ECG for high-throughput screening of multiple diseases: Proof-of-concept using multi-diagnosis deep learning from population-based datasets

Weijie Sun[1,2] **Sunil Vasu Kalmady**[1,3] **Amir Salimi**[2] **Nariman Sepehrvand**[1]

Eric Ly[1] **Abram Hindle**[2] **Russell Greiner**[2,3] **Padma Kaul**[1]

[1] Canadian VIGOUR Centre, Department of Medicine, University of Alberta, Alberta, Canada
[2] Department of Computing Science, University of Alberta, Alberta, Canada
[3] Alberta Machine Intelligence Institute, Alberta, Canada

weijie2@ualberta.ca   kalmady@ualberta.ca

Abstract

Electrocardiogram (ECG) abnormalities are linked to cardiovascular diseases, but may also occur in other non-cardiovascular conditions such as mental, neurological, metabolic and infectious conditions. However, most of the recent success of deep learning (DL) based diagnostic predictions in selected patient cohorts have been limited to a small set of cardiac diseases. In this study, we use a population-based dataset of >250,000 patients with >1000 medical conditions and >2 million ECGs to identify a wide range of diseases that could be accurately diagnosed from the patient's first in-hospital ECG. Our DL models uncovered 128 diseases and 68 disease categories with strong discriminative performance.

1   Introduction

Electrocardiogram (ECG) captures the propagation of the electrical signal in the heart and is one of the most routinely used non-invasive modalities in healthcare to diagnose cardiovascular diseases [8]. However, ECG signals can be complex, making it challenging and time-consuming to interpret, even for experts. In recent years, deep learning (DL) models have been successful in reaching near human levels of performance, however most of these studies have been limited to typical ECG abnormalities such as arrhythmias [1] and a limited set of heart diseases including valvulopathy, cardiomyopathy, and ischaemia [16].

Several clinical studies have shown strong associations of ECG abnormalities with numerous diseases beyond cardiovascular conditions, including but not limited to mental disorders : depression [18], bipolar disorder [4]; infectious conditions : HIV [15], sepsis [14]; metabolic diseases : diabetes type 2 [5], amyloidosis [2]; drug use : psychotropics [11], cannabis [23]; neurological disorders: Alzheimer disease [24], cerebral palsy [10]; respiratory diseases : pneumoconiosis [22], chronic obstructive pulmonary disease [7]; digestive system diseases : liver cirrhosis [17], alcoholic liver disease [19]; miscellaneous conditions: chronic kidney disease [13], preterm labour [3], systemic lupus erythematosus [9] etc. However, despite well established clinical associations of ECG changes with multiple diseases, very few studies have explored the information contained in ECGs that could be harnessed for prediction of non-cardiovascular conditions. A major challenge here is the lack of

availability of large training datasets of digitized ECGs that could be linked to concurrent diagnostic information across various disease types. In this context, standardized administrative health data, routinely generated at each encounter, provide a wonderful opportunity to explore the full spectrum of patient diagnoses. These data include the most responsible diagnosis, as well as any comorbidities the patient may have or develop during presentation.

In this study, we use a population-based dataset of >250,000 patients with various medical conditions and >2 million in-hospital ECGs. Here, we use diagnoses coded using the World Health Organization International Classification of Diseases (ICD) [20]. The goal of our study is to identify which diseases (with previously known or unknown associations with ECGs) can be accurately diagnosed from the patient's first ECG during an emergency department (ED) visit or hospitalization based on a learned DL model. It aims to provide a proof-of-concept for high-throughput screening of ICD-wide range of diseases based on ECG, and presents disease candidates to be explored in future ECG studies with focused investigation on specific diagnosis.

