Electrocardiographic and Vectorcardiographic Markers of Cardiac Involvement in Systemic Lupus Erythematosus: A Machine Learning Approach

Alejandro Pérez-González[1], Elisa Ramírez[1], Francisco Castells[1], M Soyfoo[2], Rubén Casado-Arroyo[3], José Millet[1]

[1]ITACA Institute, Universitat Politècnica de València, Valencia, Spain
[2]Rhumatologie, Hôpital Erasme, Brussels, Belgium
[3]Electrophysiology Lab, Hôpital Erasme, Brussels, Belgium

Abstract

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease frequently associated with cardiovascular involvement, often subclinical and difficult to detect. This study investigates differences in cardiac electrical activity between SLE patients and healthy controls by analysing standard electrocardiogram and reconstructed vectorcardiographic. ECG recordings from 64 SLE patients and 64 matched controls were processed. A statistical comparison using the Mann-Whitney U test was conducted to identify features with significant differences between groups, with T loop speed and shape among the most discriminative. The set of derived features was used for dimensionality reduction via principal component analysis (PCA) and four machine learning models were applied. The models achieved high performance, with random forest reaching up to 96% accuracy. These findings suggest that ECG and VCG analysis, combined with machine learning, could aid in the early detection of cardiac involvement in SLE.

1. Introduction

Systemic lupus erythematosus (SLE) is a chronic, multisystem autoimmune disease characterized by the production of autoantibodies and immune complex deposition, leading to inflammation and tissue damage across various organs [1]. Among its diverse manifestations, cardiovascular involvement stands out as a significant contributor to morbidity and mortality in SLE patients. Cardiac complications can encompass pericarditis, myocarditis, valvular abnormalities, and accelerated atherosclerosis, often presenting subclinically and remaining undetected until advanced stages [2].

Although there is strong evidence supporting the diagnosis of lupus through serological and immunological tests—including C-reactive protein (CRP), rheumatoid factor (RF), and antineutrophil cytoplasmic antibodies (ANCA) [3]—the relationship between lupus manifestations and cardiac electrical activity remains poorly characterized. Given its simplicity, non-invasiveness, and low cost, ECG represents a potentially valuable method for monitoring cardiac function and detecting early signs of cardiovascular involvement in SLE.

Traditional diagnostic modalities for cardiac assessment in SLE include echocardiography, which, while informative, may not detect subtle electrophysiological alterations. ECG, owing to its non-invasive nature and widespread availability, offers a valuable tool for monitoring cardiac electrical activity. Studies have identified common ECG abnormalities in SLE patients, such as ST-segment changes, T-wave inversions, and prolonged QT intervals[4].

Advances in computational methods have positioned machine learning (ML) as a promising tool to enhance ECG-based diagnostics. By detecting complex patterns in ECG signals, ML enables the early identification of cardiac abnormalities.

Early detection of subclinical cardiac involvement in SLE remains a major challenge. Subtle electrophysiological changes often go unnoticed with standard diagnostic tools. The vectorcardiogram (VCG), which can be reconstructed from the ECG [5], provides additional spatial information that is key to detecting these subtle alterations. This study applies both ECG and reconstructed VCG features, combined with ML algorithms, to identify distinctive electrical patterns in SLE patients. The goal is to improve early detection and enable interventions that could enhance patient outcomes.

2. Materials

Two datasets were used in this study, comprising a total of 128 subjects. The first dataset includes 64 patients diagnosed with SLE, collected from the Erasme Hospital in Brussels, Belgium. For each subject, standard

12-lead ECG recordings were available, acquired using the MAC16 electrocardiograph system, along with demographic data including age and sex.

The control group consists of 64 subjects obtained from an open PhysioNet database. These subjects were selected to match the lupus cohort in terms of age and sex, ensuring demographic comparability between the two groups.

3. Methods

3.1. Preprocessing and Feature Extraction

The data was filtered and explored using Python-based signal preprocessing pipelines and analyzed via the Harvard BRAVEHEART open-source platform [6]. Multiple key parameters were extracted and compared between the lupus and control groups, including morphology, amplitudes, and areas of the median 12-lead ECG beat; propagation patterns, vectors, and angles from the median VCG; and the VCG velocity profile. The feature extraction process is illustrated in Fig. 1, highlighting the transformation from raw ECG signals to derived VCG parameters. Examples of preprocessed signals from lupus patients and healthy controls are shown in Fig. 1. To statistically assess these differences, the Mann-Whitney U test was applied across all extracted features.

