Predicting the Effects of Drugs and Comorbidities on Arrhythmogenic Risk using Deep Learning*

Emma Maliar
maliars@stanford.edu

Abstract

Arrhythmias affect millions of people worldwide. In particular, arrhythmias cause 200,000 - 300,000 sudden deaths in the US per year. Doctors diagnose arrhythmias by looking at electrocardiograms that represent a patient's heartbeat. However, these diagnoses are not always accurate. For example, the average cardiologist accuracy for diagnosing atrial fibrillation is about 50%.

We developed an artificial intelligence statistical model that diagnoses arrhythmias by analyzing the patients' electrocardiogram values, comorbidities and drugs. Our model, written in python, relies on a multilayer neural network which is trained by deep learning optimization methods. We analyze the database ECG-ViEW II from South Korea that contains information on 461,000 patients, including 10,081 comorbidities and drugs. To reduce the runtime, because of the large dataset, we run the code on a supercomputer Bridges 2. Our deep learning model diagnoses arrhythmias with an overall accuracy of 83.87%, thus overperforming trained medical doctors.

Using our deep learning model to evaluate how drugs and comorbidities contribute to a patient's risk of suffering from an arrhythmia, we find that most common drugs such as aspirin and Vitamin C do not significantly affect the incidence of arrhythmias. But we also distinguished a number of drugs and comorbidities that have a strong statistically significant effect on the incidence of arrhythmias, particularly, drugs related to pregnancy, skin eruptions, and the stroke. Our deep learning analysis can aid doctors diagnose and prevent arrhythmias by informing prescriptions and by treating comorbidities that increase the risk of arrhythmias.

Keywords: Arrhythmogenic Risk, Comorbidities, Deep Learning,

Introduction

Arrhythmias, which are heart conditions characterized by an abnormal heart rate, are the underlying cause of 200,000 - 300,000 sudden cardiac deaths in the United States per year. They occur when electrical signals are disrupted from the sinus node, which regulates the heart's depolarization mechanism whose purpose is to oxygenate and supply organs with nutrients.

Arrhythmias are primarily diagnosed using electrocardiograms (ECGs), or recordings of the heart's rhythm over an interval of time, that are interpreted by cardiologists, who look for established patterns characteristic of heart conditions. Values describing aspects of the heart beat such as strength and duration of the heart's depolarization found on the ECG are observed to diagnose patients with arrythmias. Letters are used to denote sections of the heart beat pattern (Savalia et., al 2018). For example, the QT interval reflects ventricular depolarization and is usually around 440 milliseconds long, though this number can vary based on gender or slight variations.

Arrhythmias are classified by heartbeat patterns that can also either be too fast (tachycardia), or too slow (bradyarrhythmias), and by location of occurrence as either supraventricular arrhythmias or ventricular arrhythmias (Savalia et al., 2018). Each arrhythmia type is problematic in its own way; if a person's heartbeat is too slow, it may not be able to pump enough blood to the person's muscles. On the other hand, if a person's heartbeat is too fast, the ventricles are not allowed enough time to fully fill with blood, to then be pumped out to the rest of the body. The various types of arrhythmias have a large variety of consequences from chest pain and fatigue to cardiac arrest and stroke and treatments in the form of antiarrhythmic drugs, surgery, and other medical action, necessitating the accurate diagnosis of arrhythmias. Even when treated, there is currently no way to directly reset the complex sinus rhythm that regulates the heartbeat.

Currently there are few techniques to stratify patients based on arrhythmia risk in order to circumvent dangerous consequences by preventing arrhythmias before they occur. The only such test that is widely used, named CHA2DS2- VASc, stratifies patients for stroke risk that already have atrial fibrillation (Nielsen et al., 2020). This test also only considers factors independently, not taking into account the relationships between some variables. Furthermore, patients are only classified into three categories; low, medium, and high risk. These three labels are often not descriptive enough to classify complex patient risk (Nielsen et al., 2020).