2 Method

This study used population-based datasets from 26 hospitals in Alberta, Canada (2007-2020), containing information on 772,932 healthcare episodes (hospitalization and ED visits) of 260,065 patients who collectively had 13,179 unique ICD-10 codes/diseases [20]. We linked these episodes to a dataset of 2,015,808 ECGs (Philips IntelliSpace system, 12-lead, 500 Hz, 10 s) using unique patient identifiers and timing of ECG acquisition. After data cleaning and exclusions (poor signal quality[1], unlinked episodes, pacemaker and devices, < 18 years old, etc.), we used 1,514,968 ECGs that were linked to 724,074 episodes of 239,852 patients with 11,207 unique ICD codes. An ICD-10 code is 3 to 7 characters that specifies a specific disease, where the first 3 characters denote the general category of disease (e.g., 'I214' refers to 'Non-ST elevation (NSTEMI) myocardial infarction' and 'I21' refers to its broader category 'Acute myocardial infarction'). We used ICD codes and corresponding categories as labels for prediction modelling. We found 1,319 ICD codes (full code, exact match) and 699 ICD categories (match first 3 digits) that were each linked to at least 1000 ECGs.

We split our ECG dataset into the internal validation set (random 60%: 143,939 patients with 436,508 ECGs, used for training and internal validation) and external holdout set (remaining 40%: 95,913 patients with 287,566 ECGs), while ensuring that ECGs from the same patient were not shared between the sets. Whenever there were multiple ECGs in an episode, we used only the first ECG for evaluation, as it would be preferable in actual clinical practice to make a diagnostic prediction at the first point of care in the ED or hospital. We trained two DL models, for full ICD codes and ICD categories. We first trained and evaluated the performance with 80%-20% split within the internal set, and selected a list of top labels based on discriminative performance (Area under receiver operating characteristic curve (AUROC)). We then retrained the models on the entire internal set and evaluated on the external set based on the selected labels. Our DL architecture was based on ResNet [6], similar to the one used in earlier ECG modeling study [12]. Here, 12-lead ECG traces were input to the network, consisting of convolutional layer (conv), 4 residual blocks with 2 conv per block, followed by a dense layer to which age and sex features were concatenated. We used batch normalization, ReLU and dropout after each conv. The last block is then fed into a dense layer with sigmoid activation to output a 1319 (resp., 699) length vector of predicted probabilities for the codes/diseases (resp., categories). We used the Adam optimizer, learning rate of 0.001, batch size of 512, and binary cross entropy as loss function.

3 Results

In our internal validation, we found 369 out of 1319 ICD codes and 170 out of 699 ICD categories to have AUROC > 80%. Among these, 70 ICD codes and 29 ICD categories had AUROC > 90%. However, several of these labels had low precision, therefore we restricted the list to the labels with at least 5% AUPRC (area under precision-recall curve) or with an average precision that is at least 20 times greater than the prevalence of the condition. This yielded 151 ICD codes and 80 ICD categories with AUROC > 80%; and 52 ICD codes and 18 ICD categories with AUROC > 90%. Finally, we examined the replication of these lists in the external validation, and found that 128 out of

---

[1]Trace quality was ensured on muscle artifact, AC noise, baseline wander, QRS clipping, leads-off flags etc.

