Can Deep Recurrent Convolutional Neural Networks Improve the Effectiveness of Low-lead ECG Mobile Remote Health Monitoring Devices?

Brandon Dudley and Kevinpaul Gunasingh

Abstract—Abstract Cardiovascular Diseases (CVDs) are known for causing numerous death globally and in the United States. CVDs come in the forms of hypertension, coronary heart disease, heart failure, and stroke. Overall, in the United States, the population's CVD prevalence in 2010 was 36.9 percent, it reached 38.7 percent prevalence by 2020, and CVD prevalence is projected to increase to 40.5 percent by 2030, resulting in a 9.9 percent increase in CVD prevalence for the population as a whole [1]. Also, healthcare costs in CVDs alone are projected to triple by 2030 in comparison to 2010 [1]. Looking into this topic is important in order to decrease projected CVD prevalence and CVD healthcare costs. From a global perspective, 17.3 million deaths occur because of CVDs, and the number of total CVD deaths could exceed 24 million by 2030, leading to trillions (US dollars) lost in the economy [2].
Electrocardiograms (ECGs) are devices that help detect any possible symptoms for CVD, what CVD could possibly result, which helps the doctor and/or patient figure out appropriate treatment. The best brands for ECGs include CSME, ZEPHYR, POLAR, and 3M, and the brands for ECG sensors are CMSE ECG Sensor and Alive ECG Sensor [3]. After doing tests with each ECG and sensor, CMSE with the Alive ECG sensor has the lowest error rate, but CMSE with the CMSE ECG sensor has the highest error rate, and the tests were taken when the patient was at rest, walking, or jogging [3].

Index Terms—Deep Learning, Machine Learning, Cardiovascular Health, Accurate Predictions

1 Introduction

Cardiovascular Diseases (CVDs) are of great concern due to its high mortality rates, increased costs in global healthcare, and reduction of the quality of life [4]. CVDs are responsible for 17.9 million deaths worldwide annually [5]. In the recent past, healthcare organizations have looked to machine learning to enhance electrocardiograms (ECGs). ECGs can use enhancement from machine learning in order to reduce the margin of error in ECG detection, computation of scores based on ECG scans, differentiating CVD phenotypes, quantifying heart function and segmentation, and making early diagnosis of disease [6]. Due to data increase in clinical practices, machine learning can enhance cardiologist's workflow and make patient care more efficient [6]. Machine learning is useful in cardiology regarding imaging, electrocardiograms (ECGs), and risk predictions [6]. Here, we seek to present the objective of our study, prior knowledge of the topic before we started the research, and how our findings could potentially advance new knowledge or understanding of the topic. Additionally, our goals are to present the relevance between prior knowledge and current knowledge, address deficiencies in current literature related to our topic, and note potential contributions of the study.

One use for combining machine learning and cardiology is imaging. Imaging is used to scan a person's body, and the doctor tells that person whether he/she has a serious condition, and the doctor recommends appropriate treatment for the condition or disease. In cardiology, imaging is known as echocardiography. Machine learning can be used in echocardiography by applying neural network classifiers to detect patterns of heart failure based on radio graphs (echocardiographic images) [6]. Deep learning (DL), a branch of machine learning (ML), is useful in medical imaging, and unsupervised deep learning led to generative visual rationales (GVRs), and GVRs display visual features to classify a scanned image for the disease [7]. Aortic stenosis is a common heart disease for countries where the average life expectancy is high [8], and using echo cardiographs to detect aortic stenosis requires multiple metrics to check whether aortic stenosis exists, but artificial intelligence could receive metrics from echo cardiographic data [9]. ML can be implemented to enhance cardiac phenotype recognition by implementing features of cardiac tissue that show degradation [10]. Data was collected from 77 patients who had athletic physiological hypertrophy (ATH) and 63 patients who had hypertrophic cardiomyopathy (HCM), 3 ML algorithms were used on the data (artificial neural networks, support vector machines, and random forests) [10]. After conducting the test, it is safe to assume that machine learning algorithms can enhance interpretation of echo cardiographic images [10]. One of the drawbacks to electrocardiography is that human interpretation has limited its potential, so one of the challenges is to see whether computer assisted interpretation can recognize echocardiographic results (photos or videos) [11]. DL combined with convolutional neural networks (CNN) has