One study found that using this test, 18% of patients who later had a thrombus, or a possibly deadly blood clot, were misclassified as low risk and 40% were classified as medium risk (Michalska et al., 2020). Tests that predict a patient's risk for developing each type of arrhythmia are not currently widely used. Such a test could be used to predict patients at risk for arrhythmias, so that steps can be taken to decrease the patient's risk of developing an arrythmia before it occurs and possibly worsens to a more serious condition. This test would be useful to, for example, differentiate which patients with Torsades de pointes, a type of arrhythmia where the ventricles beat faster than the atrials, will develop ventricular fibrillation. Within minutes ventricular fibrillation can cause sudden cardiac arrest, which has been found to be fatal within a month for 90.5% of patients (Holmburg et., al 2000). It is important to differentiate which patients will develop ventricular fibrillation in order to more effectively monitor changes in the heart's rhythm, to predict and treat cardiac arrest with a defibrillator, therefore preventing these sudden deaths.

The exact cause of the irregular heartbeats that categorize arrhythmias is unknown, though drugs and comorbidities have been found to affect arrhythmogenic risk. Xenobiotic agents including pharmaceutical drugs, illicit drugs, and toxins can disturb the heart's rhythm, each chemical affecting the body in its own way (Sahli-Costaba et al., 2020, Stapleton et al., 2015).

Numerous chemicals can have toxic effects on the vascular system, for instance by narrowing or dilating blood vessels, consequently causing hypertension or hypotension respectively. Blockage of the potassium and calcium currents ion channels has been shown to cause early afterdepolarizations that indicate an arrhythmia. Some chemicals have also been shown to extend QT duration, and interfere with cardiac systems calcium, and sodium channels, blocking or activating them. For example in patients taking Methadone daily, a narcotic used to treat pain, 16.2% experienced a 50 millisecond increase in QT length (Ehret et al., 2006). Amiodarone and Erythromycin, and Methadone have also been shown to prolong the QT interval in a patient's heartbeat, characteristic of bradycardias (Coughtrie et al., 2017).

Some comorbidities, especially those related to the heart, have also been shown to increase risk for certain arrhythmias by changing the heart's rhythm. Hypertension, diabetes, and obesity have previously been correlated with arrhythmias (Jungen et al., 2019).

Manually investigating the correlation between each of the thousands of common drugs and comorbidities and arrhythmogenic risk in which this correlation has not previously been investigated would be costly and unfeasible. Many drugs and comorbidities that are not suspected to have a significant correlation will most likely follow this prediction and not have a statistically significant effect on arrhythmogenic risk. However, without testing all drugs and comorbidities, the few unexpected ones that do significantly affect arrhythmogenic risk will not be discovered. The solution to this seemingly impossible, yet important problem is the use of artificial intelligence that has the power to complete this task within a reasonable time frame and without the use of any expensive lab equipment.

Artificial intelligence is the remarkable capability of a computer to derive meaning from a large set of examples by finding hidden patterns. Its training process mimics how the human brain learns to perform tasks. Machine learning is the application of artificial intelligence using statistical techniques. Even though machine learning has been around for several decades, this method of analysis was previously too computationally expensive, so it has only recently become feasible with the introduction of modern computers. The introduction of new collection software has also allowed for large amounts of medical data to be collected. Both of these recent advances have led machine learning to become increasingly important as a tool to diagnose a large range of diseases (Hannun et al. 2019).

In recent years, machine learning has been used in the context of heart disease, for example to predict heart attacks, and even to find the exact location where the heart muscle is suffering from blood blockage, from only a 12 lead electrocardiogram. Machine learning has been found to be able to outperform doctors by around 10% in diagnosing arrhythmias based on ECGS. Many medical diagnosis studies that focused on heart diseases found that deep learning models, parameterized by having multilayer neural networks, outperformed other machine learning models including linear regression, decision trees, and support vector machines (Jafarian et al., 2020, Weng et al., 2017, Ribeiro et. al., 2020).

