Gerais Tropical Medicine Research Center) study focused on ChD patients and enriched with the CODE (Clinical Outcomes in Digital Electrocardiology) study focused on a general population. The performance is evaluated on two external data sets of 631 and 13,739 patients, collected in the scope of the REDS-II (Retrovirus Epidemiology Donor Study-II) study and of the ELSA-Brasil (Brazilian Longitudinal Study of Adult Health) study. The first study focuses on ChD patients and the second data set originates from civil servants from five universities and one research institute.

Findings Evaluating our model, we obtain an AUC-ROC value of 0.80 (CI 95 % 0.79-0.82) for the validation data set (with samples from CODE and SaMi-Trop), and in external validation datasets: 0.68 (CI 95 % 0.63-0.71) for REDS-II and 0.59 (CI 95 % 0.56-0.63) for ELSA-Brasil. In these external validation datasets, we report a sensitivity of 0.52 (CI 95 % 0.47-0.57) and 0.36 (CI 95 % 0.30-0.42) and a specificity of 0.77 (CI 95 % 0.72-0.81) and 0.76 (CI 95 % 0.75-0.77), respectively, in REDS-II and ELSA-Brasil. We also evaluated the model for considering only patients with Chagas cardiomyopathy as positive. In this case, the model attains an AUC-ROC of 0.82 (CI 95 % 0.77-0.86) for REDS-II and 0.77 (CI 95 % 0.68-0.85) for ELSA-Brasil.

Interpretation The results indicate that the neural network can detect patients who developed chronic Chagas cardiomyopathy (CCC) from the ECG and – with weaker performance – detect patients before the CCC stage. Future work should focus on curating large and better datasets for developing such models. The CODE is the largest dataset available to us, and their labels are self-reported and less reliable than our other data sets, i.e. REDS-II and ELSA-Brasil. This, we believe, limits our model performance in the case of non-CCC patients. We are positive that our findings constitute the first step towards building tools for more efficient detection and treatment of ChD, especially in high-prevalent regions.

Funding This research is financially supported by the Swedish Foundation for Strategic Research (SSF) via the project *ASSEMBLE* (Contract number: RIT 15-0012), by the *Wallenberg AI, Autonomous Systems and Software Program (WASP)* funded by Knut and Alice Wallenberg Foundation, by *Kjell och Märta Beijer Foundation*, by the Brazilian Agencies CNPq, CAPES, and FAPEMIG, and by projects IATS, and CIA-Saúde. The ELSA-Brasil study was supported by the Brazilian Ministries of Health and of Science and Technology (grants 01060010.00RS, 01060212.00BA, 01060300.00ES, 01060278.00MG, 01060115.00SP, and 01060071.00RJ). The SaMi-Trop and REDS-II cohort studies are supported by the National Institutes of Health (P50 AI098461-02, U19AI098461-06, 1U01AI168383-01). LG, SMB, ECS and ALPR receive unrestricted research scholarships from CNPq; ALPR received a Google Latin America Research Award scholarship. The funders had no role in the study design; collection, analysis, and interpretation of data; writing of the report; or decision to submit the paper for publication.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Research in context

Evidence before this study: Chagas disease (ChD) is a neglected tropical disease, and the diagnosis relies on blood testing of patients from endemic areas. However, there is no clear recommendation on selecting patients for serological diagnosis in those living in endemic regions. Since most of the patients with Chronic ChD are asymptomatic or oligosymptomatic, the diagnostic rates are low, preventing patients from receiving adequate treatment. The Electrocardiogram (ECG) is a widely available, low-cost exam, often available in primary care settings in endemic countries. Artificial intelligence (AI) algorithms on ECG tracings have allowed the detection of hidden conditions, such as cardiomyopathies and left ventricular systolic dysfunction.

Added value of this study: To the best of our knowledge, this is the first study that presents an AI model for the automatic detection of ChD from the ECG. As part of the model development, we utilise established large cohorts of patients from the relevant population of all-comers in affected regions in the state of Minas Gerais, Brazil. We evaluate the model on data sets with high-quality ground truth labels obtained from the patients' serological status. Our model has moderate diagnostic performance in recognition of ChD and better accuracy in detecting Chagas cardiomyopathy.

Implications of all the available evidence: Our findings demonstrate a promising AI-ECG-based model capacity for discriminating patients with chronic Chagas cardiomyopathy (CCC). However, detecting ChD patients without CCC is still insufficient, and further developments that lead to higher performance are needed. We believe this can be achieved with the addition of epidemiological questions, and that our model can be a useful tool in helping pre-selecting patients for further testing in order to determine the infection with ChD. The use of AI-ECG-based strategies for recognizing CCC patients deserves to be tested in the clinical setting.

Introduction

Worldwide it is estimated that Chagas disease (ChD) infects more than 6 million people, with thousands of deaths each year [1]. Caused by the protozoan parasite *Trypanosoma cruzi* (*T. cruzi*), the disease is endemic to countries in continental Latin America, but migration has carried ChD to new regions, including Europe and the United States [2]. The most critical consequence of ChD is chronic Chagas cardiomyopathy (CCC), which occurs in 20-40% of the infected individuals [3]. CCC comprises a wide range of manifestations, including heart failure, arrhythmias, heart blocks, sudden death, thromboembolism, and stroke [1], [3].

ChD is often a lifelong infection in which most chronically infected patients remain asymptomatic but at risk of progression to cardiac damage [4], [5]. The incidence of cardiomyopathy in those in this asymptomatic (indeterminate) form of ChD varies from 0.9 to 7% new cases annually [1] and is related to the parasite burden [5], [6]. There is no single gold-standard laboratory test for diagnosing chronic Chagas disease. Instead, at least two serological tests with different methods for detecting antibodies to *T. cruzi* and complementary sensitivity and specificity are needed to confirm infection [1], [3]. Treatment with antitrypanoso-

mal drugs such as benznidazole can prevent progression to the cardiac form [7], [8], but it does not seem to prevent death and cardiac complications in those with advanced cardiomyopathy [9]. Thus, the early recognition of chronic ChD patients is a necessary step for treatment in the early phases, when treatment success rates are higher and can prevent severe organ damage from occur [10].

Even if the newly diagnosed patient has established cardiomyopathy, an early diagnosis will allow the initiation of guideline-directed medical therapy for clinical conditions, such as heart failure and atrial fibrillation, to halt disease progression and eventually prevent death [10]. ChD patients generally have low socio-economical levels and limited access to health services, and they frequently do not realize that they are infected. The awareness of ChD among healthcare providers is also low, and there is a lack of knowledge on who to screen as well as a lack of clarity on the appropriate tests and clinical management [11], [12].

In many countries, there are detection rates below 10%, even more frequently, below 1%. The low detection rates create a barrier to the health care system, preventing patients from receiving adequate treatment [13]. The underappreciation of early diagnosis and treatment, especially at the primary healthcare level, represents a missed opportunity for modifying the natural history of the disease [10]. For this reason, the theme of World Chagas Disease Day 2022 was "finding and reporting every case to defeat Chagas disease" [13].

Here we study the possibility of using the electrocardiogram (ECG) to screen for ChD. The ECG is a widely available, low-cost exam, often provided in primary care settings in endemic countries [14]. The automated analysis of ECG is a successful technology and has already improved the analysis of this exam over the past decades [15].

The field of artificial intelligence, in particular deep learning [16], has demonstrated promising performance for automated analysis. Besides the success of classifying common ECG diagnoses with high-performance [17], [18], the technology has presented successes in predicting and screening for diseases and diagnoses which traditionally were not directly possible only from the ECG. These include detection of myocardial infarction without ST-elevation [19], predicting the future development of atrial fibrillation from sinus rhythm exams [20], [21] and the ability to screen for cardiac contractile dysfunction [22]. Indeed, there is evidence that deep learning reading of ECGs detects more than traditional features, as is indicated by studies showing good prediction of age and even the risk of death [23]–[25].

In this study, we investigate whether a deep neural network can detect ChD and CCC from ECG tracings. Being able to evaluate ChD from this exam can help to detect cases in an early stage and enables early and more effective treatment.

Methods

Data sets

We develop our model using the SaMi-Trop data set [26], [27] and the CODE data set [28]. The SaMi-Trop data set is a collection of ChD patients from the northern part of Minas Gerais, Brazil. The CODE data set [29] is more general, collected by the Telehealth Network of Minas Gerais (TNMG), Brazil [28]. For testing or external validation, we use the REDS-II data set [30] and the ELSA-Brasil data set [31]. The baseline characteristics of all four data sets are summarised in Table 1.

Definitions. Chronic ChD is diagnosed by the presence of two positive different serological tests against *T. cruzi* in both SaMi-Trop and REDS-II cohorts, as recommended by international guidelines [3]. In the ELSA-Brasil study, a cohort primarily designed to study chronic non-communicable diseases, the presence of Chagas disease was detected by the presence of only one positive serological test. In the CODE study, Chagas disease was self-reported by the patients since this electronic cohort is formed by patients under care in primary care units in the state of Minas Gerais. For SaMi-Trop, REDS-II and ELSA cohorts, ECGs were transmitted to an ECG reading center at the ‘Centro de Telessaúde in Hospital das Clínicas’ in Belo Horizonte, Minas Gerais for standardized measurement, reporting and codification according to the Minnesota coding criteria in a validated ECG data management software [32]. Major ECG abnormalities were considered according to standard definitions [33], and all tracings with a major ECG abnormality have been reviewed by an experienced cardiologist.

CODE. The Clinical Outcomes in Digital Electrocardiography (CODE) data set was developed with the database of digital ECG exams of the TNMG and a detailed description of the cohort can be obtained at [29]. The data set was collected between 2010 and 2017 from 811 counties in the state of Minas Gerais, Brazil. A subset of 15 % of this data set is available online [34].

From an initial data set of 2,470,424 ECGs, 1,773,689 patients were identified. This initial data set contains the SaMi-Trop data set. Therefore, we first remove the patients from the SaMi-Trop study to avoid any overlap. Additionally, we have to exclude the ECGs with technical problems and those from patients under age 16, resulting in a total of 2,304,596 ECG records from 1,556,767 patients.

In this data set, the labels of ChD rely on self-reported diagnoses during the consultation. A total of 47,474 ECGs (2.0 %) from 25,252 patients (1.6 %) are labelled as positive ChD cases. The serological status of the self-reported Chagas labels has not been checked, and it is also unclear whether the patient has already developed CCC or not.

SaMi-Trop. The study was conducted through a collaboration between scientists within the São Paulo-Minas Gerais Tropical Medicine Research Center (SaMi-Trop), formed with a specific research focus on ChD. [35] The study selected eligible patients with self-reported ChD diagnosis. This data set was collected in 21 Brazilian mu-

nicipalities from ECGs taken between 2010 and 2012 by the TNMG. The connection to the TNMG explains the intersection of the SaMi-Trop data set with the CODE data set. The study has a follow-up time of two years.