B. Dudley and K. Gunasingh are with the College of Computing and Software Engineering, Kennesaw State University, Marietta, GA, 30060 USA. Emails: kgunasin@students.kennesaw.edu and wdudley@students.kennesaw.edu

been proven successful in detecting anomalies in images, but DL with CNN has not been implemented in cardiology yet [11]. A method to solve this problem is to randomly select echo cardiograms, seek to achieve high accuracy on view classification, and look into the model to show that classification depends on features in the images [11]. The result was that the clustering analyses show that the neural network can sort input images based on features in the image [11]. Imaging and electrocardiograms are important in cardiology as imaging shows pictures of the cardiovascular region, but electrocardiograms detect wave patterns of the heart.

Electrocardiograms are useful for assessing cardiac problems [12]. In that experiment, a deep learning framework was trained on an image set, and it was transferred to carry out ECG diagnostics by taking into account the patient's current cardiac conditions, and these networks can be used for extracting features to find the final result [12]. After implementing the deep learning technique to classify between 3 cardiac conditions, the testing accuracy came to 92 percent and highest prediction was 98 percent, so it means that deep learning is an efficient cardiac anomaly detection mechanism, along with an easy application technique [12]. A cardiac ailment known as asymptomatic left ventricular dysfunction (ALVD) affects 5 percent of the population, but no cheap ALVD detection tools exist [13]. Potential AI application to ECGs could identify ALVD, and using 12-lead ECG and echo cardiogram data showed 85.7 percent accuracy, and it is safe to conclude that AI combined with ECGs can help detect ALVD [13]. Some patients who suffer cardiac ailments could have chronic kidney disease (CKD), leading to hyperkalemia, and those affected with that are often times asymptomatic, so implementing a deep learning model in an ECG can improve detection of hyperkalemia [14]. The experiment consisted of using either 2-lead or 4-lead ECGs, and the 2-lead ECGs had an accuracy rate of 86 percent, so incorporating AI to ECGs can improve hyperkalemia detection [14]. Electrocardiograms are important in detecting what heart disease a person may have, but they can detect whether a person is at risk.

Machine learning and cardiology is important for detecting whether a person is at risk of a cardiac disease. A problem of traditional cardiac detecting systems is that they have low sensitivity and high chance of false alarms [15]. A proposed method of invoking deep learning with ECGs could improve risk detection, and the study started with collecting data of patients at 2 hospitals from 2010 to 2017, and the test showed that logistic regression reduced false alarms by 42.1 percent [15]. This means that an ECG with deep learning has a higher sensitivity and lower false alarm rate [15]. Predicting length of hospital stay is important to keep the number of hospital beds taken and healthcare costs in control [16]. The participants in the study were adult patients admitted into King Abdulaziz Medical City Complex in Riyadh, Saudi Arabia, and the random forest showed an 80 percent accuracy, so we can conclude that machine learning models are useful in predicting possible cardiac ailment symptoms [16]. Currently, it is unknown whether ML models can address relationships with high variance variables, so one way to solve that problem is to compare random forests with support vector machine, and the results show that random forests show a 17.8 percent improvement compared to logistic regression, so it is safe to assume that ML models improved hospital readmission prediction, so machine learning models can be integrated into ECGs to predict hospital readmission [17].

Machine learning in cardiology is important because machine learning models can accurately detect what heart disease a person is diagnosed with based on pictures of scans and wave forms, and prediction of cardiac symptoms is also important. Prior to the study, I knew that machine learning can be used to detect cancer or other diseases based on given information, but I did not think that ML models can make predictions with wave forms or pictures of scans. However, I realized that combining machine learning with deep learning can help predict symptoms of cardiac diseases ,specific cardiac diseases, or treatment for that disease. I believe that in the future, this study would convince more healthcare organizations to adopt machine learning models into their electrocardiograms and echo cardiographs to make accurate predictions for patients. Some drawbacks to the study is that human interpretation has limited the potential of the ML models, and not too many researchers have looked deep into the topic, but they have proposed methods as a start, and those methods have decent accuracies.

2 Related works

To generalize, ECG classification is an important topic in machine learning, ML research is emphasized with 12-lead ECG data, leaving gaps in single-lead research, and neural networks that work both wide and deep (Long-Short Term Memories and Recurrent Neural Networks) give the best outputs. For in-office testing, ECGs with 10 electrodes and 12 leads are commonly used, and ECGs with 1 or 2 electrodes with 1 lead are used in remote ECGs. Remote ECGs are promising, but having fewer leads makes them inferior, so implementing machine learning can make single-lead remote ECGs more reliable, and going to the doctor's office would seem redundant.