Using a machine learning algorithm to classify patients based on these patterns is beneficial because it will discover all these patterns and more subtle ones that cannot be detected by humans while still taking into account the patient holistically. For example, the patient's gender will also be inputted into the model that will make the connection that women tend to have longer QT intervals.This study has several goals. One goal of this study is to detect and differentiate between various arrhythmias in order to identify patients with a high risk of developing a certain arrhythmia before it occurs.

In addition, this study aims to find correlations between various arrhythmias and medications, as well as conditions in order to inform arrhythmia risk prediction by considering the patient holistically and taking into account all factors for a more personalized accurate assessment. In particular, we use artificial intelligence to investigate the relationships between many common drugs and comorbidities and arrhythmogenic risk using a supervised deep learning statistical model.

Methods

The Electrocardiogram Vigilance with Electronic data Warehouse II dataset was used. This de-identified dataset was obtained from ecg view.org and consists of 461,000 patients 99.1% of which were from South Korea, with electrocardiogram values over the years 1998-2003. 10,081 comorbidities and drugs were tested along with other variables such as lab tests for potassium, calcium, and magnesium, age group, sex, and ecg values including RR interval, PR interval, QRS duration, QT interval, QTc interval, P axis, QRS axis, and T axis. A CITI certification was obtained in order to access the database used in this study.

A deep neural network was used to build a statistical model and it was trained using deep learning analysis. Preliminary tests were conducted to determine the optimal NN architecture including the number of hidden layers and the number of nodes. A column of nodes creates a hidden layer where every node takes input from every single node in the previous layer, including a constant bias node, and applies batch normalization and then a transformation. The values from the last hidden layer are then transformed into one value that is imputed into the sigmoid prediction function. The number of nodes is the same in each hidden layer. The network used has two hidden layers with 64 neurons in each layer. The learning rate and number of steps were also adjusted to the values that yielded the highest accuracies; the final learning rate was 1e-3 and the final number of steps was 5,000. The algorithm was written in Python, on Jupyter Notebook, using the Pandas and Numpy libraries, and pytorch, a software used in many AI applications such as image and speech recognition, self-driving cars, etc. This study can be considered a supervised classification problem, where the target variables are given as input and are categorical instead of continuous real number variables (Sevakula, 2020).

Since our deep learning model is very computationally expensive to run, and our dataset is large, bridges supercomputer based in Pittsburgh was used. It has 62,464 cores instead of the usual 4-8 cores on a laptop and the runtime on the supercomputer is 10 minutes.

First the data files were imported as data frames, and then the data was cleaned to be inputted into the neural network. Only the first five characters of the icd 10 diagnosis codes were included, excluding the location, alteration, and extension values that provide more specific information about the condition. Only drugs and conditions with more than 25 people were included in the dataset, so that the model had enough instances to predict the effect of each input in regard to other variables. Only input values before the arrhythmia diagnosis date were included in the testing input to simulate the information that would be available to diagnose an arrhythmia before its occurrence. 15 types of arrhythmias were tested with icd codes from the I47, I48, and I49 classes. Cross validation was used to give validity to the accuracy value by allocating 80% of the data to the training set, 10% to the cross validation set, and 10% to the test set, so that the high order polynomial is not fitted to the test set and the error can be generalized to future data. (Rajpurkar, 2019).

The neural network was first initialized randomly and trained (Ribeiro et al., 2020). The learning curve was then plotted to check that the model converged, or reached a point where any changes to the model produced only negligible improvements in accuracy. Learning curves can be plotted in python using the matplotlib.pyplot function, using number of iterations as the x axis and the cost outputted by the cost function as the y axis. At first our model had a high loss because it was randomly diagnosing arrhythmias, then over many iterations, the accuracy improved. The model was tested several times, while adjusting the model architecture in order to improve the model's accuracy for diagnosing various arrhythmias. The accuracy was obtained using the following cost function: $J(\theta) = 1/2m \Sigma(h\theta (x(i))-y(i))2$, where m is the number of patients.

