A proposed combined Cornell-Sokolow modeling approach to determine the left ventricular hypertrophy accurately based on multi-domain information processing

Muhtasheem Ajwad Mahin\textsuperscript{1} and Mohiuddin Ahmad\textsuperscript{2}

\textsuperscript{1,2}Dept. of Electrical and Electronic Engineering, Khulna University of Engineering & Technology (KUET), Khulna-9203, Bangladesh

E-mail: muhtasheem1mahin@gmail.com, ahmad@eee.kuet.ac.bd

Abstract. Left Ventricular Hypertrophy (LVH) is associated with cardiomyopathy and many other heart diseases. In this paper, the authors have proposed a combined Cornell-Sokolow (CCS) methodology to achieve significant improvement in detecting LVH by ECG/EKG with multi-domain analysis. Considering two famous voltage criteria simultaneously helps prevent the concealing of LVH in the patient’s data. The advancement of multi-domain analysis gives access to reach the algorithm from any common and cost-effective platform. This multi-domain analysis consists of both image and signal processing of the ECG data. In a MATLAB environment, these processes are quite accurate and optimized. With the help of the CCS model, the outcome of this research is satisfactory, both in terms of feature detection and LVH observation.

Keyword- LVH, Cornell-Sokolow (CCS) model, Multiple leads analysis, Image processing, Image to matrix processing, Baseline wandering, Signal processing, MATLAB platform.

1. Introduction
To the report of the world health organization in 2016, around 56.6 million people had died from Ischemic heart disease, which is also associated with stroke. In the fast-changing environment and our aloof lifestyle is significantly increasing in heart diseases. Till now, cardiovascular diseases are the fundamental causes of death around the world [1]. The research is focused on developing an electrical process capable of effectively detecting cardiac disease more accurately. This research follows to extract information from the electrocardiogram (ECG) to pre-process and process to accomplish decisions as accurately as possible. The main feature is to understand and study LVH (Left Ventricular Hypertrophy) [2] by the ECG/EKG multi-lead analysis with the Combine Cornell-Sokolo (CCS) model. There are many kinds of cardiac diseases like Congenital heart disease, Coronary artery disease Arrhythmia, Myocardial infarction, Hypertrophic cardiomyopathy, and many others [3]. The human heart is segmented into four different parts. The enlargement of the left ventricle is known as Left Ventricular Hypertrophy. It causes when the heart left ventricle (known as the main pumping chamber) wall thickened up to 10-12mm. This occurs to supply the excessive blood pressure that has been needed due to some cardiac abnormalities. This cardiac abnormality leads to serious problems like chronic pressure, coronary artery disease, heart block, heart failure, and even stroke. Many of the cardiac diseases can find out by a proper understanding of LVH. By analysing the ECG signal, this LVH can be detected. When the myocardium is hypertrophied, its large mass causes electrical activation, so the QRS complex’s amplitude, representing ventricular depolarization, is increased. But the problem is that the
presents of LVH can't be detected due to signal interferences, sometimes other cardiac activities that interrupt or eradicate the information from the signal. Even with the presence of LVH and no system error, the LVH might not be detected. A combined Cornell-Sokolow (CCS) modeling of an electrical process can solve this problem. This research has been developed a bunch of algorithms both in image and signal processing to identify the presents of LVH in patient's data.

In the last few decants, many research articles have been published to detect LVH following various methods. Even new technology has been introduced, such as magnetic resonance imaging (MRI) [4], cardiovascular magnetic resonance (CMR), and some other different voltage criteria method. But our focus is to develop a more accurate method suitable for common and cost-effective medical equipment. This research has included the two most commonly used methods, Cornell voltage Criteria, and Sokolow-Lyon voltage Criteria, to detect LVH by feature detection. Features are represented by P, Q, R, S, T waves from the ECG signal. These waves carry most of the information about heart disease by their height, depth, and shape. The doctors mostly use both of these voltage criteria for standard LVH study. The proposed CCS model allows us to check most of the complex cases in a short time. Which will be very convenient for the doctors. The information source of this research is based on a database that has provided by PhysioNet.Org [7] and some online/offline ECG reports.

2. Method of LVH Detection

2.1. Cornell Voltage Criteria
The Cornell voltage criterion is one of the most used criteria among the doctors in the study of LVH. This criterion suggests if the sum of the R wave in aVL and the S wave in V3 is greater than 28 millimeters in males or greater than 20 mm in females, LVH is present. Point to be noted that the 1mm square box is equal to 0.1mV and 0.04sec in the time scale of the standard ECG report.

2.2. Sokolow-Lyon Voltage Criteria
The Sokolow-Lyon, another famous method which is also very popular among doctors. In some studies, it has been said Sokolow-Lyon criterion is more accurate than the Cornell voltage criteria [5]. In these criteria, if the sum of S wave in V1 or V2 and R wave in V5 or V6 is greater than 35 mm, LVH is present. So this analysis will include V1, V2, V5, and V6’s lead signal.