- Early detection of silent ischaemic heart disease by 24-hour electrocardiographic monitoring of active subjects: Shlomo Stern and Dan Tzivoni [18] took electrocardiogram recordings that were 24 hour nonstop on 84 patients with precordial symptoms. They conducted the experiment on 43 men from 41 to 76 years of age and 41 women from 41 to 75 years of age, and all these test subjects have praecordial symptoms, normal test result for 12-lead test, and negative from 2-step master test. The 2 results were classified as a negative 24-hour reading and a positive 24-hour reading. A negative 24-hour reading is where the patients had chest pain, and a follow-up test after 6 months shows improvement on the patient's condition. A positive 24-hour reading is where the patients experience heart problems, and the conditions got worse after the 6 month follow-up test. What the tests show is that through continuous 24 hour recording is that everyday stress and physiological conditions cause fluctuations in the ECG recordings. However, a weakness I noticed is that the data set was biased. The negative test result group consisted mostly of women, and the positive test result group consisted mostly of men. This is biased because

men have a higher incidence rate of getting arteriosclerotic coronary disease. In my opinion, I would

- Deep learning and the electrocardiogram: Review of the current state-of-the-art: Sulaiman Somani et. al [19] talk about how deep learning gained popularity in the recent past and its use in healthcare. Deep learning is used in cardiovascular aspects, and some uses of deep learning regarding the heart is to develop an easy-to-understand ECG analysis for clinicians and clinical perspectives on cardiology's potential advances.

- Analysing and improving the diagnosis of ischaemic heart disease with machine learning: Ischaemic heart disease [20] has one of the highest death rates, and improvements to treat the disease are crucial. The four diagnostic levels of CVD testing include resting stage, sequential during exercise, blood examination, and checking blood vessels to the heart.

- Machine Learning for Cardiovascular Disease Improves When Social, Environmental Factors Are Included: Rachel Harrison of NYU [21] talks about how machine learning can make accurate predictions regarding CVD, but incorporating social factors can make better predictions for groups of different ethnic groups. CVD is responsible for numerous deaths worldwide, but social determinants are incorporated into CVD causes, and the social determinants can influence diet and exercise. Rumi Chunara discovered that CVD is increasing in middle to low income countries, and US communities that consist of minorities have seen tremendous increase in CVD deaths. Chunara believes that since CVD deaths are rising in these groups over a short period of time, so it's predictable that social and environmental issues could be the reason why. Healthcare organizations have incorporated machine learning (ML) into cardiology (cardiovascular research) for detecting data patterns to find out any underlying disease symptoms exist, statistics are essential for identifying CVD risks and treatments, and accurate models help notify risks and potential treatments. Usually, CVD risk is measured with blood pressure and cholesterol levels, but healthcare workers seldom consider social factors. Chunara aims to incorporate social factors into machine learning models, but one problem is that social factors have unusual and non-linear patterns with CVD prediction.

- Applications of Machine Learning in Cardiac Electrophysiology: Rahul Muthalaly and Robert Evans [22] discuss how machine learning algorithms can help patients in different stages of patient care in electrophysiology, and that machine learning can improve accuracy and efficiency, but its use is fairly new. Cardiology contains big data and needs to address and diagnose issues to mitigate CVD, and ML can handle big data, make accurate predictions regarding CVD risks, and recommend appropriate treatments. The big data aspect is nothing but a large data set, and heath care organizations generate high amounts of data. Big data has 3 V's, and the 3 V's are volume (higher volume of data is better), velocity (data generation speed), and variety (types and features in data). Machine learning can be applied in cardiology (especially electrocardiology) in surface electrocardiography, intracardiac mapping, and cardiac implantable electronic devices. Surface electrocardiography (Surface ECG) can help concise, detailed, and affordable information, and that helps public health, along with available wearable technology to improve raw data quality, but the validity must be verified before jumping to conclusions.