To assess how strong the correlation between arrhythmogenic risk and each drug or comorbidity is, a new score was predicted using the trained model, assuming that the patient did not take a certain drug or have a certain condition. This new score reflects the likelihood that a patient has each type of arrhythmia based on the information inputted and was then compared to the original diagnosis values. The mean and standard deviations of the change between the original score and the new score were then computed for each patient, for each arrhythmia, to infer the p value of each drug and comorbidity considered. This score reflects how strong the correlation of each drug or comorbidity is with arrhythmogenic risk and can be correlated with an increase or decrease in arrhythmogenic risk.

Results

Our model had a final accuracy of 83.87 percent for diagnosing various arrhythmias.

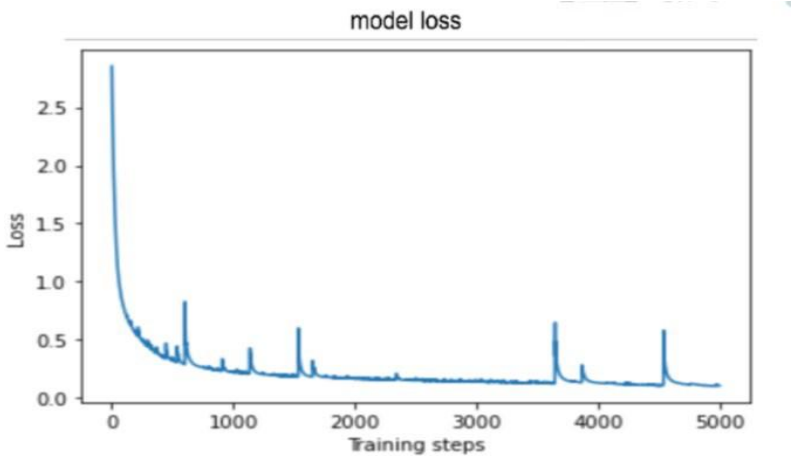

Fig 1: The model loss saw a characteristic J learning curve, that signifies that the training of neural network converged. Below, we summarize the conditions that lead to the highest risk of arrhythmias.

Drug or condition	Description	Risk Change	P value
Hypertrophic cardiomyopathy	Harder for heart to pump blood	3.28	.029
Pregnancy	The condition of being pregnant	12.43	.020
Stroke	Blood and oxygen supply to brain tissue interrupted	13.78	.007
Tissue plasminogen activator	Breaks down of blood clots	16.09	.032
Generalized skin eruption	Due to drugs and medicaments	18.56	.036
Non-S Televation myocardial infarction (heart attack)	Artery supplying blood and oxygen to heart blocked	28.64	.046

Fig 2: The change in risk for several of the conditions with the highest such values is shown above. We next show the percentage changes in the strength of association for several most influential drugs.