A total of 2,157 patients were assessed in the study. Among the patients from the original SaMi-Trop study, we removed 22 patients with an undefined serological status, and the remaining 83 for not having a paired ECG recording. After the exclusions, the resulting data set comprises 2,054 patients with 1,910 ChD positive patients (93.4 %). The positive patients consist of 1,111 patients with CCC (54.1 % of total sample) and 799 without (38.9 % of total sample).

Some of the patients have taken multiple ECG recordings during an exam which we utilize during development as a form of data augmentation. Hence, we have 5,019 SaMi-Trop ECG traces available including 2,693 traces with CCC (53.7 %) and 1,961 traces without (39.1 %).

REDS-II. The Retrovirus Epidemiology Donor Study-II (REDS-II) data set was collected to observe the natural history of ChD patients in São Paulo and Montes Carlos, Brazil from blood donors. Seropositive and seronegative patients examined in 1996-2002 were re-examined in 2008-10 [4] with ECG exams and again in 2018-19 [30]. The data set consists of 631 patients that performed an ECG in the last visit in 2018-19, including 348 ChD patients (55.8 %), of which 149 patients had CCC (23.6 % of the total sample). The model is evaluated using a single exam from each patient (the first one).

ELSA-Brasil. The Brazilian Longitudinal Study of Adult Health (ELSA-Brasil) aimed to examine risk factors and the long-term incidence of chronic diseases with focus on cardiovascular diseases and diabetes. The baseline evaluation was performed in 2008-2010 and recruited active and retired civil servants from five universities and research institutes from 6 different Brazilian states. ChD serological status and standardized ECG were obtained from all participants [36], [37].

The data set consists of 15,105 patients in total. We remove 27 patients where the ChD serological status is not available, 12 patients where the serological status is inconclusive, and 1,327 patients from which the ECG traces are not available. After the exclusions, we have a data set with a total of 13,739 patients. ChD was confirmed in 280 of the patients (2.0 %), of which 46 had CCC (0.3 % of the total sample). The model is evaluated using a single exam from each patient (the first one).

Model

Data preprocessing. The ECG signals have been re-sampled such that all ECGs have the same sampling frequency of 400 Hz. Each input ECG has 4,096 time samples for each of the 12 standard ECG leads. Original signals of a shorter time span have been extended through zero-padding. The output data comprises binary scalar variables corresponding to positive or negative diagnose. We combine positive cases with and without CCC in our model in order to focus on the class of positive ChD cases in general.

Table 1: Data set baseline characteristics. For CODE the Chagas patient reports their own condition, while for the others the blood sample is used to determine the serological status. Smoking gives the current smoking habits of the patient. CCC: Chronic Chagas cardiomyopathy which is not available (n.a.) for the CODE data set. MI: Myocardial infarction. Dislipidemia is not available (n.a.) for SaMi-Trop or REDS-II data sets.

Development data sets						
	CODE (Patients, n = 1,556,767)			SaMi-Trop (Patients, n = 2,054)		
	Not reported	Chagas		Seronegative	Chagas	CCC
Patients, n	1,524,766	32,001		144	1,910	1,111
Sex (male), n (%)	922,780 (39.7 %)	11705 (36.6 %)		60 (41.7 %)	629 (32.9 %)	398 (35.8 %)
Age (years), mean (sd)	52 (18)	59 (14)		64.0 (14.7)	59.2 (12.9)	60.4 (13.0)
Hypertension, n (%)	482,184 (31.6 %)	21,849 (68.3 %)		112 (77.8 %)	1219 (63.8 %)	762 (68.6 %)
Diabetes, n (%)	102803 (6.7 %)	4,398 (13.7 %)		31 (21.5 %)	195 (10.2 %)	107 (9.6 %)
Smoking, n (%)	108,815 (7.0 %)	4,190 (13.1 %)		49 (38.9 %)*	49 (38.9 %)*	301 (29.2 %)*
Previous MI, n (%)	13,693 (0.9 %)	1,308 (4.1 %)		12 (8.3 %)	90 (4.7 %)	69 (6.2 %)
Dislipidemia, n (%)	62,686 (4.1 %)	3,823 (11.9 %)		n.a.	n.a.	n.a.
Obesity, n (%)	86,500 (5.7 %)	3,321 (10.4 %)		24 (21.8 %)*	222 (15.0 %)*	106 (12.8 %)*
Exams, n	2,257,122	47,474		365	4,654	2,693

Test data sets						
	REDS-II (Patientes, n = 631)			ELSA-Brasil (Patientes, n = 13,739)		
	Seronegative	Chagas	CCC	Seronegative	Chagas	CCC
Patients, n	283	348	149	13,459	280	46
Sex (male), n (%)	140 (49.5 %)	171 (49.1 %)	82 (55.0 %)	167 (59.6 %)	19 (41.3 %)	6,256 (45.5 %)
Age (years), mean (sd)	58.2 (9.6)	56.1 (9.8)	56.4 (9.8)	52.1 (9.2)	57.4 (9.1)	61.1 (6.8)
Hypertension, n (%)	101 (35.7 %)*	136 (39.1 %)*	62 (41.6 %)*	4,814 (35.8 %)*	135.0 (48.2%)*	27.0 (58.7 %)*
Diabetes, n (%)	38 (13.4 %)*	45 (12.9 %)*	16 (10.7 %)*	2,160 (16.0 %)	57.0 (20.4 %)	15.0 (32.6 %)
Smoking, n (%)	94 (33.2 %)	98 (28.2 %)	42 (28.2 %)	1 769 (13.1 %)	32 (11.4 %)	4 (8.7 %)
Previous MI, n (%)	9 (3.2 %)*	13 (3.7 %)*	8 (5.4 %)*	242 (1.8 %)	9 (3.2 %)	5.0 (10.9 %)
Dislipidemia, n (%)	n.a.	n.a.	n.a.	8,853 (65.8 %)	201 (71.8 %)	36 (78.3 %)
Obesity, n (%)	68 (24.6 %)*	66 (19.4 %)*	28 (19.0 %)*	3,079 (22.9 %)	67 (23.9 %)	10 (21.7 %)
Exams, n	283	348	149	13,459	280	46

*For CODE, comorbidity data is self-reported and might be underrepresented. There is missing data regarding smoking (seronegative=18/chagas=148/ccc=81 missing entries) and obesity (34/431/281 missing entries) in the SaMi-Trop data set; regarding Hypertension (12/9/3 missing entries), Diabetes (2/8/4 missing entries), Previous MI (0/5/1 missing entries) and Obesity (7/8/2 missing entries) in the REDS-II data set; and, regarding Hypertension (16/0/0 missing entries) fin the ELSA-Brasil. For these entries, we report the percentages that consider the total of patients without missing data.

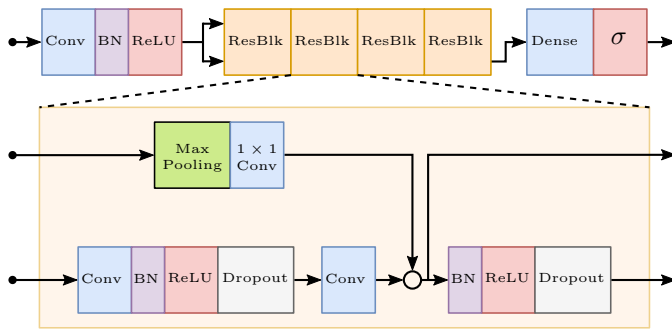


Figure 1: **Network architecture.** The figure was originally illustrated in [17].

Architecture. The deep learning model consists of a residual neural network (ResNet) adapted to uni-dimensional signals, and includes convolutional layers both before and within the residual blocks. Our network architecture is visualised in Figure 1. We make use of the same network architecture as [17], where the CODE data set was utilised to classify multiple ECG abnormalities; we refer to that work for further details and note that we have modified the final output layer in adaptation to our binary classification. The model is implemented in PyTorch [38], building upon code used in related work [39], [40].

Parameter tuning. The learnable parameters of the neural network are chosen through minimisation of the binary cross-entropy loss function. For increased computational efficiency, we split the training data into mini-batches of size 32.

We use both the CODE and SaMi-Trop data sets during the training phase. This way, we utilise the size of the CODE data set — with many examples of negative diagnoses — as well as the high-quality (mainly positive) entries of SaMi-Trop. Both data sets contribute with 50% of the data that the model experience in each mini-batch. The validation data is an independent mix of 30% of the SaMi-Trop entries and twice as many entries from CODE.

The dropout rate is 0.5, and we use a weight decay of 0.001 to reduce the risk of overfitting. The learning rate is initially set to 0.001 and is decreased in a step-wise manner by a factor 10 when the validation loss has not decreased for ten subsequent epochs (counted with respect to SaMi-Trop) — we terminate the optimisation if the learning rate drops below 10^{-7} . We apply early stopping by using the network parameter values associated with the lowest validation loss for testing.

To reduce the sensitiveness of the weight initialisation, we use an ensemble approach by running the optimisation 15 times with different random seeds, and then averaging the outputs of the final models. The progression of the losses evaluated on the training and validation data sets are displayed in Figure 2.

Threshold selection. The model output is a value between 0 and 1 and can loosely be interpreted as the predicted probability of ChD being present in the exam analysed. The Chagas diagnose is predicted as positive when the model output is above a given classification threshold.

We consider two different approaches to selecting the

threshold. The first one is by maximising the F1 score (i.e. the harmonic mean of precision and recall) on the validation data. This threshold is suitable for balanced or moderately imbalanced data sets where the main interest is to diagnose the patients under consideration.

The second approach is to choose the threshold by requiring a certain specificity on the validation data. The higher the specificity, the more likely is the model to correctly diagnose a negative patient. As a high specificity typically is desired for screening purposes, this approach for threshold selection is motivated on highly imbalanced data sets (which reflects the Chagas prevalence in the population as a whole).

The first approach is used on the REDS-II test set since this data set is only moderately imbalanced (55.8% ChD and 23.6% CCC ECGs). On the ELSA-Brasil test set the threshold is selected according to the second approach since this data set is more imbalanced (2.0% ChD and 0.3% CCC ECGs). We select the threshold by requiring a 90% specificity on the validation data.

Evaluation

Metrics. Recall (also known as sensitivity), specificity and precision are threshold-dependent metrics that we used to evaluate and report the model performance. Recall or sensitivity specifies the ratio of true positive predictions to positive cases (i.e. the ratio of the positive cases that are indeed predicted as positive); specificity denotes the ratio of true negative predictions to negative cases; and precision is the ratio of true positive predictions to all positive predictions (the ratio of all positive predictions that are correct).

We also report two threshold-independent metrics. The AUC-ROC (also known as c-statistics) is the integral of the receiver-operator characteristics (ROC), and can be interpreted as the probability that a randomly chosen sample with positive label is assigned a higher output than a randomly chosen sample with negative label. Lastly, we report the average precision, which is obtained by integrating the precision-recall curve and thereby summarising it into a single value.