3 Proposed method

The main purpose of this study is to show how reliable single-lead data is for serious cardiac condition diagnosis by comparing it to superior, more accurate, and precise twelve-lead data, and if there are better ways of obtaining the data. It's without question that single-lead can truly never trump the performance of twelve-lead, but could it at least be sufficient to outweigh the shortcomings of twelve-lead devices in the ways stated prior? We hope to answer the following questions:

- Are single-lead ECG recordings sufficient for diagnosing serious CVDs?
- Can twelve-lead data help give us insight into which lead is better for CVD detection?.
- How well do models designed for single-lead data perform when trained using twelve-lead data?
- Can models trained on single-leads work produce correct ouputs for twelve-lead testing data?

As also stated earlier, twelve-lead ECG remains to be the most common and most accurate methods for detecting and diagnosing CVDs. As expected, this means that the amount of data that exists for twelve-lead far exceeds that of single-lead datasets. In order to test any hypothesis between two drastically different datasets, we will have to consider some important steps to eliminate bias and ensure that our results can be validated. Before we discuss our experimental design, reasoning, and processes, we will first discuss the datasets we used as well as some information that attributes to our baseline methodology. Dataset 1 – PhysioNet, CNC 2017 - AF Classification from a Short Single Lead ECG Recording

TABLE 1: The number of recordings, data mean, standard deviation, max value, median value, min value, and the sample distribution

Type	Records	Mean	SD	Max	Median	Min
Normal	5154	31.9	10.0	61.0	30	9.0
AF	771	31.6	12.5	60	30	10.0
Other	2557	34.1	11.8	60.9	30	9.1
Noise	46	27.1	9.0	60	30	10.2
Total	8,528	32.5	10.9	61.0	30	9.0

This dataset was used in the 2017 PhysioNet challenge and contains 8,528 single-lead recordings from an AliveCor ECG device. All ECG samples in this dataset have a sample frequency of 300 Hz. However, due to the "Noise" class not existing in any twelve-lead dataset nor being represented well (aprox. 0.5Dataset 2 - PhysioNet, CNC 201 - Classification of 12-lead ECGs This dataset was used in the 2020 PhysioNet challenge and contains 8 sub datasets that total around 80,000 individual samples. We used the PTB-XL sub dataset with 21,837 samples with a sample frequency of 500 Hz. All data is muti-labeled from a set of 27 different diagnosed cardiac events from medical professionals. Due the incredible difference in the labeling between datasets, it was imperative that we homogenize them. The only two approaches we could take to achieve this would be to use

unsupervised machine learning methods to artificially label the single-lead dataset with more specific cardiac events or investigate how to reduce the number of labels from 27 to 3. The method we chose was the latter as providing more in-depth labeling of the single-lead data is incredibly difficult. In the next section, we will explain our procedure for doing this in the "Experimental Setup" section. Furthermore, to be more consistent with the best and most sensitive lead electrode placements that are used in single-lead setups, we will be only using the leads V1, V2, V3, V4, V5 and V6, also known as the horizontal plane, in the twelve-lead dataset. [23] [24] We chose these sets of data specifically for a couple reasons. The first being that PhysioNet, through the CinC competitions, have offered some of the highest-quality ECG datasets available to the public. The samples are studied and labeled by expert medical professionals and curated by experienced data scientists. Furthermore, both datasets offer quality data in regard to size, accuracy and bias from the large number of unique patients and wide demographics of patients (race, gender, etc.). However, one glaring bias is the use of a single device manufacturer for all our single-lead samples. The twelve-lead devices are numerous and have different manufactures but the specificity of that information was largely omitted.

3.1 Baseline Method

One way of looking at and understanding high lead data compared to small lead data is that high lead data is simply a large set of single leads that were taken from the same source, or patient, at the same time. The functionality of the individual leads on multi-lead machines and their corresponding electrodes do not necessarily differ much from those on small or single-lead machines [25]. We will treat the individual leads on our twelve-lead dataset as twelve different samples, each one being added to twelve new datasets that only contain samples from the same lead for the purpose of training, testing and validating a model while using those findings to give insight into creating a different model for training on data generated with single-lead devices. By cross-examining both datasets between models, we may be able to get insight into how well each model trains on both types of data and if the models can still classify samples that were not generated from the same source as the data it trained on.

The baseline model we used was a decently wide and very shallow convolutional network with very minimal layers and parameters. The first model we trained was using the 2017 Physionet dataset. We used a base model reflecting on the winner of the 2020 Physionet challenge which can be viewed in figures 1 and 2.