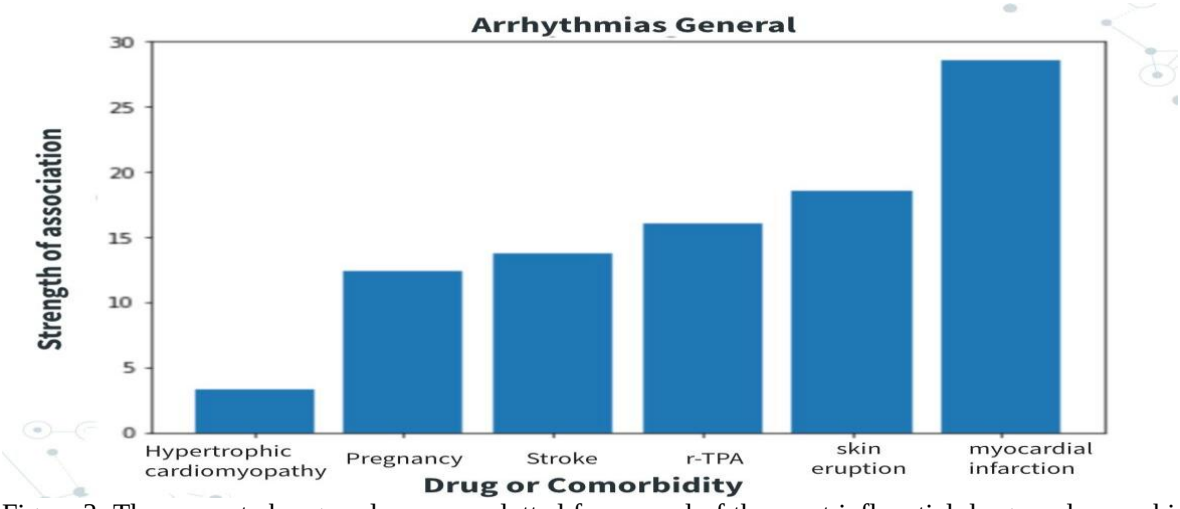

Figure 3: The percent change values were plotted for several of the most influential drugs and comorbidities in regard to arrhythmogenic risk.

Drugs and comorbidities including hypertrophic cardiomyopathy, pregnancy, stroke, tissue plasminogen activator, skin eruptions, and myocardial infarctions were statistically significant at a significance level of 0.05 and increased a patient's risk of developing various arrhythmias. The top common drugs in the dataset, including aspirin, digoxin, heparin sodium, adenosine, and vitamin B did not have statistically significant effects on arrhythmogenic risk.

Discussion

The following trends in the data were distinguished.

General skin eruptions due to medications taken internally were associated with a higher risk of developing an arrhythmia. This is shown to be consistent with literature, as antiarrhythmics, like amiodarone and Quinidex, have been shown to be associated with dermatological complications in publications. Hyperpigmentation, phototoxic reactions, and pseudoporphyria are all common dermatological complications from prolonged amiodarone intake and can indicate that a patient is taking antiarrhythmic medication (Jaworski et al., 2014).

Pregnancy, and pregnancy complications as well as drugs associated with pregnancy were very strongly correlated with many different arrhythmia types and generally increased one's arrhythmogenic risk. As shown in the table, pregnancy significantly increased arrhythmogenic risk by 12.43% on average. During pregnancy, changes in blood flow, hormones, and nervous system occur, that could trigger an arrhythmia. Literature finds that there is an increase in heart rate during pregnancy by around 15%, especially in the 3rd trimester which may predispose a patient to an arrhythmia, therefore supporting our results. (Enriquez et al., 2014).

As expected, comorbidities of the circulatory system showed the strongest correlation with arrhythmogenic risk. For example, a type of myocardial infarction was correlated with a 28.64% increase in arrhythmogenic risk. Both arrhythmias, and myocardial infarction have been found to be correlated with many similar comorbidities, specifically those that generally have a negative impact on the circulatory system. For example, patients with type 2 diabetes have been found to

be more susceptible to developing cardiac arrhythmias (Jungen et al., 2019). Similarly, type 2 diabetes patients have been established as a major risk group for myocardial infarction, because of poor glycemic control (Edqvist et al., 2019). Another heart condition, Hypertrophic cardiomyopathy which was correlated with a 3.28% change in arrhythmogenic risk, was also found to impact myofilament Ca2+-sensitivity, which has been found to increase the susceptibility of developing an arrhythmia (Wijnker et al., 2019). Furthermore, tissue plasminogen activator medication, which was correlated with a 16.09% increase in arrhythmogenic risk, is used for the breakdown of blood clots that can lead to stroke and is therefore correlated to stroke. Stroke was correlated with a 13.78% increase in arrhythmogenic risk. The association between stroke and arrhythmogenic risk is established in literature. One study found that 21.9% of patients with ischemic strokes developed arrhythmias (Daniele et al., 2002). In addition, arrhythmias are correlated with heart tissue damage from stroke or heart conditions and are common side effects of heart surgery (Nielsen et al., 2020).

Our analysis has some limitations. The study population in our data was 99.1% Korean and therefore may not be generalizable to the greater population. Stigmatized diagnoses, such as HIV, STDs, and congenital malformations were also removed, which may affect some of the outcomes.