Analysis of the results in groups. As part of the model analysis we evaluate the model performance in different subgroups of patients. We stratify the patients by age group {16-40, 40-49, 50-59, 60-69, 70+} and sex {male, female}. Bootstrapping [41] is used to analyse the empirical distribution of the metrics in each subgroup. We generate 1,000 different data sets by sampling with replacement from the test set (each with the same number of samples as in the test set). Using the bootstrapped data sets, we compute the evaluation metrics described above and present the results in box plots.

Visualisation tools. To identify possible patterns in the classification, we highlight parts of the ECG that the model focuses on for its prediction using an adaptation of the Grad-CAM visualisation method [42]. Visualisations are generated in two steps: in a forward pass we compute the activations of the neural network in an intermediary layer (we use the first convolutional layer of the first residual

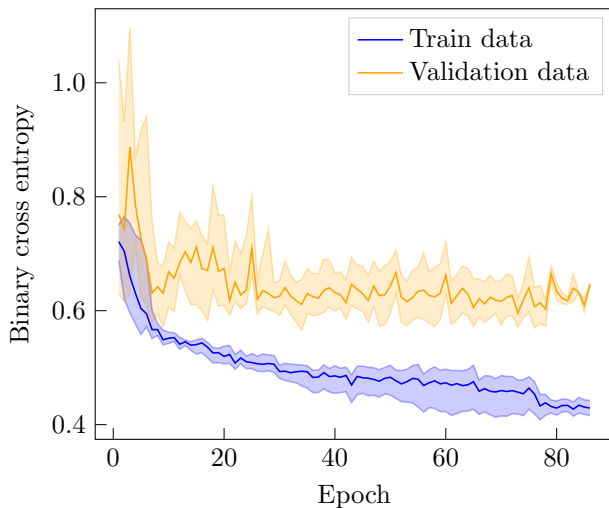


Figure 2: **Loss function evaluation.** The shaded regions correspond to the maximum and minimum values of 15 separate learning processes with different weight initialisations. The solid lines are the averages.

block), and in a backward pass we compute the gradients corresponding to these activations. The gradients are averaged to get the relative importance of each channel, which is then used to compute a proportional mean of the activations.

In essence, these plots highlight which parts of the ECG the network assigns particularly high importance. We generated the Grad-CAM plots for 20 cases (10 with CCC and 10 without) with the highest probability among the true positive cases. These plots were then inspected and analysed by a cardiologist for possible medical patterns.

Results

We evaluated the model performance on the validation data and the external test data sets. The ROC curve performance is displayed in Figure 3. The model attains AUC-ROC values of 0.80 (CI 95 % 0.79-0.82) for the validation data set, 0.68 (CI 95 % 0.63-0.71) for REDS-II and 0.59 (CI 95 % 0.56-0.63) for ELSA-Brasil. The confidence intervals have been formed by bootstrapping the output of the ensemble model. Table 2 lists all performance metrics evaluated on the validation data for two different thresholds selected through the aforementioned approaches. The same metrics evaluated on the test data sets are listed in Table 3. Additionally, we also analysed the precision-recall curve and the empirical probabilities predicted by the model. These results are displayed in the Supplementary Material Figures S.1-S.4. The metrics for subgroups stratified by age and sex are displayed in Figure 5.

We also evaluated the model for considering only patients with CCC as positive. In this case, the model attains an AUC-ROC of 0.82 (CI 95 % 0.77-0.86) for REDS-II and 0.77 (CI 95 % 0.68-0.85) for ELSA-Brasil. All metrics for this configuration are included in Table 3.

In the Supplementary Material Figure S.5 and Table S.1,

we show the additional results for another test set configuration. Namely where the patients with CCC have been excluded; the remaining patients where ChD was detected are here constituting the positive cases (this configuration is indicated “no CCC”). In the Supplementary Material, we also show the result of a model trained to detect CCC (with all others being considered negative). Figure S.6 shows the training curve, Figure S.7 shows the ROC curves, precision-recall curves and empirical distribution of the probabilities, and finally, Table S.2-S.3 give the performance metrics in this case.

The Grad-CAM analysis is presented in Figure 6, which shows three representative leads of a patient with CCC from the ELSA-Brasil data set. The shaded regions illustrate what parts of the signals the model considers to be of particular importance for the prediction. In the Supplementary Material Figure S.8 we include the equivalent plots for another three patients with positive Chagas diagnose, one with and two without CCC.

Discussion

Deep neural network-enabled analysis of the ECG is a topic of intense research [19]–[25]. Such methods have shown promising potential in detecting diverse conditions that are not traditionally diagnosed from the ECG, such as contractile dysfunction [22] or non-STEMI myocardial infarction [19]. ChD is the parasitic disease with the most impact in South America [43] and it affects the lives of millions of individuals worldwide. Early detection of this disease can therefore have a huge impact. Antiparasitic drugs are most effective in the early stage of the disease, however, most patients only become aware that they are infected much later when the patient is already in the later stage of the disease and presents other manifestations. Providing early treatment and the usage of advanced artificial intelligence or machine learning methods for the detection of this disease presents itself as a promising alternative. To the best of the authors’ knowledge, this is the first study to present such an application.

The development of data-driven methods for automatic diagnosis of neglected diseases presents a challenge of its own. These diseases usually affect areas where the population is underprivileged and have little access to the healthcare system. The data might not come in well-organised databases or might not even be stored in electronic format. In this sense, the CODE, SaMi-Trop, ELSA-Brasil and REDS-II cohorts are extremely valuable: they are medium or large-size and well-kept data sets that can be used for developing and testing such tools.

The results we present are promising and indicate that the model is capable of detecting patients with CCC from the ECG tracings with high discrimination. For patients without CCC, the discrimination is lower.

In light of the results, it is natural to ask if we can further improve the performance with respect to patients with CCC. Therefore, we restrict the positive diagnoses to patients with CCC during the training phase and consider all patients without CCC as negatives (this implies that

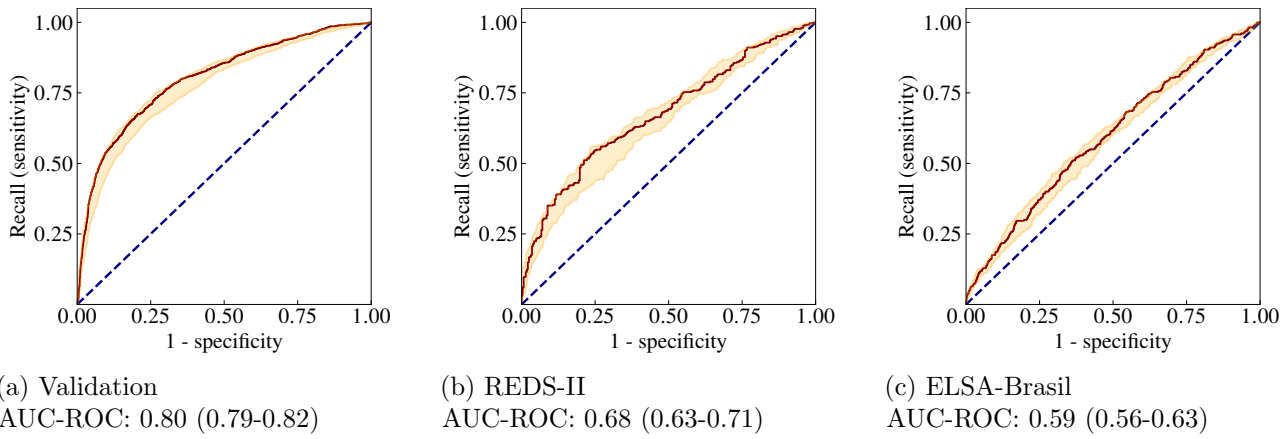


Figure 3: **ROC curves on validation and test data: ChD vs normal.** Receiver operating characteristics (ROC) computed on the validation data set (SaMi-Trop + CODE) (a) and the test data sets, REDS-II (b) and ELSA-Brasil (c). The shaded regions encapsulate the maximum and minimum values corresponding to 15 different weight initialisations. The outputs of the 15 trained models are averaged to produce the output of the ensemble model, the result of which is shown by the solid lines. The 95% CI is obtained by bootstrapping the ensemble model. The dotted blue lines correspond to completely random assignment of class probabilities.

Table 2: **Results on validation data.** Metrics and 95% confidence intervals evaluated on the validation data set for two different classification thresholds: 0.60 (selected by maximising the F1 score) and 0.71 (corresponding to 90% specificity).

Metric \ Threshold	F1 score-based: 0.60	Specificity-based: 0.71
Recall	0.67 (0.65-0.70)	0.54 (0.51-0.57)
Specificity	0.80 (0.78-0.81)	0.90 (0.89-0.91)
Precision	0.62 (0.59-0.64)	0.73 (0.70-0.75)
F1 score	0.64 (0.62-0.66)	0.62 (0.60-0.64)
AUC-ROC	0.80 (0.79-0.82)	
Average precision	0.68 (0.65-0.70)	

ChD positive patients without CCC are considered negative in this scenario). The result of this approach is given in the Supplementary Material. All metrics considered, except for the recall, are indeed improved. Thus, this model might be the preferable choice for CCC detection.

Chagas cardiomyopathy is characterised by a group of typical ECG abnormalities, frequently combining conduction disturbances, especially right bundle branch block with left anterior hemiblock, associated with rhythm disorders, such as ventricular ectopic beats and atrial fibrillation [44]–[46]. Thus, it is unsurprising that our Grad-CAM analysis depicts exactly the late portion of the QRS in cases with a bundle branch block. It is interesting that the Grad-CAM map also depicts the QRS complex when recognising the ChD patients with CCC, maybe related to the presence of high frequency, low amplitude abnormalities typical of fibrosis, which can occur early in the natural history of ChD [47]. However, this type of analysis has clear limitations [48], [49] since heatmaps can provide information on where the critical area for the neural network model is to make a decision but not inform if the abnormality is related to changes in voltage, duration or morphology modification of the ECG tracing. Moreover, recurrent features, like the

RR interval, are not shown in this kind of heatmaps. Our analysis here is also limited to a small set of correct model predictions and does not represent a statistical analysis. Hence, we cannot deduct general rules for the diagnosis of ChD but we can identify from the unsurprising areas where the model focuses on that it does not use some unrelated proxy information to make its predictions.

Comparing the two test data sets, we obtain similar performance for discrimination in terms of AUC-ROC, but very different precision. This indicates that our model predicts many false positives for the ELSA-Brasil data set. Given the vast difference in prevalence for ChD patients in ELSA-Brasil (2.0%) and REDS-II (55.1%), it is reasonable that for ELSA-Brasil our model will by default have lower precision. We can also observe the large portion of false positive cases in Figure S.3c when choosing a threshold of 0.60 (based on F1 score) or even 0.71 (based on 90% specificity). We believe the performance could be improved with the addition of epidemiological questions, and that our model can be a useful tool in helping pre-selecting patients for further testing in order to determine the infection with ChD.