3.2 Proposed Model

Despite the success of the baseline model by the authors, the model performed nominally on our 12-lead dataset and extremely poor on the single-lead due to the scale of the model being too large relative to our data. To compensate for this, we scaled back the number of filters per layer by a factor of 8, the kernel width by 2, and reduced the number of convolutional layers by 3. The result was a model that was 6.5x smaller and trained well to our single-lead data. However, the model was shown to be too small for the 12-lead set so we increased the number of filters by about 33 percent and received better results.

3.3 Implementation Details

Each dataset is saved in their respective folders containing the .mat and .hea. For our implementation, we are only concerned with the .hea files. These files are able to be accessed and loaded using the WFDB Python package. For the single-lead dataset, we simply loaded each of the sample's signal into a Numpy array of size 1714. The final matrix size is 8,538 x 1,714. The same process was applied to the twelve-lead dataset except, for each sample, the leads were taken individually. For a dataset of 21,838 samples, we ended up with 262,056 total, and a matrix of 262,056 x 1714.

4 Experimental Setup

In order to evaluate our results with consistency and in tandem to the research questions stated in section 3.0, we will train the model using thirteen total datasets composing of the single-lead dataset and the twelve individual single-lead subsets from the twelve-lead dataset. To refrain from introducing potentially strong bias to the model, we are not using any trained weights from previous tests for any subsequent tests after. Each training session will be distinct such that we have a total of 6 trained models.

4.1 Research Questions

Twelve-lead ECG machines continue to be the most accepted and widely acclaimed methods for obtaining accurate cardiac information for diagnostics [26]. However, these machines are expensive, heavy, require much setup and require a training clinician to operate correctly. Consequentially, patients would also need to visit a hospital or clinic to receive an ECG which is costly and puts strain on medical infrastructure. The point of this study is to understand if mobile, low-lead ECG devices at least sufficient in producing data for accurate and life-saving diagnostics. Additionally, in order to test our hypothesis, we would also need to develop and verify current state-of-the-art methods for ECG classification and identify if these strengths can be used universally over many different datasets generated from different sources.

4.2 Dataset and Preprocessing Techniques

The two datasets that we used for this study are the 2017 Physionet and 2020 Physionet ECG databases [27] [28] Additionally, we will discuss the strategy alterations to the datasets we used in order to make their labels homogeneous.

For 2017 Physionet dataset contains around 8,500 single-lead samples between 4 classes: Normal, Atrial Fibrillation and Noise. The 2020 Physionet dataset is a multi-labled dataset that contains around 21,837 12-lead samples with anywhere from 1 to 87 total labels, each indicating a diagnosis for a very particular atrial event. However, a random-forest classifier was ran on the data to identify the 27 most important labels that indicate a serious cardiac condition

[29]. In order to homogenize both of our datasets, we decided to split the 12-lead dataset into 3 parent classes to be consistent with our single-lead dataset. Furthermore, we dropped the noise class from our single-lead dataset due to it being heavily under sampled and because there were no noise samples in the 12-leads. The following table details our relabeling scheme. Due to many of the samples having multiple labels, we applied a precedence such that: atrial fibrillation ¿ other arrhythmia ¿ normal, Many of the samples in the "Normal" label are common atrial events that are consistent with conditions related to stress, lack of sleep, malnutrition, or other short-term conditions that do not indicate serious cardiac conditions.

TABLE 2: Relabeling scheme to make the twelve-lead dataset consistent with the single-lead

Original Label	Relabel
1st degree av block	normal
atrial fib	atrial fib
atrial flutter	atrial fib
bradycardia	normal
complete right Bundle Branch Block	arrythmia
inomplete ight Bundle Branch Block	normal
left anterior fascicular block	normal
left axis deviation	normal (unless also labeled with LBBR)
left bundle branch block	normal
low qrs voltages	arrhythmia
premature atrial contraction	arrhythmia
premature ventricular contractions	arrhythmia
prolonged pr interval	normal
prolonged qt interval	arrythmia
right axis deviation	normal
right bundle branch block	arrythmia
sinus rhythm	normal
sinus tachycardia	normal
sinus bradycardia	normal
qwave abnormal	arrythmia
supraventricular premature beats	normal
t wave abnormal	normal
t wave inversion	arrythmia
ventricular premature beats	normal

The datasets are considered well-organized and the samples have been monitored by ECG clinicians for basic normalization, data representation and correctness. For both datasets, each sample has a frequency rate of 500Hz, or 500 signal readings-per-second.