Conclusion

A deep learning model was developed that can diagnose various arrhythmias with an accuracy of 83.87%. Our analysis facilitates a non invasive, reproducible, and reliable procedure to diagnose various arrhythmias, to predict the effect of drugs and comorbidities, and to inform decisions on treatment that a patient should undertake, if any. Several drugs and comorbidities that are significantly correlated with arrhythmogenic risk were identified, in particular drugs related to pregnancy and circulatory diseases. Our deep learning model can be used to aid doctors in diagnosing, preventing and treating arrhythmias.

Future research should be done to investigate how comorbidities and drugs affect a patient's electrocardiogram in order to better understand the cause of the correlations between each arrhythmia type and identified drugs and comorbidities that were previously believed not to have a statistically significant correlation with arrhythmogenic risk. Common drugs and comorbidities identified as not having a statistically significant correlation with arrhythmogenic risk should be reevaluated using patients from different ethnicities and age groups. The algorithm should also be run on another dataset for validation.

Acknowledgements

I would like to thank Nicole Spinelli for her help in developing this project throughout all the stages of its completion. Our many discussions on the study's design and implementation have guided many aspects of this project. I would also like to thank Amber Sun for many useful comments and suggestions.

Bibliography

Amin A, Houmsse A, Ishola A, Tyler J, Houmsse M. The current approach of atrial fibrillation management. Avicenna J Med. 2016 Jan-Mar;6(1):8-16. doi: 10.4103/2231-0770.173580. PMID: 26955600; PMCID: PMC4759971.

Coughtrie AL, Behr ER, Layton D, et al. Drugs and life-threatening ventricular arrhythmia risk: results from the DARE study cohort. BMJ Open. 2017 Oct;7(10):e016627. DOI: 10.1136/bmjopen-2017- 016627.

Daniele O, Caravaglios G, Fierro B, Natalè E. Stroke and cardiac arrhythmias. J Stroke Cerebrovasc Dis.

2002 Jan-Feb;11(1):28-33. doi: 10.1053/jscd.2002.123972. PMID: 17903852.

Edqvist, J., Rawshani, A., Adiels, M., Björck, L., Lind, M., Svensson, A. M., Gudbjörnsdottir, S., Sattar, N., & Rosengren, A. (2019). Contrasting Associations of Body Mass Index and Hemoglobin A1c on the Excess Risk of Acute Myocardial Infarction and Heart Failure in Type 2 Diabetes Mellitus. Journal of the American Heart Association, 8(24), e013871. https://doi.org/10.1161/JAHA.119.013871

Emara MK, Saadet AM. Transient atrial fibrillation in hypertensive patients with thiazide induced hypokalaemia. Postgrad Med J. 1986 Dec;62(734):1125-7. doi: 10.1136/pgmj.62.734.1125. PMID: 3658850; PMCID: PMC2418946.

Enriquez AD, Economy KE, Tedrow UB. Contemporary management of arrhythmias during pregnancy. Circ Arrhythm Electrophysiol. 2014 Oct;7(5):961-7. doi: 10.1161/CIRCEP.114.001517. PMID: 25336366.

Harris PR. The Normal Electrocardiogram: Resting 12-Lead and Electrocardiogram Monitoring in the Hospital. Crit Care Nurs Clin North Am. 2016 Sep;28(3):281-96. doi: 10.1016/j.cnc.2016.04.002. Epub 2016 Jun 22. PMID: 27484657.

Hannun, A.Y., Rajpurkar, P., Haghpanahi, M. et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 25, 65–69 (2019). https://doi.org/10.1038/s41591-018-0268-3

Holmberg M, Holmberg S, Herlitz J. Incidence, duration and survival of ventricular fibrillation in out-of- hospital cardiac arrest patients in sweden. Resuscitation. 2000 Mar;44(1):7-17. doi: 10.1016/s0300-9572(99)00155-0. PMID: 10699695.