As previously mentioned, the ChD status in the CODE

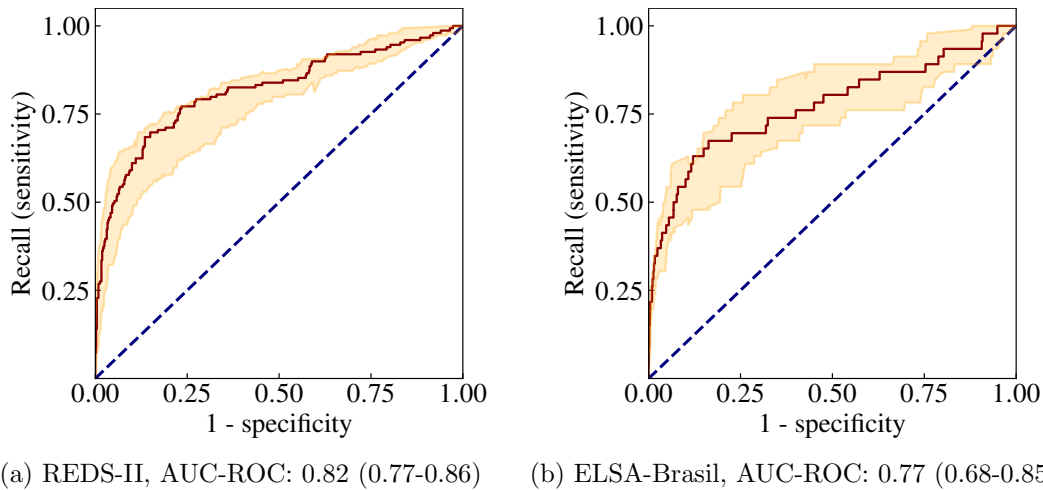


Figure 4: **ROC curves on test data: CCC vs all.** Receiver operating characteristics computed in REDS-II and ELSA-Brasil for predicting Chagas Cardiomyopathy. The shaded regions encapsulate the maximum and minimum values corresponding to 15 different weight initialisations – the outputs of these models are averaged to produce the output of the ensemble model, the result of which is given by the solid lines. The dotted blue lines correspond to completely random assignment of class probabilities.

Table 3: **Results on test data.** Metrics and 95 % confidence intervals were evaluated on two different configurations of the test data sets. Left we consider ChD and CCC as positive. Right we only consider CCC as positive. The classification thresholds are 0.60 for REDS-II and 0.71 for ELSA-Brasil.

Metric \ Test data	REDS-II	ELSA-Brasil	REDS-II (CCC)	ELSA-Brasil (CCC)
Recall	0.52 (0.47-0.57)	0.36 (0.30-0.42)	0.79 (0.72-0.85)	0.70 (0.56-0.82)
Specificity	0.77 (0.72-0.81)	0.76 (0.75-0.77)	0.73 (0.69-0.76)	0.76 (0.75-0.77)
Precision	0.73 (0.68-0.79)	0.03 (0.02-0.04)	0.47 (0.41-0.53)	0.01 (0.01-0.01)
F1 score	0.61 (0.57-0.66)	0.06 (0.05-0.07)	0.59 (0.53-0.64)	0.02 (0.01-0.03)
AUC-ROC	0.68 (0.63-0.71)	0.59 (0.56-0.63)	0.82 (0.77-0.86)	0.77 (0.68-0.85)
Average precision	0.74 (0.69-0.78)	0.04 (0.03-0.06)	0.69 (0.61-0.76)	0.10 (0.03-0.19)

data set is based on self-reporting by the patients, and the labels are thus suffering from notable uncertainty. Thus, testing on these labels might be uninformative and we have used more reliable databases such as ELSA-Brasil and REDS-II to get a better estimate of our model performance. Nonetheless, the labels in CODE still contain a sufficient amount of information to learn about CCC patients and the data set was indeed useful in developing a better-performing model. Methods designed to reduce the impact of label noise (see e.g. [50], [51]) could potentially be employed for more efficient use of the CODE data.

Our model could be even more insightful if we could test it on other openly available data sets. However, data sets about neglected diseases are scarce and both ELSA-Brasil as well as REDS-II are valuable but also medium to large-scale sources to rigorously test the model. Furthermore, a comparison with other models or software for Chagas detection would be useful, but unfortunately, it is not possible — to the best of our knowledge, this is the first work that tackles automatic diagnosis of Chagas directly from the ECG. Therefore, this study serves as a first baseline that opens a new line of work for further improvements.

Our findings are particularly valuable under the scantiness of validated strategies to detect ChD patients in endemic regions. Current recommendations for screening include all patients who were born in or have lived for an extended period in ChD endemic zones [44], which can be challenging, especially in endemic countries, since it can encompass the whole population of a region. A risk score was developed specifically to answer the question, “Does my patient have chronic Chagas disease?” but it seems to have limited practical value since it includes 13 variables obtained from clinical and epidemiological history and from a conventionally analysed 12-lead ECG [52]. It implies that the best approach would merge conventional and non-conventional methods [53], including the use of rapid point-of-care serological tests [54].

A clinical study would be particularly valuable, as the performance of the model could be evaluated directly by clinicians and patients. At this stage, we foresee our model as a pre-selection method of patients for further screening of the serological status. It is important to underline that more available data will enable improvements of the model that can be adapted into its daily clinical practice. We

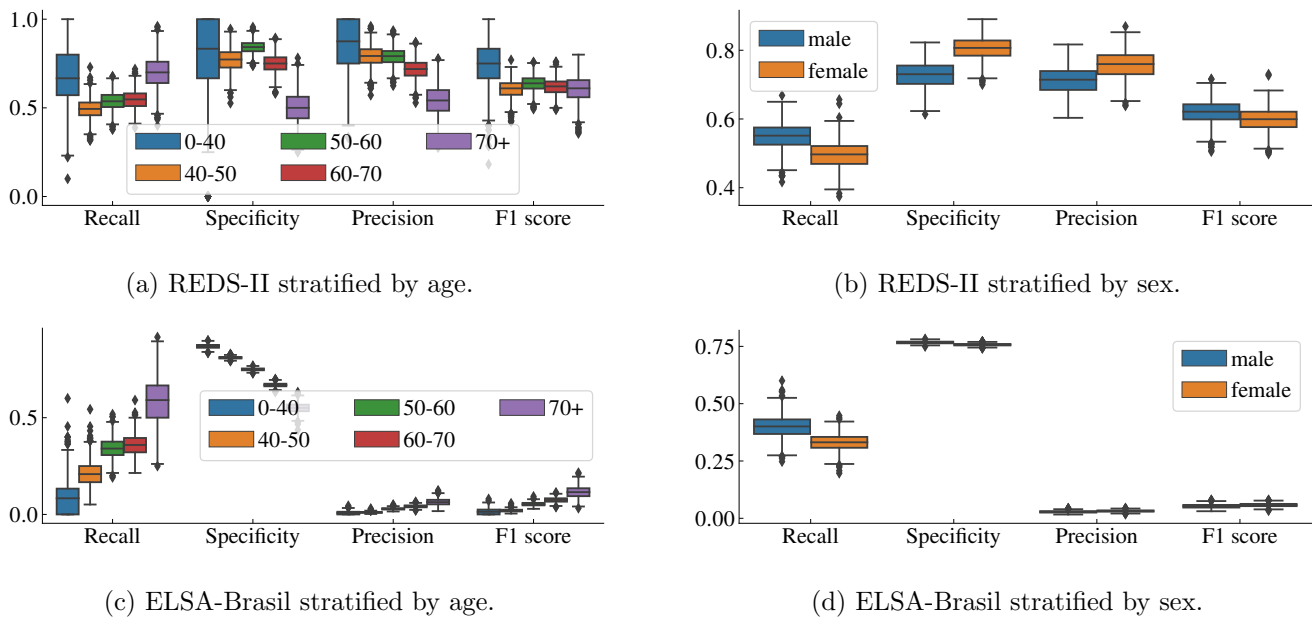


Figure 5: **Results stratified by subgroup.** Box plots of the model performance on the REDS-II (top row) and ELSA-Brasil (bottom row) test sets stratified by age (left column) and sex (right column). The box plots give the performance on 1000 bootstrapped samples.

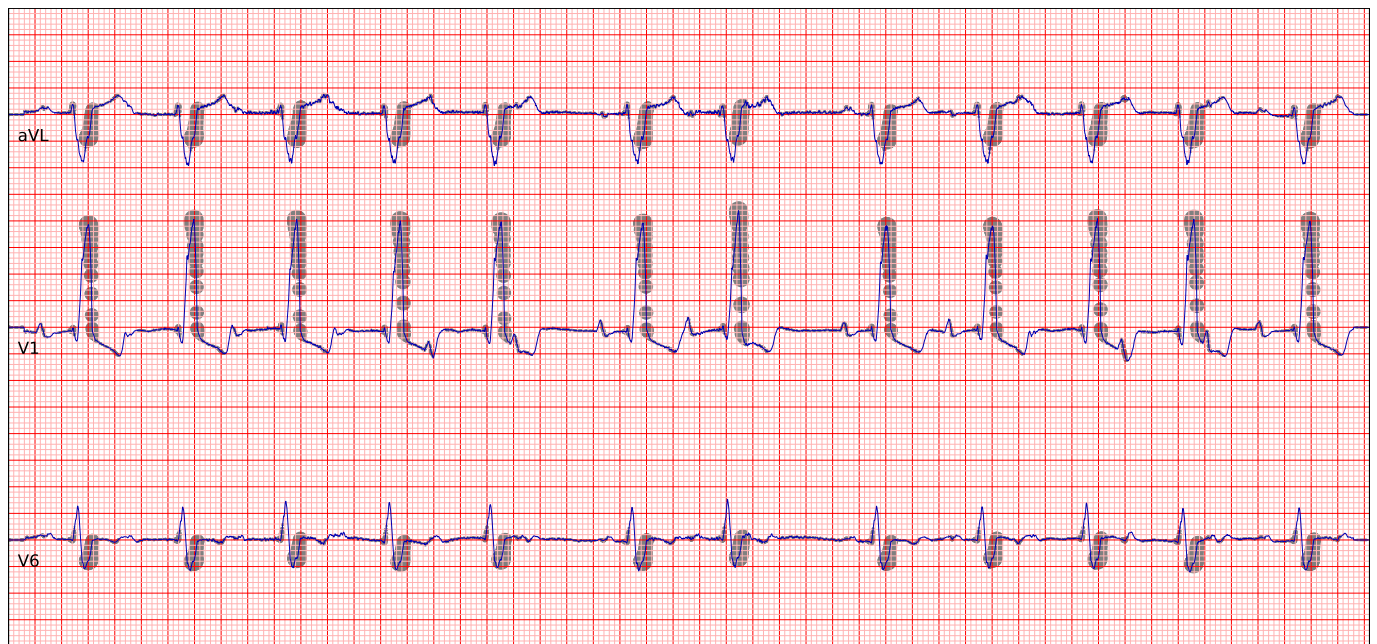


Figure 6: **Grad-CAM analysis.** Grad-CAM plot for a patient with CCC from the ELSA-Brasil data set, correctly classified by the model as Chagas positive. This plot includes three representative leads (top to bottom: aVL, V1 and V6). The shading indicates regions that the model assigns particular importance for its prediction.

hope that a future study will evaluate the clinical relevance of our model to improve the early diagnosis of ChD.