3.2. Dimensionality Reduction

Principal component analysis (PCA) was applied to reduce the feature space prior to classification. A total of 22 principal components were selected in which the variance explained was >90% of the total variance. The first principal component (PC1) was largely influenced by features derived from the QRST and T complexes, particularly in the VCG (e.g the features representing the area under the QRST complex in Lead II and the Y-component VCG area yielded a combined loading score of $\mathbf{0.3}$).

3.3. ML Models and Hyperparameters

Of the total data, a subdivision was made into 80% for traning and 20% for testing. Of the 80% for training, 20% was reserved for validation and hyperparameter tuning. A set of four supervised ML models were implemented and evaluated on the PCA-reduced dataset. The hyperparameters were tuned within a comprehensive range using grid search. The models and corresponding hyperparameters are summarized in Table 1.

The full set of extracted and PCA-reduced features was used to train and test four classification models—Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and Decision Tree—to assess their discriminative potential.

Table 1: Model hyperparameters. Selected values are marked with an asterisk (*).

Model	Hyperparameters tested
Logistic Regression	C: 0.01–**1***–10 penalty: **l2*** solver: **lbfgs***, saga
SVM	kernel: **linear***, rbf, poly C: **0.1***–10 gamma: scale, auto (**N/A***)
Random Forest	n_estimators: **50***–200 max_depth: **None***–20 max_features: **sqrt***, log2
Decision Tree	max_depth: **None***–20 min_samples_split: **2***–10 criterion: gini, **entropy***

3.4. Performance Evaluation

Confusion matrices were computed for each model. Metrics such as accuracy (ACC), sensitivity (SE), and specificity (SP) were calculated to assess classification performance.

4. Results

4.1. Statistical Analysis

Key ECG and VCG features were statistically compared between lupus patients and healthy controls. The most relevant features, selected according to the Mann–Whitney U test (p ¡ 0.001), are shown in Fig. 2 for visual comparison. These include:

– **T median speed:** propagation speed across the T wave loop,

– **VCG T length:** loop length in VCG space,

– **VM tpeak_tend:** interval between the T wave peak and end in the median VM lead,

– **SAI Z:** area under the absolute Z-component of the QRST complex.

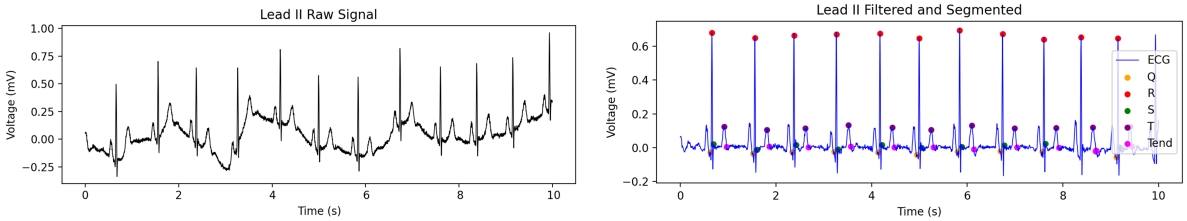
The boxplots in Fig. 2 reveal clear distributional differences between the two groups, suggesting these features may serve as potential biomarkers.

4.2. ML Models

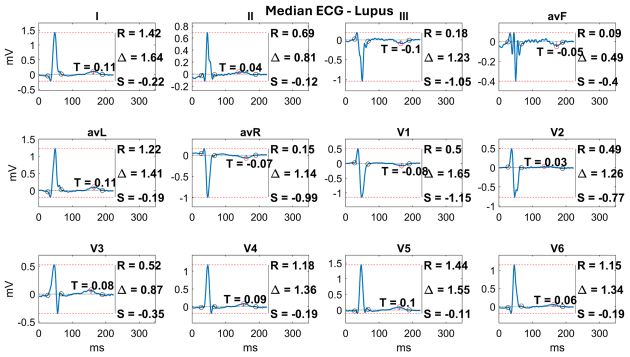
The classification outcomes are visualized through confusion matrices in Fig. 3.

Table 2 summarizes the overall performance of each model using standard evaluation metrics: ACC, SE and SP. LR and SVM demonstrated consistent performance with accuracy values of 0.91. RF achieved the highest ACC (0.96) and perfect SP, while the Decision Tree model exhibited lower overall performance, particularly in SP.