Due to the readings being different in levels and equipment used, we normalized all samples in each dataset using their z-score for basic feature scaling. The z-score normalizes the sample values by using it's standard mean and mean in order to keep the magnitude (the "loudness") of the wave samples consistent throughout the recordings and dataset.

4.3 Evaluation Metrics

For our evaluation metrics, we primarily focused on our achieved F1 scores, training and validation losses and confusion matrices. The model uses a k = 5 cross-entropy validation algorithm where 20% of each fold is used as validation data while the rest is used for training. At the end of the training epoch, the best performing set of weights, as determined by which fold had the lowest validation loss, is saved and loaded for the next epoch.

5 Results and Discussion

5.1 Model Evaluation

For our initial tests, we achieved the following results as displayed in table 3 for the single-lead dataset.

TABLE 3: The results from training on the single-lead dataset

Lead	Val acc	Val loss	F1(N)	F1 (A)	F1(O)
-	0.8137	0.5692	0.5603	0.8641	0.6408

From table 3, we can observe that the single-lead data ended with a validation accuracy of 81.37%, a loss of 0.5692 and a highest F1 sore of 0.8641 for the atrial fib score.

TABLE 4: The results from training on leads V1, V2, V3m V4, and V5

Lead	Val acc	Val loss	F1(N)	F1 (A)	F1(O)
V1	0.7633	0.5381	0.6034	0.819	0.8494
V2	0.7415	0.5922	0.5558	0.7887	0.8268
V3	0.7439	0.5661	0.6396	0.7938	0.8308
V4	0.7973	0.4976	0.6556	0.8111	0.8538
V5	0.713	0.6162	0.6402	0.7514	0.832

From this table, we can inquire a few things about how the data performed and how well the relabelling process worked. The V5 lead had the highest validation loss while V4 had the lowest. V4 also had the highest F1 score for all 3 classes.

Lastly, we tested the V4-lead data on the model trained only on the single-lead dataset and we received the results as shown in table 5.

TABLE 5: The results from testing the V4-lead data on the model trained only on single-lead data.

-	Precision	Recall	F1-score	Support
Normal	0.0670	0.6228	0.1211	1514
Afib	0.4	0.0007	0.0015	7772
Other	0.5743	0.3552	0.4390	12551
Acc	0.2476	0.2476	0.2476	0.2476
Macro avg	0.3471	0.3262	0.18721	21837
Weighted	0.4771	0.2476	0.2612	21837

6 Conclusion

We will refer back to our original research questions and answer them iteratively based off our results posted in the previous section.

- Are single-lead ECG recordings sufficient for diagnosing serious CVDs?

Diagnosing critical CVDs from single-lead devices is a very difficult topic that has been undergoing research for an extensive period of time. Previous research shows that although alarming detections can be made, it is very hard to clinically diagnose conditions and formulate a prognosis. At best, single-lead devices can show indications of serious issues but more tests must be ran in clinics before official diagnostics can be made.

- Can twelve-lead data help give us insight into which lead is better for CVD detection?

From our data, the V4 lead performed exceptionally well compared to the other tested leads. This may indicate the

horizontal V4 lead may be an ideal electrode placement for future single-lead tests.

- How well do models designed for single-lead data perform when trained using twelve-lead data?

Comparing the single-lead data to the 5 leads, we can see a few clear distinctions. The single lead had an overall higher validation accuracy and significantly higher F1 score on the atrial fibrillation class compared to all of 5 of the other leads. However, whether this result is a result from better feature detection, potential bias, or from the distributions of data remains to be further concluded.

- Can models trained on single-leads work produce correct ouputs for twelve-lead testing data?

From our data in table 5, we can see that our model performed exceptionally poor with this test. There are many possible reasons this may be the case. A few could be miss-handling on the relabeling procedure or that the sources of our data could be too drastically different for feature detection between datasets.

In conclusion, our study may indicate that there is promise in prioritizing the V4th horizontal lead when recording data from a single-lead device. If we are to attempt to transition to mobile-health monitoring devices with smaller, less sensitive data, we must be sure to use the best possible method of obtaining crucial, life-saving data. The overall idea that we can phase out twelve-lead devices for smaller, more mobile ones is yet unknown.