Contributors

AHR, CJ, ALPR were responsible for the study design. ALPR, TBS, AHR conceived the project and acted as the project leader. CJ, DG, AHR choose the neural network architecture, implemented, and tuned the deep neural network. CJ, DG did all the statistical tests. ALPR, CSC, CLO interpreted the

results and provided clinical interpretation. CJ, AHR, DG were responsible for preprocessing the training data. ECS, CSC, CLO, AMF were responsible for cohort design and management, data acquisition, follow-up, and ECG exams in SaMi-Trop and REDS-II cohorts. LG and SMB were responsible for cohort design and management, data acquisition, follow-up, and ECG exams in ELSA cohort. ALPR and AHR were responsible for management of the CODE cohort. AHR, CJ, ALPR, DG, TBS contributed to the writing and all authors revised it critically for important intellectual content. All authors read and approved

the submitted manuscript.

Competing interests

There are no competing interests.

Data sharing

SaMi-Trop cohort was made openly available (<https://doi.org/10.5281/zenodo.4905618>). The CODE-15% cohort was also made openly available (<https://doi.org/10.5281/zenodo.4916206>). The data sets contain information about Chagas condition mortality, age, sex, the ECG tracings, and the flag indicating whether the ECG tracing is normal. The DNN model parameters that give the results presented in this paper are also available (<https://doi.org/10.5281/zenodo.7371623>). This should allow the reader to partially reproduce the results presented in the paper. Restrictions apply to additional clinical information on the CODE-15% and SaMi-Trop cohorts; to the full CODE cohort; to the REDS-II dataset and, to the ELSA-Brasil cohort. Researchers affiliated to educational or research institutions might make requests to access the data sets. Requests should be made to the corresponding author of this paper. They will be forwarded and considered on an individual basis by the Telehealth Network of Minas Gerais, SaMi-Trop and ELSA-Brasil Steering Committee. An estimate for the time needed for data access requests to be evaluated is three months. If approved, any data use will be restricted to non-commercial research purposes. The data will only be made available on the execution of appropriate data use agreements.

Code availability

The code for the model training, evaluation and statistical analysis is available at the GitHub repository <https://github.com/carji475/ecg-chagas>.

References

- [1] M. Nunes, A. Beaton, H. Acquatella, *et al.*, “Chagas cardiomyopathy: An update of current clinical knowledge and management: A scientific statement from the American Heart Association,” *Circulation*, vol. 138, e169–e209, 2018.
- [2] C. Bern, “Chagas’ Disease,” *The New England Journal of Medicine*, vol. 373, no. 5, pp. 456–466, 2015.
- [3] M. C. P. Nunes, W. Dones, C. A. Morillo, J. J. Encina, A. L. Ribeiro, and Council on Chagas Disease of the Interamerican Society of Cardiology, “Chagas disease: An overview of clinical and epidemiological aspects,” *Journal of the American College of Cardiology*, vol. 62, no. 9, pp. 767–776, 2013.
- [4] E. C. Sabino, A. L. Ribeiro, V. M. Salemi, *et al.*, “Ten-Year Incidence of Chagas Cardiomyopathy Among Asymptomatic Trypanosoma cruzi-Seropositive Former Blood Donors,” *Circulation*, vol. 127, no. 10, pp. 1105–1115, 2013.
- [5] M. C. P. Nunes, L. F. Buss, J. L. P. Silva, *et al.*, “Incidence and predictors of progression to chagas cardiomyopathy: Long-term follow-up of trypanosoma cruzi-seropositive individuals,” *Circulation*, vol. 144, no. 19, pp. 1553–1566, 2021.
- [6] A. L. Basquiera, A. Sembaj, A. M. Aguerri, M. Omelianiuk, S. Guzmán, J. Moreno Barral, T. F. Caeiro, R. J. Madoery, and O. A. Salomone, “Risk progression to chronic chagas cardiomyopathy: Influence of male sex and of parasitaemia detected by polymerase chain reaction,” *Heart*, vol. 89, no. 10, pp. 1186–1190, 2003.
- [7] R. Viotti, C. Vigliano, B. Lococo, G. Bertocchi, M. Petti, M. G. Alvarez, M. Postan, and A. Armenti, “Long-term cardiac outcomes of treating chronic chagas disease with benznidazole versus no treatment,” *Annals of Internal Medicine*, vol. 144, no. 10, pp. 724–734, 2006.
- [8] C. S. Cardoso, A. L. P. Ribeiro, C. D. L. Oliveira, *et al.*, “Beneficial effects of benznidazole in chagas disease: Nih sami-trop cohort study,” *PLOS Neglected Tropical Diseases*, vol. 12, no. 11, pp. 1–12, 2018.
- [9] C. A. Morillo, J. A. Marin-Neto, A. Avezum, *et al.*, “Randomized trial of benznidazole for chronic chagas’ cardiomyopathy,” *New England Journal of Medicine*, vol. 373, no. 14, pp. 1295–1306, 2015.
- [10] L. E. Echeverría, R. Marcus, G. Novick, *et al.*, “Whf iasc roadmap on chagas disease,” *Global Heart*, vol. 15, 2020.
- [11] A. F. Miranda-Arboleda, E. J. Zaidel, R. Marcus, M. J. Pinazo, L. E. Echeverría, C. Saldarriaga, a. Sosa Liprandi, A. Baranchuk, on behalf of the Neglected Tropical Diseases, and other Infectious Diseases affecting the Heart (NET-Heart) project, “Roadblocks in chagas disease care in endemic and nonendemic countries: Argentina, colombia, spain, and the united states. the net-heart project,” *PLOS Neglected Tropical Diseases*, vol. 15, no. 12, pp. 1–12, 2022.
- [12] R. F. Damasceno, E. C. Sabino, A. M. Ferreira, A. L. P. Ribeiro, H. F. Moreira, T. E. C. Prates, C. A. Sampaio, and D. S. Haikal, “Challenges in the care of patients with chagas disease in the brazilian public health system: A qualitative study with primary health care doctors,” *PLOS Neglected Tropical Diseases*, vol. 14, no. 11, pp. 1–13, 2020.
- [13] *World chagas disease day 2022 - finding and reporting every case to defeat chagas disease*, <https://www.who.int/news-room/events/detail/2022/04/14/default-calendar/world-chagas-disease-day-2022---finding-and-reporting-every-case-to-defeat-chagas-disease>, 2022.
- [14] M. B. Alkmim, C. B. G. Silva, R. M. Figueira, D. V. V. Santos, L. B. Ribeiro, M. C. da Paixão, M. S. Marcolino, J. C. Paiva, and A. L. Ribeiro, “Brazilian national service of telediagnosis in electrocardiography,” *Studies in health technology and informatics*, vol. 264, 1635–1636, 2019.

- [15] P. W. Macfarlane and J. Kennedy, "Automated ecg interpretation—a brief history from high expectations to deepest networks," *Hearts*, vol. 2, no. 4, pp. 433–448, 2021.
- [16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [17] A. H. Ribeiro, M. H. Ribeiro, G. M. M. Paixão, *et al.*, "Automatic diagnosis of the 12-lead ecg using a deep neural network," *Nature communications*, vol. 1760, no. 11, 2020.
- [18] K. C. Siontis, P. A. Noseworthy, Z. I. Attia, and P. A. Friedman, "Artificial intelligence-enhanced electrocardiography in cardiovascular disease management," *Nat Rev Cardiol*, vol. 18, no. 7, pp. 465–478, 2021.
- [19] S. Gustafsson, D. Gedon, E. Lampa, A. H. Ribeiro, M. J. Holzmann, T. B. Schön, and J. Sundström, "Development and validation of deep learning ECG-based prediction of myocardial infarction in emergency department patients," *Scientific Reports*, vol. 12, no. 1, 2022.
- [20] Z. I. Attia, P. A. Noseworthy, F. Lopez-Jimenez, *et al.*, "An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: A retrospective analysis of outcome prediction," *The Lancet*, 2019.
- [21] S. Biton, S. Gendelman, A. H. Ribeiro, G. Miana, C. Moreira, A. L. P. Ribeiro, and J. A. Behar, "Atrial fibrillation risk prediction from the 12-lead ECG using digital biomarkers and deep representation learning," *European Heart Journal - Digital Health*, 2021.
- [22] Z. I. Attia, S. Kapa, F. Lopez-Jimenez, *et al.*, "Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram," *Nature Medicine*, vol. 25, no. 1, pp. 70–74, 2019.
- [23] S. Raghunath, A. E. Ulloa Cerna, L. Jing, *et al.*, "Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network," *Nat Med*, vol. 26, no. 6, pp. 886–891, 2020.
- [24] S. Raghunath, A. E. Ulloa Cerna, L. Jing, *et al.*, "Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network," *Nature Medicine*, 2020.
- [25] E. M. Lima, A. H. Ribeiro, G. M. Paixão, *et al.*, "Deep neural network estimated electrocardiographic age as a mortality predictor," *Nature Communications*, vol. 12, 2021.
- [26] C. S. Cardoso, E. C. Sabino, C. D. L. Oliveira, *et al.*, "Longitudinal study of patients with chronic chagas cardiomyopathy in brazil (sami-trop project): A cohort profile," *BMJ Open*, vol. 6, no. 5, 2016.
- [27] A. L. P. Ribeiro, A. H. Ribeiro, G. M. Paixao, *et al.*, *Sami-Trop: 12-lead ECG traces with age and mortality annotations*, version 1.0.0, 2021.
- [28] M. Alkmim, R. Figueira, M. Marcolino, *et al.*, "Improving patient access to specialized health care: The telehealth network of minas gerais, brazil," *Bulletin of the World Health Organization*, vol. 90, pp. 373–8, 2012.
- [29] A. L. P. Ribeiro, G. M. Paixão, P. R. Gomes, *et al.*, "Tele-electrocardiography and bigdata: The code (clinical outcomes in digital electrocardiography) study," *Journal of Electrocardiology*, vol. 57, S75–S78, 2019.
- [30] M. C. P. Nunes, L. F. Buss, J. L. P. Silva, *et al.*, "Incidence and Predictors of Progression to Chagas Cardiomyopathy: Long-Term Follow-Up of Trypanosoma cruzi-Seropositive Individuals," *Circulation*, vol. 144, no. 19, pp. 1553–1566, 2021.
- [31] E. M. L. Aquino, S. M. Barreto, I. M. Bensenor, *et al.*, "Brazilian longitudinal study of adult health (ELSA-Brasil): Objectives and design," *American Journal of Epidemiology*, vol. 175, no. 4, pp. 315–324, 2012.
- [32] P. R. Gomes, G. M. Paix, E. M. Lima, M. S. Marcolino, L. B. Ribeiro, G. Chequer, M. B. Alkmim, P. W. Macfarlane, and A. L. Ribeiro, "Electrocardiogram report system: The importance of decision-making tools," *Journal of Electrocardiology*, vol. 69, p. 87, 2021.
- [33] P. Denes, "Major and minor ECG abnormalities in asymptomatic women and risk of cardiovascular events and mortality," *JAMA*, vol. 297, no. 9, p. 978, 2007.
- [34] A. H. Ribeiro, G. M. Paixao, E. M. Lima, M. Horta Ribeiro, M. M. Pinto Filho, P. R. Gomes, D. M. Oliveira, W. Meira Jr, T. B. Schön, and A. L. P. Ribeiro, *CODE-15%: a large scale annotated dataset of 12-lead ECGs*, version 1.0.0, 2021.
- [35] C. S. Cardoso, E. C. Sabino, C. D. L. Oliveira, *et al.*, "Longitudinal study of patients with chronic Chagas cardiomyopathy in Brazil (SaMi-Trop project): A cohort profile," *BMJ Open*, vol. 6, no. 5, e011181, 2016.
- [36] B. A. Resende, A. M. Beleigoli, A. L. P. Ribeiro, B. Duncan, M. I. Schmidt, J. G. Mill, A. C. Goulart, A. da Costa Pereira, S. M. Barreto, and M. D. F. H. S. Diniz, "Chagas disease is not associated with diabetes, metabolic syndrome, insulin resistance and beta cell dysfunction at baseline of brazilian longitudinal study of adult health (ELSA-brasil)," *Parasitology International*, vol. 85, p. 102440, 2021.
- [37] M. M. Pinto-Filho, L. C. Brant, M. Foppa, *et al.*, "Major electrocardiographic abnormalities according to the minnesota coding system among brazilian adults (from the ELSA-brasil cohort study)," *The American Journal of Cardiology*, vol. 119, no. 12, pp. 2081–2087, 2017.