2.3. A Combine Model Proposition
When the research begins in the first place, the authors thought that if a patient were subjected to LVH, they would simultaneously match both of the voltage criteria [6]. As the research is going on, several cases were found where the patients confirm LVH positive, not showing both of the criteria but only following one criterion. This finding helps our research to stay on the right path. So our algorithms have been designed to process and analyze a Combine Cornell Sokolow methodology. Because each of the voltage criteria inspects different signals, they are entirely independent of one another. We design an algorithm that considers both criteria for a patient’s data. We like to name it the CCS model.

3. Multi-Domain Dependency
It is conventional to work with direct signal (both time and frequency domain) in the bioelectronics research field. Signals are accurate, easy to process, less complex, rather than converting the image to the signal process. Initially, this research was based on signal processing. However, it faces a lack of data availability when it is necessary to analyze multiple leads data. ECG reports are easy to collect from local hospitals and diagnostic centers with 12 lead signals. So the image processing technique was introduced. It helps to analyze data is not only for this research but also for any other future research in this kind of situation.

4. Pre-processing Of ECG Data
The preprocessing is essential for the computation because data contain some misleading information. Though the input data can be signal or image, data can be direct from the 12 lead ECG signal, or it can
be any type of image file, just like a scanned copy of the available ECG report. Before processing, it should maintain a standard for the algorithm. The algorithm itself is limited to some parameters.

4.1. ECG Lead Signal Pre-processing.
Most of the time, the signal contains noise or different shapes; even it has some power line harmonics. The algorithm cannot detect the specific information with this noise. So, the processing begins with 15 lead ECG signals from the PhysioNet.Org where the database is named PTB Diagnostic ECG Database (PTBdb) [7]. The database contains about 290 patient’s information. The reason for choosing this database is that it has 15 leads of data set for every subject in its collection. The preprocessing includes:
(i) Filtering the Signal (high-frequency component removed, smoothing)
(ii) Base Line Wandering.
Filtering is an important preprocess for any signal processing. These ECG signals contain lots of high-frequency components. So before feeding the signal into the algorithm, it needs to be free from any interference. A low pass filter can be applicable. Figure 1 shows the filtering that we have applied in the preprocessing for aVL lead. For our easy computation, we use Savitzky-Golay filtering. It is a polynomial filter. It can provide a satisfactory result for ECG signal filtering.
Baseline wandering is another complicated and challenging task for ECG signal to process. When the raw signal has been received from the leads, the signals offend get distorted. The baseline can be shift due to positional displacement. In medical equipment has a primary baseline wandering mechanism. We have to perform baseline correction using a piecewise polynomial function for the Matlab environment.

![Raw aVL input](image1)

**Figure 1.** Filtered Signal and its Baseline Correction

When the signal is noise-free, we can use it for baseline wandering. In our algorithm, we have used pathfinding algorithms [8]. This algorithm is based on polynomial fitting.
This is the filtered signal of lead aVL in figure 1, it can be seen that the baseline wander has done. Any kind of amplitude or feature detection in this preprocess has been avoided.

4.2. Image Data Pre-processing
Our research uses both kinds of information; the signal from the database as well as EGC reports as images. Similar to the signal, the image is used as input data. It may contain noises also, the alignment can be missing. The algorithm cannot detect the specific information with this limitation.
In ECG reports pre-processing, we can provide a scanned copy or a good resolution ECG report in Matlab. Most of the time it is available in red-yellow color, RGB code: RGB (225, 196, 188) and HEX
code is: #e1c4bc. This background is important for the color adjustment level in the algorithm. It is not necessary to exactly match the color, but it should be close enough around it.

![Image](image.png)

**Figure 2.** Selected Window before the sequential process

From figure 2, a window has been selected with a minimum of two peaks of R from V6. That completes our pre-processing. The algorithm will follow different steps on that selected window in the processing.

5. **Combine Cornell-Sokolow (CCS) Model**

As we have already discussed, the proposed research is built on the most useful two methods, Cornell voltage criteria and Sokolow-Lyon voltage criteria. So this time we developed a model that can represent the combined logic for these voltage criteria.

The user has to input the information about data in the algorithm, whether it is image or signal. If it is a signal, the algorithm will arrange a series of processes to extract the selected lead feature. We have the advantage of the PTB database because the lead sequence is the same for every record. Now for an image, the information inside the selected window follows a series of morphological operations. This operation has to be complete before computing for a new matrix. This new matrix represents the time domain signal from the image. So signals are ready from two different origins. Now remain part of the algorithm support all necessary calculations for the CCS model.