Jaworski, K., Walecka, I., Rudnicka, L., Gnatowski, M., & Kosior, D. A. (2014). Cutaneous adverse reactions of amiodarone. Medical science monitor : international medical journal of experimental and clinical research, 20, 2369–2372. https://doi.org/10.12659/MSM.890881

Wijnker, P., Sequeira, V., Kuster, D., & Velden, J. V. (2019). Hypertrophic Cardiomyopathy: A Vicious Cycle Triggered by Sarcomere Mutations and Secondary Disease Hits. Antioxidants & redox signaling, 31(4), 318–358. https://doi.org/10.1089/ars.2017.7236

Jungen, C., Scherschel, K., Flenner, F., Jee, H., Rajendran, P., De Jong, K. A., Nikolaev, V., Meyer, C., Ardell, J. L., & Tompkins, J. D. (2019). Increased arrhythmia susceptibility in type 2 diabetic mice related to dysregulation of ventricular sympathetic innervation. American journal of physiology. Heart and circulatory physiology, 317(6), H1328–H1341. https://doi.org/10.1152/ajpheart.00249.2019

Krummen DE, Ho G, Villongco CT, Hayase J, Schricker AA. Ventricular fibrillation: triggers, mechanisms and therapies. Future Cardiol. 2016 May;12(3):373-90. doi: 10.2217/fca-2016-0001. Epub 2016 Apr 27. PMID: 27120223.

Ludhwani D, Goyal A, Jagtap M. Ventricular Fibrillation. [Updated 2020 Aug 10]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2020 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK537120/

Morrone, D., & Morrone, V. (2018). Acute Pulmonary Embolism: Focus on the Clinical Picture. Korean circulation journal, 48(5), 365–381. https://doi.org/10.4070/kcj.2017.0314

Ng GA. Neuro-cardiac interaction in malignant ventricular arrhythmia and sudden cardiac death. Auton Neurosci. 2016 Aug;199:66-79. doi: 10.1016/j.autneu.2016.07.001. Epub 2016 Jul 5. PMID: 27423297.

Rajpurkar, Pranav & Irvin, Jeremy & Zhu, Kaylie & Yang, Brandon & Mehta, Hershel & Duan, Tony & Ding, Daisy & Bagul, Aarti & Langlotz, Curtis & Shpanskaya, Katie & Lungren, Matthew & Ng, Andrew. (2017). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning.

Reno CM, Bayles J, Huang Y, Oxspring M, Hirahara AM, Dosdall DJ, Fisher SJ. Severe Hypoglycemia- Induced Fatal Cardiac Arrhythmias Are Mediated by the Parasympathetic Nervous System in Rats. Diabetes. 2019 Nov;68(11):2107-2119. doi: 10.2337/db19-0306. Epub 2019 Aug 22. PMID: 31439645; PMCID: PMC7118248.

Salaminia, S., Sayeh Miri, F., Angha, P., Sayeh Miri, K., & Motedayen, M. (2018). Evaluating the effect of magnesium supplementation and cardiac arrhythmias after acute coronary syndrome: a systematic review and meta-analysis. BMC cardiovascular disorders, 18(1), 129. https://doi.org/10.1186/s12872-018-0857-6

Sultana R, Sultana N, Rashid A, Rasheed SZ, Ahmed M, Ishaq M, Samad A. Cardiac arrhythmias and left ventricular hypertrophy in systemic hypertension. J Ayub Med Coll Abbottabad. 2010 Oct- Dec;22(4):155-8. PMID: 22455286.

Wang TJ, Parise H, Levy D, et al. Obesity and the Risk of New-Onset Atrial Fibrillation. JAMA.2004;292(20):2471–2477. doi:10.1001/jama.292.20.2471

Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N (2017) Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS ONE 12(4): e0174944. https://doi.org/ 10.1371/journal.pone.0174944

Zheng T, Xie W, Xu L, He X, Zhang Y, You M, Yang G, Chen Y. A machine learning-based framework to identify type 2 diabetes through electronic health records. Int J Med Inform. 2017 Jan;97:120-
127. doi: 10.1016/j.ijmedinf.2016.09.014. Epub 2016 Oct 1. PMID: 27919371; PMCID: PMC5144921.