- [38] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *NIPS-W*, 2017.
- [39] E. M. Lima, A. H. Ribeiro, G. M. M. Paixão, *et al.*, "Deep neural network-estimated electrocardiographic age as a mortality predictor," *Nature communications*, vol. 5117, no. 12, 2021.
- [40] A. H. Ribeiro, *Ecg-age-prediction*, <https://github.com/antonior92/ecg-age-prediction>, 2021.
- [41] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap*. CRC press, 1994.
- [42] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, 2020.
- [43] World Health Organization, "Chagas disease in Latin America: An epidemiological update based on 2010 estimates," *Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire*, vol. 90, no. 06, pp. 33–44, 2015.
- [44] M. C. P. Nunes, A. Beaton, H. Acquatella, *et al.*, "Chagas cardiomyopathy: An update of current clinical knowledge and management: A scientific statement from the american heart association," *Circulation*, vol. 138, no. 12, e169–e209, 2018.
- [45] A. L. P. Ribeiro, M. S. Marcolino, R. J. Prineas, and M. F. Lima-Costa, "Electrocardiographic abnormalities in elderly chagas disease patients: 10-year follow-up of the bambui cohort study of aging," *J. Am. Heart Assoc.*, vol. 3, no. 1, e000632, 2014.
- [46] B. O. d. F. Brito and A. L. P. Ribeiro, "Electrocardiogram in chagas disease," *Rev. Soc. Bras. Med. Trop.*, vol. 51, no. 5, pp. 570–577, 2018.
- [47] A. L. P. Ribeiro, P. S. Cavalvanti, F. Lombardi, M. d. C. P. Nunes, M. V. L. Barros, and M. O. d. C. Rocha, "Prognostic value of signal-averaged electrocardiogram in chagas disease," *J. Cardiovasc. Electrophysiol.*, vol. 19, no. 5, pp. 502–509, 2008.
- [48] M. Ghassemi, L. Oakden-Rayner, and A. L. Beam, "The false hope of current approaches to explainable artificial intelligence in health care," *Lancet Digit Health*, vol. 3, no. 11, e745–e750, 2021.
- [49] W. Meira, A. L. P. Ribeiro, D. M. Oliveira, and A. H. Ribeiro, "Contextualized interpretable machine learning for medical diagnosis," *Commun. ACM*, vol. 63, no. 11, pp. 56–58, 2020.
- [50] G. Patrini, A. Rozza, A. Krishna Menon, R. Nock, and L. Qu, "Making deep neural networks robust to label noise: A loss correction approach," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1944–1952.
- [51] M. Lukasik, S. Bhojanapalli, A. Menon, and S. Kumar, "Does label smoothing mitigate label noise?" In *International Conference on Machine Learning*, PMLR, 2020, pp. 6448–6458.
- [52] P. E. A. A. d. Brasil, S. S. Xavier, M. T. Holanda, A. M. Hasslocher-Moreno, and J. U. Braga, "Does my patient have chronic chagas disease? development and temporal validation of a diagnostic risk score," *Rev. Soc. Bras. Med. Trop.*, vol. 49, no. 3, pp. 329–340, 2016.
- [53] M. Romero, J. Postigo, D. Schneider, J.-P. Chippaux, J. A. Santalla, and L. Brutus, *Door-to-door screening as a strategy for the detection of congenital chagas disease in rural bolivia*, 2011.
- [54] L. E. Zamora, F. Palacio, D. S. Kozlowski, K. Doraivelu, C. M. Dude, D. J. Jamieson, and L. B. Hadad, "Chagas disease screening using Point-of-Care testing in an At-Risk obstetric population," *Am. J. Trop. Med. Hyg.*, vol. 104, no. 3, pp. 959–963, 2020.

Supplementary material

Table S.1: **Results on test data: no-CCC configuration.** Metrics and 95 % confidence intervals evaluated on the no-CCC configuration of the test data sets (see the text for details). The classification thresholds are 0.60 for REDS-II and 0.71 for ELSA-Brasil.

Metric \ Test data	REDS-II (no CCC)	ELSA-Brasil (no CCC)
Recall	0.33 (0.27-0.39)	0.29 (0.23-0.35)
Specificity	0.77 (0.72-0.82)	0.76 (0.75-0.77)
Precision	0.50 (0.41-0.58)	0.02 (0.02-0.03)
F1 score	0.39 (0.33-0.46)	0.04 (0.03-0.05)
AUC-ROC	0.56 (0.51-0.61)	0.56 (0.52-0.59)
Avg. prec.	0.45 (0.39-0.52)	0.02 (0.02-0.02)

Table S.2: **Results on validation data: CCC-specific training.** Equivalent to Table 2 when the training is adapted to specifically target patients with CCC. The classification thresholds are 0.51 (selected by maximising the F1 score) and 0.33 (corresponding to 90 % specificity).

Metric \ Threshold	F1 score-based: 0.51	Specificity-based: 0.33
Recall	0.72 (0.69-0.75)	0.79 (0.76-0.82)
Specificity	0.95 (0.94-0.95)	0.90 (0.89-0.91)
Precision	0.76 (0.74-0.79)	0.66 (0.63-0.68)
F1 score	0.74 (0.72-0.76)	0.72 (0.69-0.74)
AUC-ROC	0.93 (0.92-0.93)	
Avg. prec.	0.80 (0.77-0.82)	

Table S.3: **Results on test data: CCC-specific training.** Equivalent to Table 3 when the training is adapted to specifically target patients with chronic Chagas cardiomyopathy. The classification thresholds are 0.51 for REDS-II and 0.33 for ELSA-Brasil.

Metric \ Test data	REDS-II (CCC)	ELSA-Brasil (CCC)
Recall	0.70 (0.63-0.77)	0.59 (0.44-0.73)
Specificity	0.98 (0.96-0.99)	0.96 (0.96-0.96)
Precision	0.91 (0.85-0.95)	0.05 (0.03-0.07)
F1 score	0.79 (0.74-0.84)	0.09 (0.06-0.12)
AUC-ROC	0.90 (0.86-0.93)	0.91 (0.85-0.95)
Avg. prec.	0.83 (0.76-0.89)	0.10 (0.05-0.19)

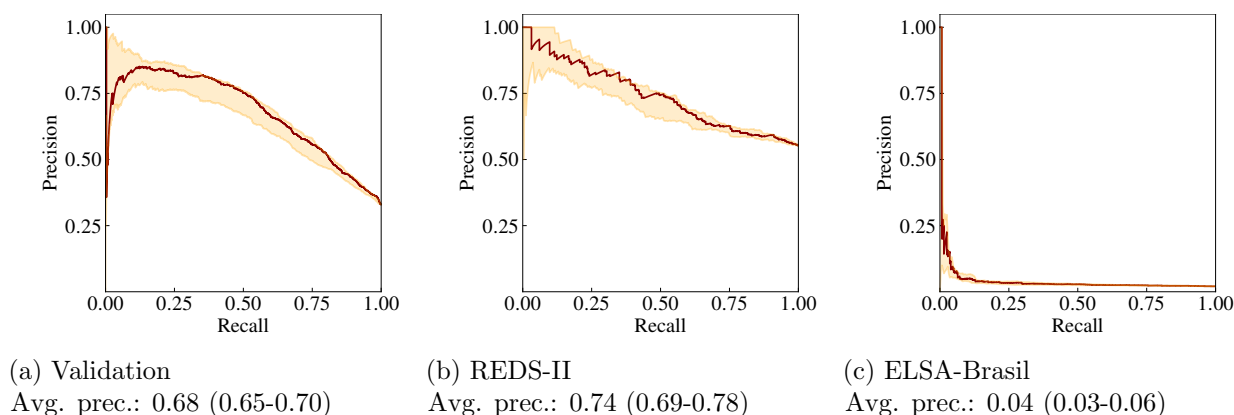


Figure S.1: **Precision-recall curves on validation and test data: ChD+CCC vs normal.** The shaded regions encapsulate the maximum and minimum values corresponding to 15 different weight initialisations — the outputs of these models are averaged to produce the output of the ensemble model, the result of which is given by the solid lines.