6. **CCS Model and the Proposed Algorithm**

As the algorithm follows the CCS model just after preprocess. It is a combination of two different domains of data processing. So for both image and signal processing, the design of the algorithm has completely different structures. We can discuss these two process briefly to understand its logical analysis.

6.1. **Lead Signal Processing.**

The signals are collected from physionet.org [7] and the database is named PTB. When the specific leads are selected, we can apply a function for detecting the peak. In the case of aVL, V5, and V6 we need to detect the R feature or known as an R wave, which is also the peak values for those lead’s signals. Now we can apply a building function with threshold voltage and minimum peak distance. In return, it allows us to identify the peak values in the signal. Every signal was recorded with a gain factor. This gain factor has been adjusted comparing with the graphical view for better computing. Most of the patients have 11 to 15 pulse in 10sec. For better accuracy, we have taken reading for the first 10sec. One thing must be noted; in aVL, V5, or V6 the highest peak is R wave in any heart condition. The function “Findpeaks” needs to adjust the threshold and minimum peak distance for detecting all other peaks in the total samples. Though the human heart rate is 60 to 100 bpm, the minimum peak distance should depend on sampling frequency and heart rate.
Figure 3 shows the detection of the R wave from aVL. Similarly, we have detected R wave also for V5 & V6. When it comes to detecting the S wave, we can simply invert the whole signal and use the same function. For LVH, the S wave in V1, V2, and V3 is the lowest point. So an inverted signal will contain the S feature as a positive peak [9].

6.2. Image Processing (ECG Report As Input)
As we already discussed, two peak waves need to fit into the selected window. The sequential process can be performed in that area of the ECG report. It has been called the sequential process because the whole process is a series of morphological operations and some logical data analysis.

Figure 4 shows a sequential morphological operation in the selected window. From left to right, we can describe the operation as below: At first, the RGB image was taken with a minimum of two peaks, figure 4(A). The background color range should be near RGB (225, 196, 188). The RGB image is converted to an HSV image figure 4(B). The HSV image is again converted to the BW image by utilizing range selection from HSV values, which generate the third image in the bottom left corner, figure 4(C).
Finally, noise is removed by connected components function with factors 6 to 8 depending on the HSV ranging, figure 4(D).

When the black and white image is ready, it is time to image matrix conversion [10] [11]. A flow chart can help with the idea. From the BW image, we separated the coordinate for both black and white pixels. Then the coordinates are put into different two matrixes. This matrix contains some repetition of data. The design algorithm checks and rechecks those data. Having relevant data generates a final matrix. As the output matrix is a 2D signal, we have to ensure a single value for each sample. After the processing is done, data is ready to be plotted.

**Figure 5.** HSV to matrix block in detail.

Figure 5 represents the HSV to Matrix sequential block. This block diagram shows the back-end operation of the new matrix [12].

**Figure 6.** Generated new matrix, Detected R Feature of V6.
Figure 6 shows the graphical presentation of the new matrix from Fig. 4. There is some point to be noted. The peak value of this convention can be a shift in the few millivolt range. So we have to introduce Precision Adjust Factor (PAF) into the algorithm. That scale factor adjusts the amplitude of the voltage. From Fig. 2, we can see the RV6 wave is about 1.800mV. So from the image to matrix processing, we get around 1.8306 mV, close enough for computing. The second thing that needs to consider, the image has no physical voltage level. So there is no zero potential. In the new matrix, all the points are scale down with a value limit. This value has been decided by the subtraction of all individual samples from their median. In lower to a higher order, most of the values are near the baseline. This technique helps to get zero potential or baseline with easy computation.

When features are detected, calculations can be done for the CCS model. Here we are using the data from signal only. Because we cannot be able to find any LVH positive case in our collected ECG report. The algorithm performs quite accurately, with 92.58% average accuracy. From the PTB database, patient number 58, 52 years old female, recorded her data in 1991. The data is lodges as “s0216lrem” in the .mat file. Detected features are:

| Lead Signal's Feature | Voltage(mV) | Standard Height (mm) |
|-----------------------|------------|----------------------|
| aVL(R)                | 1.073mV    | 10.73mm              |
| V3(S)                 | 1.3436mV   | 13.436mm             |
| V1(S)                 | 1.9909mV   | 19.909mm             |
| V6(R)                 | 0.2450mV   | 2.45mm               |

Table 1 describes about the detected values. Now we can perform the calculation for CCS model, 10.73+13.43= 24.16 mm > 20mm – LHV Positive by Cor. Voltage Criteria 19.9+2.45= 22.35 mm < 35 mm – LHV Negative by Sokolow-Lyon Criteria

So, the patient is subjected to LHV by Cornell voltage criteria though it is not following by the Sokolow-Lyon method. In the calculation, 10.73+13.43= 24.16 mm, and this is greater than 20mm for women. So the CCS model suggested that patient no: 58 were subjected to LVH. This result shows the accuracy of detection has been increasing if the model is combined. Regarding system accuracy, we can be narrowing down our results of feature detection.