Chapter: IX Diseases of the circulatory system Number of ICD codes / categories: > 80% AUROC: 34 , > 90% AUROC: 14	Acute myocardial infarction (I210, I211, I212, I214, I219) Acute myocarditis (I409, I400) Acute pericarditis (I301, I309) Angina pectoris (I200, I2088) Aortic aneurysm and dissection (I712) Atrial fibrillation and flutter (I480, I481, I4890, I4891) Atrioventricular and left bundle-branch block (I442, I447, I441) Cardiac arrest (I460) Cardiomyopathy (I420, I421, I422, I426, I428) Cardiomyopathy in diseases classified elsewhere (I43) Certain current complications following acute myocardial infarction (I2381) Chronic ischaemic heart disease (I2511, I255, I2510) Complications and ill-defined descriptions of heart disease (I514, I518) Heart failure (I500) Nonrheumatic aortic valve disorders (I350, I352) Nonrheumatic mitral valve disorders (I340) Other cardiac arrhythmias (I4900) Other conduction disorders (I456) Other diseases of pericardium (I313, I319) Other disorders of circulatory system in diseases classified elsewhere (I982, I983) Other pulmonary heart diseases (I272) Rheumatic mitral valve diseases (I050)
Chapter: I Certain infectious and parasitic diseases Number of ICD codes / categories: > 80% AUROC: 3 , > 90% AUROC: 0	Chronic viral hepatitis (B182) Other sepsis (A419) Unspecified human immunodeficiency virus [HIV] disease (B24)
Chapter: IV Endocrine, nutritional and metabolic diseases Number of ICD codes / categories: > 80% AUROC: 10 , > 90% AUROC: 2	Amyloidosis (E85) Obesity (E662, E668) Other disorders of fluid, electrolyte and acid-base balance (E875) Type 1 diabetes mellitus (E1010, E1023, E1042) Type 2 diabetes mellitus (E1123, E1128) Unspecified diabetes mellitus (E1410)
Chapter: V Mental and behavioural disorder Number of ICD codes / categories: > 80% AUROC: 15 , > 90% AUROC: 5	Eating disorders (F50) Multiple drug use and use of other psychoactive substances (F191) Mental and behavioural disorders due to use of alcohol (F100, F101, F102, F103) Mental and behavioural disorders due to use of cannabinoids (F121, F125) Mental and behavioural disorders due to use of cocaine (F14) Mental and behavioural disorders due to use of opioids (F11) Mental and behavioural disorders due to use of other stimulants (F151) Schizophrenia (F209) Specific personality disorders (F603) Unspecified dementia (F03) Unspecified nonorganic psychosis (F29)
Chapter: VI Diseases of the nervous system Number of ICD codes / categories: > 80% AUROC: 6 , > 90% AUROC: 0	Alzheimer disease (F00) Cerebral palsy (G809) Other degenerative diseases of nervous system, not elsewhere classified (G312) Other disorders of brain (G931) Primary disorders of muscles (G71) Sleep disorders (G4730)
Chapter: X Diseases of the respiratory system Number of ICD codes / categories: > 80% AUROC: 8 , > 90% AUROC: 1	Other chronic obstructive pulmonary disease (J440, J441) Other interstitial pulmonary diseases (J841) Pneumoconiosis due to dust containing silica (J620) Respiratory conditions due to other external agents (J70) Respiratory failure, not elsewhere classified (J960, J9691, J969)
Chapter: XI Diseases of the digestive system Number of ICD codes / categories: > 80% AUROC: 8 , > 90% AUROC: 5	Alcoholic liver disease (K701, K703, K704) Fibrosis and cirrhosis of liver (K746) Hepatic failure, not elsewhere classified (K721, K729) Other diseases of liver (K767, K766)
Chapter: XIV Diseases of the genitourinary system Number of ICD codes / categories: > 80% AUROC: 6 , > 90% AUROC: 1	Chronic kidney disease (N185) Excessive, frequent and irregular menstruation (N920) Glomerular disorders in diseases classified elsewhere (N083, N0835, N0839) Hyperplasia of prostate (N40)
Misc Chapters: II, XIII, XV, XVII Neoplasms | Diseases of the musculoskeletal system and connective tissue | Pregnancy, childbirth and the puerperium | Congenital malformations, deformations and chromosomal abnormalities Number of ICD codes / categories: > 80% AUROC: 8 , > 90% AUROC: 4	Secondary malignant neoplasm of respiratory and digestive organs (C782) Osteoporosis without pathological fracture (M819) Systemic lupus erythematosus (M321) Other maternal diseases complicating pregnancy, childbirth and puerperium (O99) Labour and delivery complicated by fetal stress [distress] (O68) Preterm labour and delivery (O60) Congenital malformations of aortic and mitral valves (Q231) Congenital malformations of cardiac septa (Q211)

Figure 1: Validated list of top performing ICD codes and categories that could be predicted from the patient's first in-hospital ECG using deep learning.

151 (84.8%) ICD codes and 68 out of 80 (85.0%) ICD categories were replicated to have AUROC > 80%; and 40 out of 52 (76.9%) ICD codes and 16 out of 18 (88.9%) ICD categories were replicated to have AUROC > 90%. We present this final validated list in Figure 1, under different ICD sections.