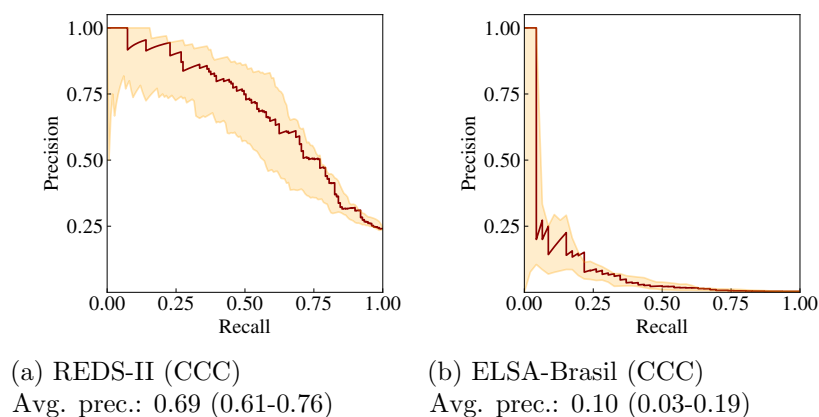


Figure S.2: **Precision-recall curves on test data: CCC vs rest.** We consider only CCC as positive and ChD as well as normal as negative here.

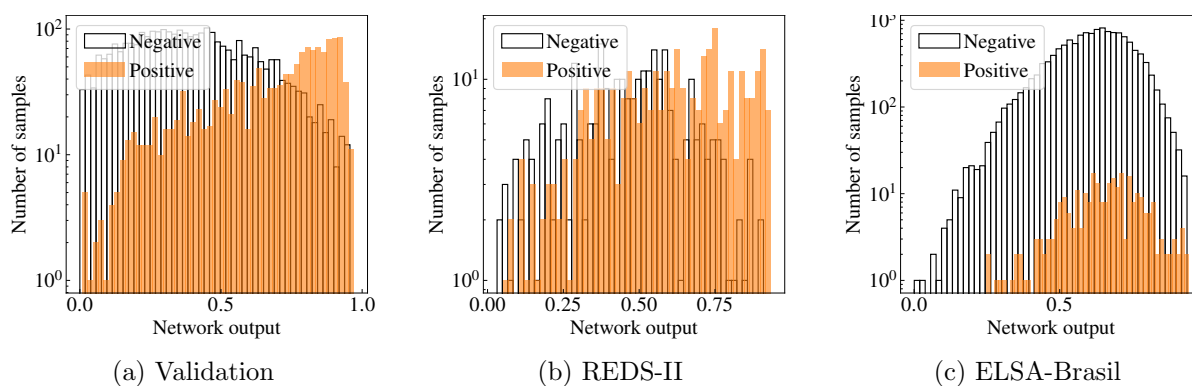


Figure S.3: **Output histograms: ChD+CCC vs normal.** Histograms are computed on the validation data and the test data. Note the logarithmic scale of the y -axis. We can see the number of false positive/negatives when applying the selected thresholds on the x -axis: 0.60 for REDS-II and 0.71 for ELSA-Brasil.

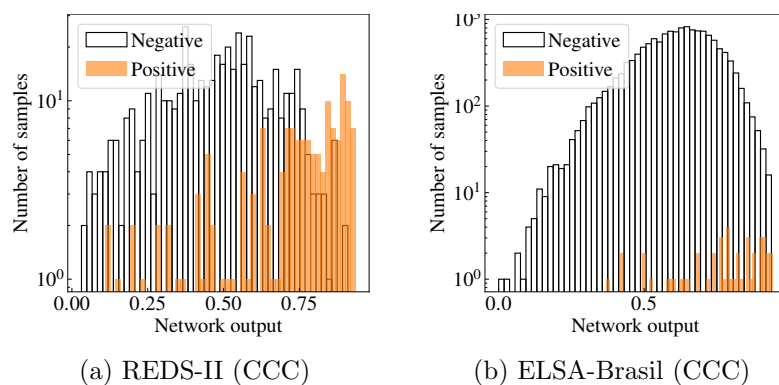


Figure S.4: **Output histograms: CCC vs rest.** Histograms are computed on the test data where we only consider CCC as positive and ChD as well as normal as negative. Note the logarithmic scale of the y -axis.

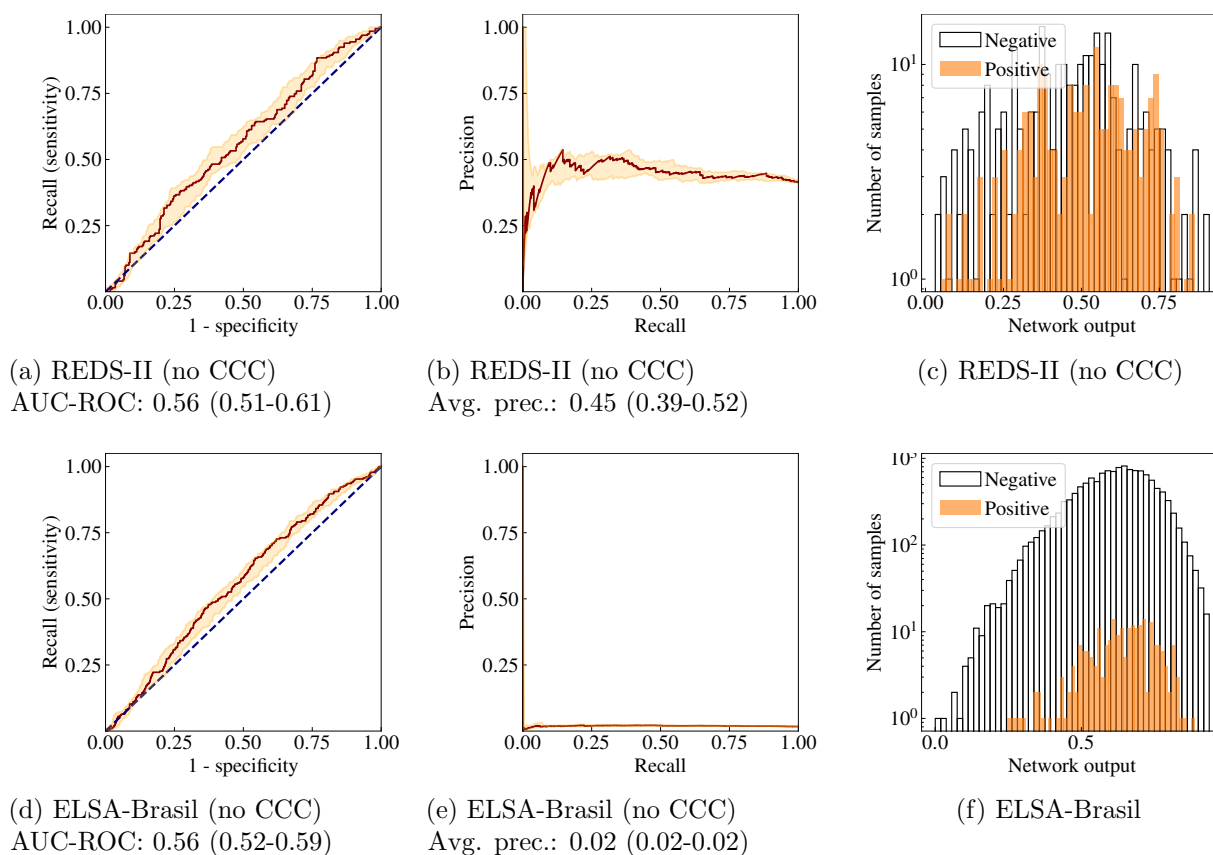


Figure S.5: **Results on test data: no-CCC configuration.** Receiver operating characteristics (left), precision-recall curves (middle) and output histograms (right) computed on the test data for the no-CCC configuration (see the text for details). This set removed the CCC cases and shows ChD vs normal.

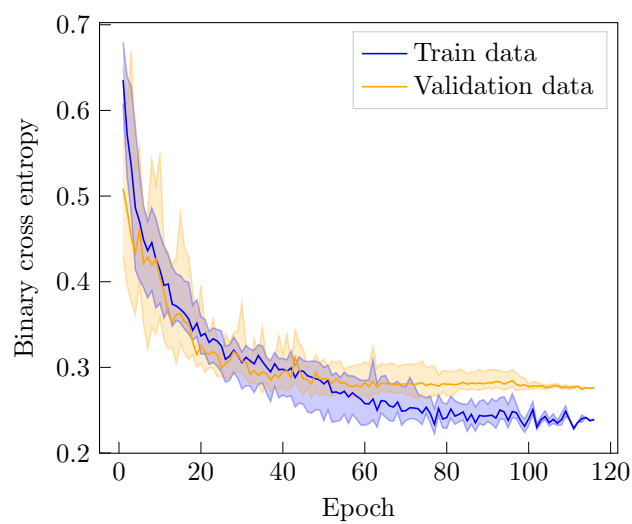


Figure S.6: **Loss function evaluation: CCC-specific training.** Equivalent to Figure 2 when the training is adapted to specifically target patients with CCC.