| CCS Model | Picante data | aVL | S in V3 | S in V1or V2 | R in V5 or V6 |
|-----------|--------------|-----|---------|--------------|--------------|
| Signal    | s0181lrem (50)| 96.93% | 96.04% | 88% (v2)    | 90.84% (v6)  |
| Signal    | s0216lrem (58)| 97.55% | 98.6%  | 98.53% (v2) | 98.08% (v6)  |
| Image     | 2x_cf.jpg    | 98.98% | 81.7%  | 91.35% (v1) | 98.31% (v6)  |

Table 2 shows the accuracy in detecting features by the algorithm. We have inculcated two signals from the PTB database and compare the detected value from the graphical plot by an expert physician [13]. The patient number has been indicated with data. We have also compared data accuracy for image ECG report input. However, we face a lack of right resolution images, so we have included one result here.

In this study, we have identified significant points that we need to consider. We have gathers many ECG report where signals are overlapping with each other. Especially in the case of LVH, the R peaks are extremely tall. So that they sometimes overlap with another signal at the top.
7. Conclusion
In this paper, the authors had combined two well-known methods of Left Ventricular Hypertrophy studies. Individualistic voltage criteria offer a better LVH understanding of this model. So, the proposed CCS model has given a promising result. On the other hand, multi-domain information processing gives access to the use of the CCS model with any kind of information. Even with a simple ECG report, a patient can analyze his or her cardiac abnormalities very easily. Utilizing proper optimization and machine learning, the CCS model can lead to a fast, more accurate, and worldwide heart disease diagnostic network.

8. References
[1] who. int by Newsroom/ Factsheet/ Detail (2018, 24th of May) “Top 10 Causes of Death” can be found at https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death
[2] Eskerud I, Lønnebakken M T, Gerds T March 2019 Left ventricular hypertrophy contributes to Myocardial Ischemia in Non-obstructive Coronary Artery Disease (the MicroCAD study) International Journal of Cardiology 286 pp 1–6
[3] Simone G, Palmieri V, Bella J N, Celentano A, Hong Y, Oberman A, Kitzman D W February 2002 Association of left ventricular hypertrophy with metabolic risk factors, the HyperGEN study Journal of Hypertension 20 no. 2 pp 323–331
[4] Habermann C R, Weiss F, Lilje C, and Weil J March 2004 MRI of Congenital Heart Disease in Childhood RöFo - Fortschritte auf dem Gebiet der R 176(2) pp 191–9
[5] Su F, Li Y, Lin Y, Lee C and Wang C June 2017 A comparison of Cornell and Sokolow-Lyon electrocardiographic criteria for left ventricular hypertrophy in a military male population in Taiwan: the Cardiorespiratory fitness and Hospitalization Events in armed Forces study Cardiovascular Diagnosis and Therapy 7 no. 3 pp 244–251
[6] Rodrigues S L, D’Angelo L, Pereira A C, Krieger L E and Mill J G January 2008 Revision of the Sokolow-Lyon-Rappaport and Cornell voltage criteria for left ventricular hypertrophy Arquivos Brasileiros de cardiologia 90 no. 1 pp 46–53
[7] Physionet by National Institute of General Medical Sciences (2020, June 22), PTB database can be found at https://archive.physionet.org/cgi-bin/atm/ATM
[8] Rahman M A, Milu M M H, Anjum A, Khanam F and Ahmad M 2017 Baseline wandering removal from ECG signal by wandering path finding algorithm 3rd Int. Conf. on Electrical Information and Communication Technology (EICT)
[9] Diker A, Avci E, Gedikpinar M May 2017 Determination of R-peaks in ECG signal using Hilbert transform and pan-Tomkins algorithms 25th Signal Processing and Communications Applications pp 1–4
[10] Qin Q, Yinggao Y J, Chengyu L, January 2019 An adaptive and time-efficient ECG R-peak detection algorithm International Journal of Recent Technology and Engineering (IJRTE) 7 pp 1–14
[11] Patil R and Karandikar R October 2017 Digitization of Documented Signal Using Vertical Scanning National Seminar on Information Systems and Technology (SISPOTEK) 1 pp 239–242 ISSN 2597-3584.
[12] Priyanka R K, Shrivastava, Panbude S, Narayanan G November 2014 Digitization of ECG Paper Records Using MATLAB International Journal of Innovative and Exploring Engineering (IJITEE) 4(6) pp 1–3 ISSN: 2278-3075.
[13] Healio.com by SLACK Incorporation (2020, June 22) “Left Ventricular Hypertrophy (LVH) ECG Review” can be found at https://www.healio.com/cardiology/learn-the-heart/ecg-review/ecg-topic-reviews-and-criteria/left-ventricular-hypertrophy-review.