4   Discussion

To the best of our knowledge, this is the first study which explores the ECG based predictability of multiple diseases over the ICD-wide diagnostic landscape. Our DL models trained and validated on population scale datasets demonstrate excellent AUROC (i.e high sensitivity & specificity) for several diseases, however their precision (PPV) might be limited, partially owing to their low prevalence rates (89.8% of diseases had <1% positive class ECGs) [21]. Therefore, model predictions for such diseases might be more suitable for 'rule out' screening, rather than 'rule in' diagnostics. Population-based records enable learning from high volume healthcare data, although diagnostic labels obtained from these records may not be considered as gold-standard ground truth without proper adjudication. Also, like any other DL model, the latent ECG features used for prediction in our models may not be directly related to underlying pathology of diseases, and could be attributed to patient's comorbidities, medication usage and lifestyle factors that are naturally correlated with disease states in the population. Finally, although most adult patients get an ECG at some point during their lifetime, there is a potential for selection bias in our cohort as it is restricted to patients who had undergone at least one ECG in the 13 year period (2007-2020). Therefore, these results should be considered preliminary proof-of-concept for further investigation of specific diseases by future studies. Our next steps will focus on ensuring the generalizability across the hospitals (evaluating on ECGs from hospital sites that were not used during training) and robustness against possible model biases, such as towards certain gender or ethnic groups. That said, the current study demonstrates an exciting potential for state-of-the-art DL models trained on a ubiquitous diagnostic test (ECG) linked to routinely collected health data to transform high-throughput diagnostics for a wide range of diseases.

5   Potential negative societal impact

Our results are primarily based on AUROC which is commonly used in clinical models as concordance index, even though it does not account for real world costs associated with individual diseases. Moreover, given the low precision in identifying the majority of disease conditions, a positive result may trigger a cascade of invasive or non-invasive diagnostic measures that may inflict high costs on the healthcare systems. However, early diagnosis of conditions could provide opportunity to control disease progress and prevent further complications.

References

[1] Erick A Perez Alday, Annie Gu, Amit J Shah, Chad Robichaux, An-Kwok Ian Wong, Chengyu Liu, Feifei Liu, Ali Bahrami Rad, Andoni Elola, Salman Seyedi, et al. Classification of 12-lead ECGs: the Physionet/computing in cardiology challenge 2020. *Physiological measurement*, 41(12):124003, 2020.

[2] Zhongwei Cheng, Kongbo Zhu, Zhuang Tian, Dachun Zhao, Quancai Cui, and Quan Fang. The findings of electrocardiography in patients with cardiac amyloidosis. *Annals of Noninvasive Electrocardiology*, 18(2):157–162, 2013.

[3] M Gemelli, F De Luca, R Manganaro, R Leonardi, F Rando, A Agnetti, C Mami, and G Di Pasquale. Transient electrocardiographic changes suggesting myocardial ischaemia in newborn infants following tocolysis with beta-sympathomimetics. *European journal of pediatrics*, 149(10):730–733, 1990.

[4] Brandon Hage, Briana Britton, David Daniels, Keri Heilman, Stephen W Porges, and Angelos Halaris. Low cardiac vagal tone index by heart rate variability differentiates bipolar from major depression. *The World Journal of Biological Psychiatry*, 20(5):359–367, 2019.

[5] Peter P Harms, Amber A van der Heijden, Femke Rutters, Hanno L Tan, Joline WJ Beulens, Giel Nijpels, Petra Elders, et al. Prevalence of ECG abnormalities in people with type 2 diabetes: The Hoorn diabetes care system cohort. *Journal of Diabetes and its Complications*, 35(2):107810, 2021.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[7] Marte Strømsnes Larssen, Kjetil Steine, Janne Mykland Hilde, Ingunn Skjørten, Christian Hodnesdal, Knut Liestøl, and Knut Gjesdal. Mechanisms of ECG signs in chronic obstructive pulmonary disease. *Open heart*, 4(1):e000552, 2017.

[8] Aurore Lyon, Ana Mincholé, Juan Pablo Martínez, Pablo Laguna, and Blanca Rodriguez. Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances. *Journal of The Royal Society Interface*, 15(138):20170821, 2018.

[9] Gihyun Myung, Lindsy J Forbess, Mariko L Ishimori, Sumeet Chugh, Daniel Wallace, and Michael H Weisman. Prevalence of resting-ECG abnormalities in systemic lupus erythematosus: a single-center experience. *Clinical rheumatology*, 36(6):1311–1316, 2017.

[10] Carlos Alberto Pastore, Nelson Samesima, Rodrigo Imada, Marta Reis, Maria Teresa Santos, Maria Cristina Ferreira, Cesar Grupi, Fernanda Fumagalli, Jaqueline Wagenfuhr, and Maira Chammas. Characterization of the electrocardiographic pattern of individuals with cerebral palsy. *Journal of electrocardiology*, 44(2):138–141, 2011.