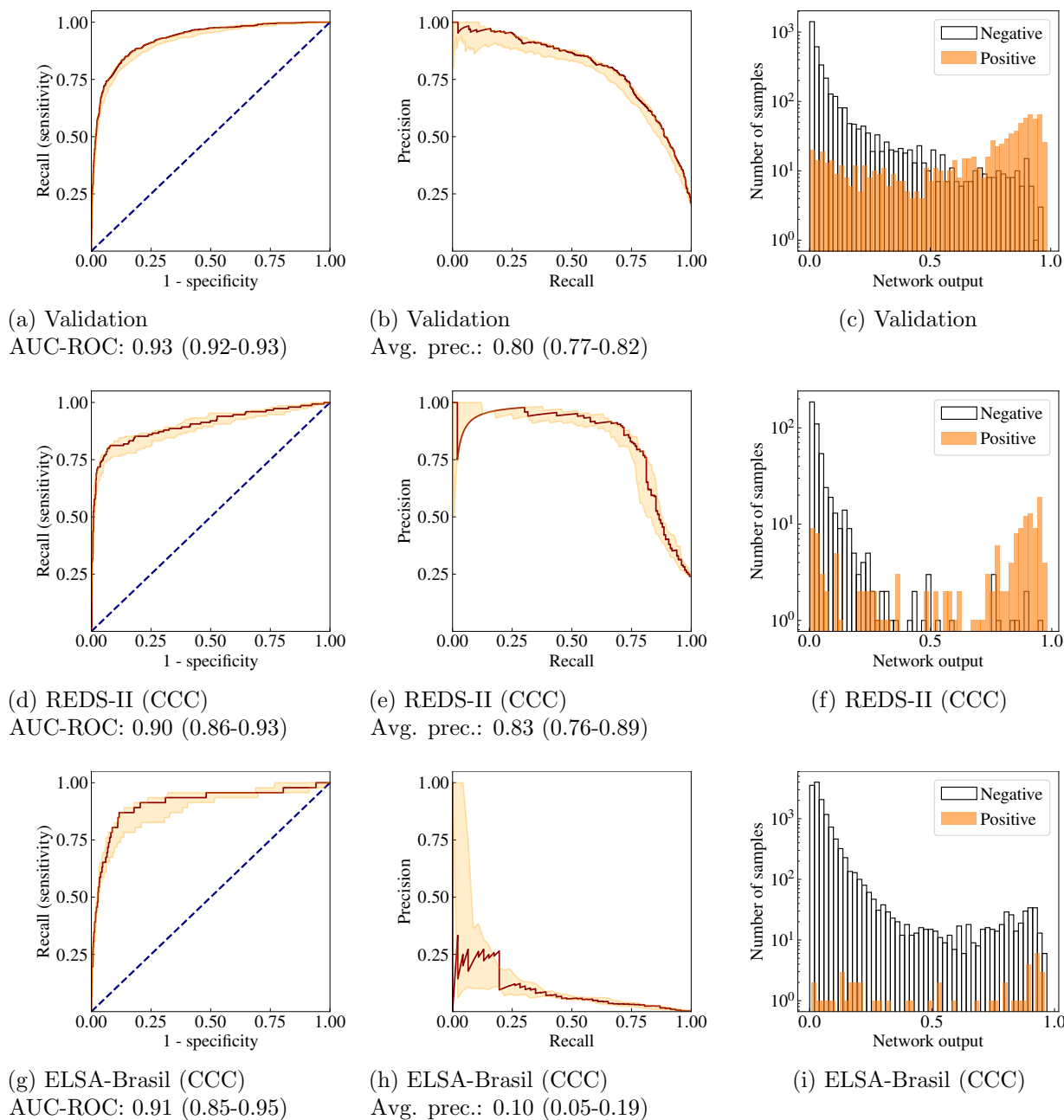
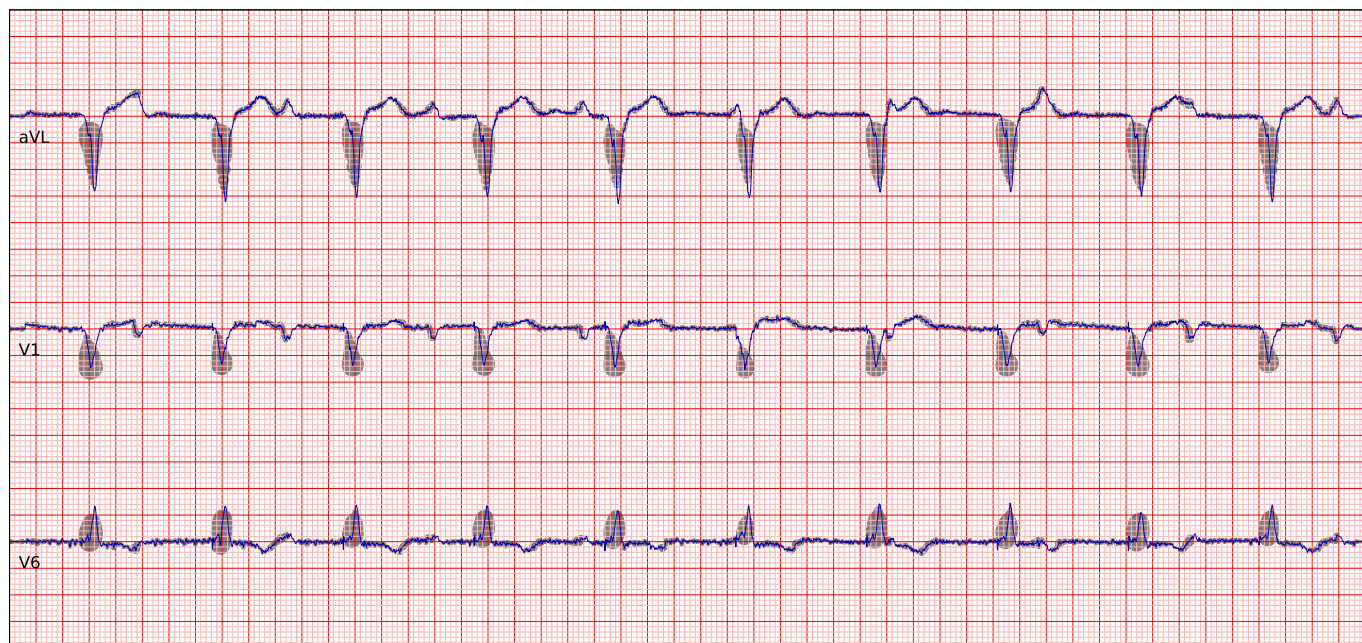
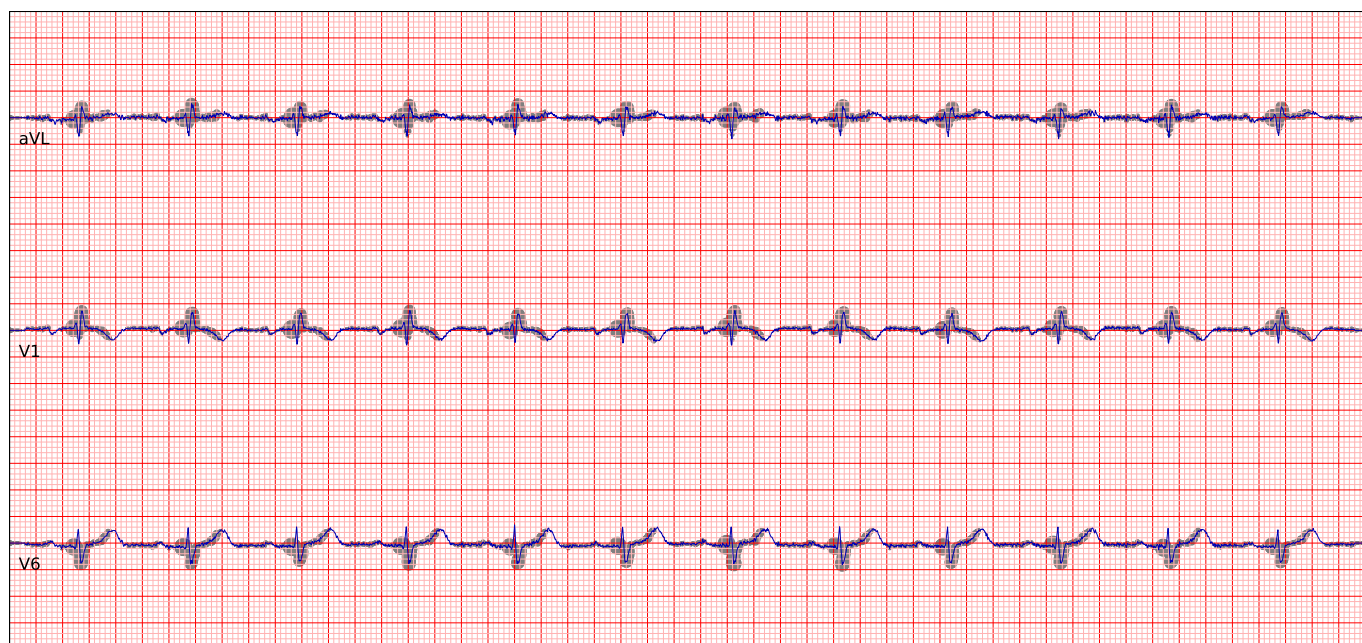


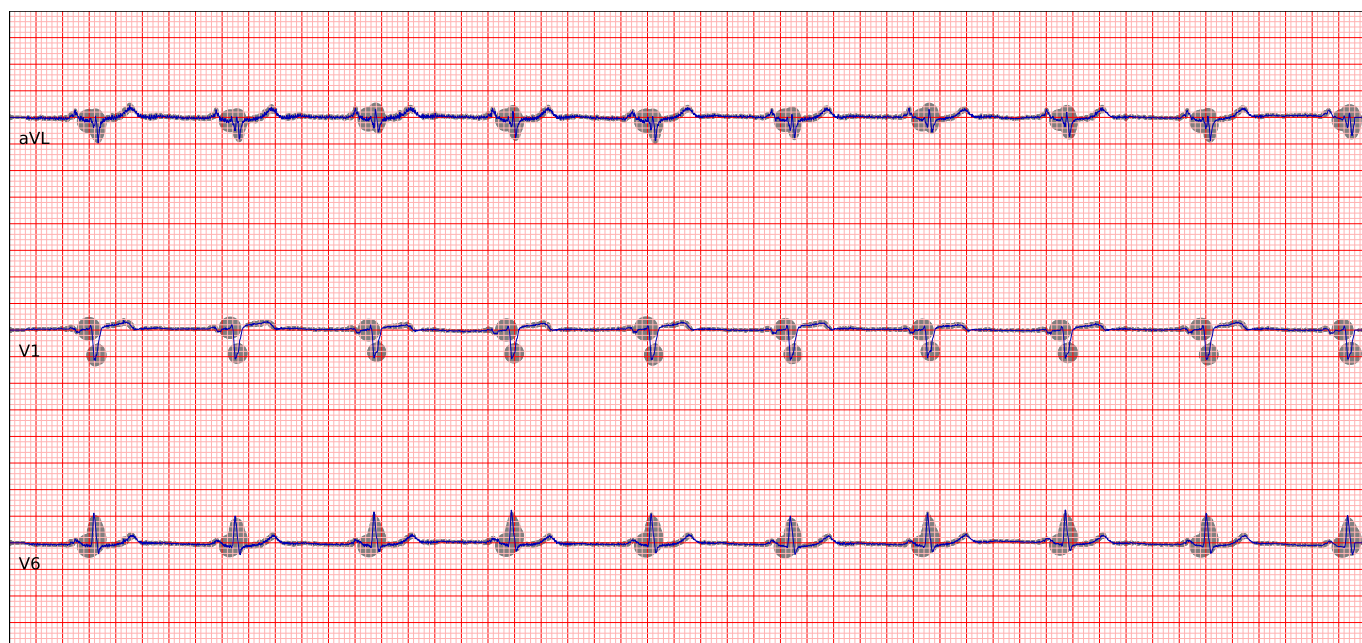
Figure S.7: **Results on test data: CCC-specific training.** Receiver operating characteristics (left), precision-recall curves (middle) and output histograms (right) computed on the validation data and the test data when the training is adapted to specifically target patients with CCC.



(a) Positive patient with CCC.



(b) Positive patient without chronic Chagas cardiomyopathy.



(c) Positive patient without chronic Chagas cardiomyopathy.

Figure S.8: **Grad-CAM analysis: additional patients.** Complementing Figure 6 with another three Grad-CAM plots for patients from the ELSA-Brasil data set, correctly classified by the model as Chagas positive. We here include one patient with chronic Chagas cardiomyopathy (a), and two without (b-c). The plots include three representative leads (top to bottom: aVL, V1 and V6). The shading indicates regions that the model assigns particular importance.