**HIGH PRECISION DIGITIZATION OF PAPER-BASED ECG RECORDS: A STEP TOWARD MACHINE LEARNING.**

**ABSTRACT**

 **Introduction:** The electrocardiogram (ECG) plays an important role in the diagnosis of heart diseases. However, most patterns of diseases are based on old datasets and stepwise algorithms that provide limited accuracy. Improving diagnostic accuracy of the ECG can be done by applying machine learning algorithms. This requires taking existing scanned or printed ECGs of old cohorts and transforming the ECG signal to the raw digital (time (milliseconds), voltage (mill volts)) form.

**Objectives:** We present a MATLAB-based tool and algorithm that converts a printed or scanned format of the ECG into a digitized ECG signal.

**Methods:** 30 ECG scanned curves are utilized in our study. An image processing method is first implemented for detecting the ECG regions of interest and extracting the ECG signals. It is followed by serial steps that digitize and validate the results.

**Results:** The validation demonstrates very high correlation values of several standard ECG parameters: PR interval 0.984 +/- 0.021 (p-value < 0.001), QRS interval 1+/- SD (p-value < 0.001), QT interval 0.981 +/- 0.023 p-value <0.001, and RR interval 1 +/- 0.001 p-value <0.001.

**Conclusion:** Digitized ECG signals from existing paper or scanned ECGs can be obtained with more than 95% of precision. This makes it possible to utilize historic ECG signals in machine learning algorithms to identify patterns of heart diseases and aid in the diagnostic and prognostic evaluation of patients with cardiovascular disease.

**Index Terms—** Electrocardiogram, Digitization, Matlab Tool, Image Processing.