[11] Christoffer Polcwiartek, Kristian Kragholm, Steen M Hansen, Brett D Atwater, Daniel J Friedman, Carlo A Barcella, Claus Graff, Jonas B Nielsen, Adrian Pietersen, Jimmi Nielsen, et al. Electrocardiogram characteristics and their association with psychotropic drugs among patients with schizophrenia. *Schizophrenia bulletin*, 46(2):354–362, 2020.

[12] Antônio H Ribeiro, Manoel Horta Ribeiro, Gabriela MM Paixão, Derick M Oliveira, Paulo R Gomes, Jéssica A Canazart, Milton PS Ferreira, Carl R Andersson, Peter W Macfarlane, Wagner Meira Jr, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature communications*, 11(1):1–9, 2020.

[13] Salman Shafi, Mohammad Saleem, Roshina Anjum, Wajid Abdullah, and Tahir Shafi. ECG abnormalities in patients with chronic kidney disease. *Journal of Ayub Medical College Abbottabad*, 29(1):61–64, 2017.

[14] Supreeth P Shashikumar, Matthew D Stanley, Ismail Sadiq, Qiao Li, Andre Holder, Gari D Clifford, and Shamim Nemati. Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. *Journal of electrocardiology*, 50(6):739–743, 2017.

[15] Elsayed Z Soliman, Ronald J Prineas, Mollie P Roediger, Daniel A Duprez, Franck Boccara, Christoph Boesecke, Christoph Stephan, Sally Hodder, James H Stein, Jens D Lundgren, et al. Prevalence and prognostic significance of ECG abnormalities in hiv–infected patients: results from the strategies for management of antiretroviral therapy study. *Journal of electrocardiology*, 44(6):779–785, 2011.

[16] Sulaiman Somani, Adam J Russak, Felix Richter, Shan Zhao, Akhil Vaid, Fayzan Chaudhry, Jessica K De Freitas, Nidhi Naik, Riccardo Miotto, Girish N Nadkarni, et al. Deep learning and the electrocardiogram: review of the current state-of-the-art. *EP Europace*, 2021.

[17] Letitia Toma, Adriana Mercan Stanciu, Anca Zgura, Nicolae Bacalbasa, Camelia Diaconu, and Laura Iliescu. Electrocardiographic changes in liver cirrhosis—clues for cirrhotic cardiomyopathy. *Medicina*, 56(2):68, 2020.

[18] Yiming Wang, Xun Zhao, Adrienne O'Neil, Alyna Turner, Xingde Liu, and Michael Berk. Altered cardiac autonomic nervous function in depression. *BMC psychiatry*, 13(1):1–7, 2013.

[19] M Wehr, J Hess, B Noll, and JC Bode. Cardiac findings in alcoholic liver disease. *Medizinische Klinik (Munich, Germany: 1983)*, 85(11):629–36, 1990.

[20] WHO. *International statistical classification of diseases and related health problems*, volume 2. World Health Organization, 2016.

[21] Alexander William Wong. Classification and analysis of 12-lead electrocardiograms. Master's thesis, University of Alberta, 2021.

[22] Qiuyun Wu, Lei Han, Ming Xu, Hengdong Zhang, Bangmei Ding, and Baoli Zhu. Effects of occupational exposure to dust on chest radiograph, pulmonary function, blood pressure and electrocardiogram among coal miners in an eastern province, china. *BMC public health*, 19(1):1–8, 2019.

[23] Ella Yahud, Gideon Paul, Michael Rahkovich, Lubov Vasilenko, Yonatan Kogan, Eli Lev, and Avishag Laish-Farkash. Cannabis induced cardiac arrhythmias: a case series. *European Heart Journal-Case Reports*, 4(6):1–9, 2020.

[24] Roberto Zulli, Franco Nicosia, Barbara Borroni, Chiara Agosti, Paola Prometti, Paolo Donati, Massimiliano De Vecchi, Giuseppe Romanelli, Vittorio Grassi, and Alessandro Padovani. Qt dispersion and heart rate variability abnormalities in alzheimer's disease and in mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(12):2135–2139, 2005.