REFERENCES

[1] P. A. Heidenreich, J. G. Trogdon, O. A. Khavjou, J. Butler, K. Dracup, M. D. Ezekowitz, E. A. Finkelstein, Y. Hong, S. C. Johnston, A. Khera *et al.*, "Forecasting the future of cardiovascular disease in the united states: a policy statement from the american heart association," *Circulation*, vol. 123, no. 8, pp. 933–944, 2011.

[2] L. J. Laslett, P. Alagona, B. A. Clark, J. P. Drozda, F. Saldivar, S. R. Wilson, C. Poe, and M. Hart, "The worldwide environment of cardiovascular disease: prevalence, diagnosis, therapy, and policy issues: a report from the american college of cardiology," *Journal of the American College of Cardiology*, vol. 60, no. 25S, pp. S1–S49, 2012.

[3] H. Ding, A. Sarela, R. Helmer, M. Mestrovic, and M. Karunanithi, "Evaluation of ambulatory ecg sensors for a clinical trial on outpatient cardiac rehabilitation," in *IEEE/ICME International Conference on Complex Medical Engineering*. IEEE, 2010, pp. 240–243.

[4] M. Amini, F. Zayeri, and M. Salehi, "Trend analysis of cardiovascular disease mortality, incidence, and mortality-to-incidence ratio: results from global burden of disease study 2017," *BMC Public Health*, vol. 21, no. 1, pp. 1–12, 2021.

[5] A. Natarajan, Y. Chang, S. Mariani, A. Rahman, G. Boverman, S. Vij, and J. Rubin, "A wide and deep transformer neural network for 12-lead ecg classification," in *2020 Computing in Cardiology*. IEEE, 2020, pp. 1–4.

[6] R. Cuocolo, T. Perillo, E. De Rosa, L. Ugga, and M. Petretta, "Current applications of big data and machine learning in cardiology," *Journal of geriatric cardiology: JGC*, vol. 16, no. 8, p. 601, 2019.

[7] J. C. Seah, J. S. Tang, A. Kitchen, F. Gaillard, and A. F. Dixon, "Chest radiographs in congestive heart failure: visualizing neural network learning," *Radiology*, vol. 290, no. 2, pp. 514–522, 2019.

[8] B. A. Carabello and W. J. Paulus, "Aortic stenosis," *The lancet*, vol. 373, no. 9667, pp. 956–966, 2009.

[9] D. Playford, E. Bordin, L. Talbot, R. Mohamad, B. Anderson, and G. Strange, "Analysis of aortic stenosis using artificial intelligence," *Heart, Lung and Circulation*, vol. 27, p. S216, 2018.

[10] S. Narula, K. Shameer, A. M. Salem Omar, J. T. Dudley, and P. P. Sengupta, "Machine-learning algorithms to automate morphological and functional assessments in 2d echocardiography," *Journal of the American College of Cardiology*, vol. 68, no. 21, pp. 2287–2295, 2016.

[11] A. Madani, R. Arnaout, M. Mofrad, and R. Arnaout, "Fast and accurate view classification of echocardiograms using deep learning," *NPJ digital medicine*, vol. 1, no. 1, pp. 1–8, 2018.

[12] A. Isin and S. Ozdalili, "Cardiac arrhythmia detection using deep learning," *Procedia computer science*, vol. 120, pp. 268–275, 2017.

[13] Z. I. Attia, S. Kapa, F. Lopez-Jimenez, P. M. McKie, D. J. Ladewig, G. Satam, P. A. Pellikka, M. Enriquez-Sarano, P. A. Noseworthy, T. M. Munger *et al.*, "Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram," *Nature medicine*, vol. 25, no. 1, pp. 70–74, 2019.

[14] C. D. Galloway, A. V. Valys, J. B. Shreibati, D. L. Treiman, F. L. Petterson, V. P. Gundotra, D. E. Albert, Z. I. Attia, R. E. Carter, S. J. Asirvatham *et al.*, "Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram," *JAMA cardiology*, vol. 4, no. 5, pp. 428–436, 2019.

[15] J.-m. Kwon, Y. Lee, Y. Lee, S. Lee, and J. Park, "An algorithm based on deep learning for predicting in-hospital cardiac arrest," *Journal of the American Heart Association*, vol. 7, no. 13, p. e008678, 2018.

[16] T. A. Daghistani, R. Elshawi, S. Sakr, A. M. Ahmed, A. Al-Thwayee, and M. H. Al-Mallah, "Predictors of in-hospital length of stay among cardiac patients: A machine learning approach," *International journal of cardiology*, vol. 288, pp. 140–147, 2019.

[17] B. J. Mortazavi, N. S. Downing, E. M. Bucholz, K. Dharmarajan, A. Manhapra, S.-X. Li, S. N. Negahban, and H. M. Krumholz, "Analysis of machine learning techniques for heart failure readmissions," *Circulation: Cardiovascular Quality and Outcomes*, vol. 9, no. 6, pp. 629–640, 2016.

[18] S. Stern and D. Tzivoni, "Early detection of silent ischaemic heart disease by 24-hour electrocardiographic monitoring of active subjects." *British heart journal*, vol. 36, no. 5, p. 481, 1974.

[19] S. Somani, A. J. Russak, F. Richter, S. Zhao, A. Vaid, F. Chaudhry, J. K. De Freitas, N. Naik, R. Miotto, G. N. Nadkarni *et al.*, "Deep learning and the electrocardiogram: review of the current state-of-the-art," *EP Europace*, 2021.

[20] M. Kukar, I. Kononenko, C. Grošelj, K. Kralj, and J. Fettich, "Analysing and improving the diagnosis of ischaemic heart disease with machine learning," *Artificial intelligence in medicine*, vol. 16, no. 1, pp. 25–50, 1999.

[21] N. W. Communications, "Machine learning for cardiovascular disease improves when social, environmental factors are included," Jul 2021. [Online]. Available: https://www.nyu.edu/about/news-publications/news/2021/july/machine-learning-for-cardiovascular-disease-improves-when-social.html

[22] R. G. Muthalaly and R. M. Evans, "Applications of machine learning in cardiac electrophysiology," *Arrhythmia & Electrophysiology Review*, vol. 9, no. 2, p. 71, 2020.

[23] G. Landesberg, M. Mosseri, Y. Wolf, Y. Vesselov, and C. Weissman, "Perioperative myocardial ischemia and infarction: identification by continuous 12-lead electrocardiogram with online ST-segment monitoring," *Anesthesiology*, vol. 96, no. 2, pp. 264–270, Feb 2002.

[24] R. Duarte, A. Stainthorpe, J. Mahon, J. Greenhalgh, M. Richardson, S. Nevitt, E. Kotas, A. Boland, H. Thom, T. Marshall, M. Hall, and Y. Takwoingi, "Lead-I ECG for detecting atrial fibrillation in patients attending primary care with an irregular pulse using single-time point testing: A systematic review and economic evaluation," *PLoS One*, vol. 14, no. 12, p. e0226671, 2019.

[25] K. Khunti, "Accurate interpretation of the 12-lead ecg electrode placement: A systematic review," *Health Education Journal*, vol. 73, no. 5, pp. 610–623, 2014. [Online]. Available: https://doi.org/10.1177/0017896912472328

[26] K. G. Harmon, M. Zigman, and J. A. Drezner, "The effectiveness of screening history, physical exam, and ECG to detect potentially lethal cardiac disorders in athletes: a systematic review/meta-analysis," *J Electrocardiol*, vol. 48, no. 3, pp. 329–338, 2015.

[27] F. Andreotti, O. Carr, M. A. F. Pimentel, A. Mahdi, and M. De Vos, "Comparing feature-based classifiers and convolutional neural networks to detect arrhythmia from short segments of ecg," in *2017 Computing in Cardiology (CinC)*, 2017, pp. 1–4.

[28] E. A. Perez Alday, A. Gu, A. J Shah, C. Robichaux, A. K. Ian Wong, C. Liu, F. Liu, A. Bahrami Rad, A. Elola, S. Seyedi, Q. Li, A. Sharma, G. D. Clifford, and M. A. Reyna, "Classification of 12-

lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020," *Physiol Meas*, vol. 41, no. 12, p. 124003, 01 2021.

[29] H. Zhu, Y. Zhao, Y. Pan, H. Xie, F. Wu, and R. Huan, "Robust heartbeat classification for wearable single-lead ecg via extreme gradient boosting," *Sensors*, vol. 21, no. 16, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/16/5290