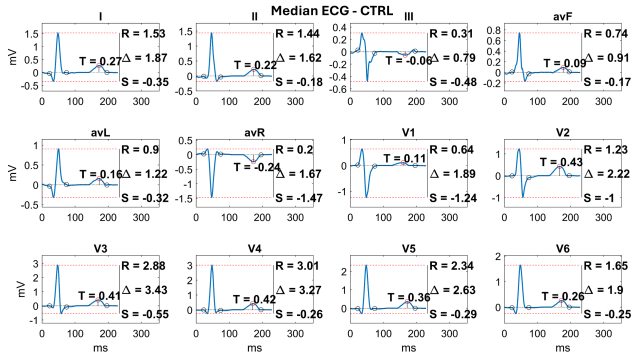
This study demonstrates that VCG and ECG parameters can effectively distinguish between lupus patients and



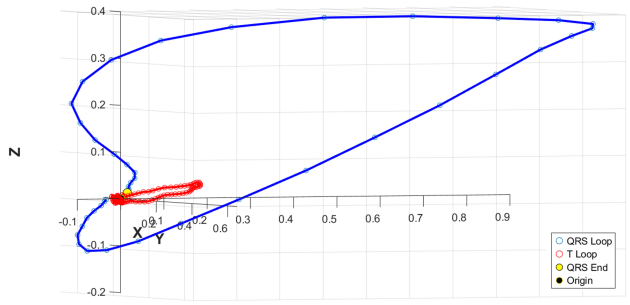
(a) Lupus: Raw ECG, filtered and segmented



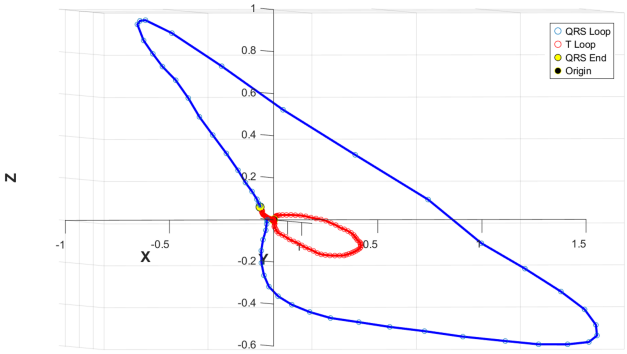
(b) Lupus: Median ECG beat morphology



(c) Control: Median ECG beat morphology



(d) Lupus: Segmented VCG loop (axes in mV)



(e) Control: Segmented VCG loop (axes in mV)

Figure 1: ECG signal processing pipeline and morphology comparison between a lupus patient and a healthy control. (a) Raw ECG and filtered version with segmented beats. (b–e) show the median ECG beats and VCG loops derived from the same subjects.

Table 2: Performance metrics of the evaluated models on the test set.

Model	Accuracy	Sensitivity	Specificity
Logistic Regression	0.91	0.91	0.92
SVM	0.91	0.91	0.92
Random Forest	0.96	0.91	1.00
Decision Tree	0.78	0.91	0.67

healthy controls, revealing underlying cardiac alterations associated with the disease. Among the most discriminative features were T wave loop speed, T loop length, T peak-to-end interval, and the area under the Z component of the QRST complex (SAI Z), all of which showed statistically significant differences between groups.

From a physiological perspective, the VCG visualizations illustrate that lupus patients tend to exhibit faster propagation during the T wave, as evidenced by narrower T loops in the vectorcardiogram, as shown in Fig. 1. This may reflect altered repolarization dynamics, potentially linked to autonomic dysregulation commonly observed in SLE. Additionally, the reduced T loop length and altered VM $t_{\text{peak-end}}$ intervals further support the hypothesis of disrupted repolarization processes in lupus.

The classification models, particularly LR, SVM, and RF, showed strong discriminative performance in distinguishing lupus cases from healthy controls, each achieving

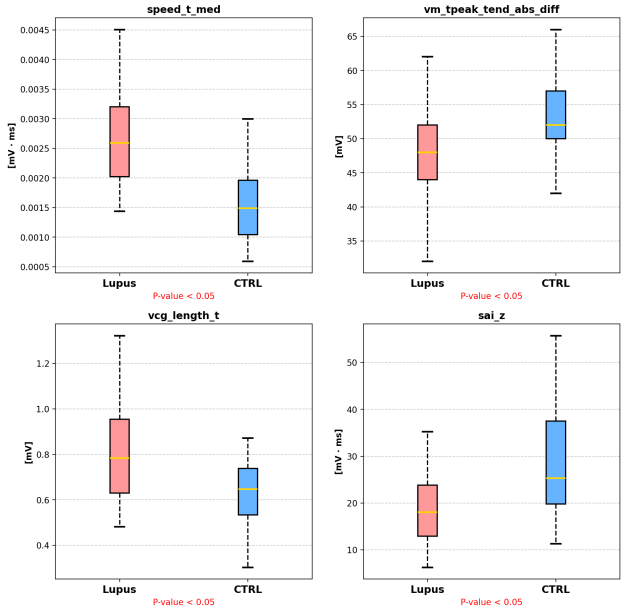


Figure 2: Boxplots of selected ECG and VCG features showing significant differences between lupus and control groups.

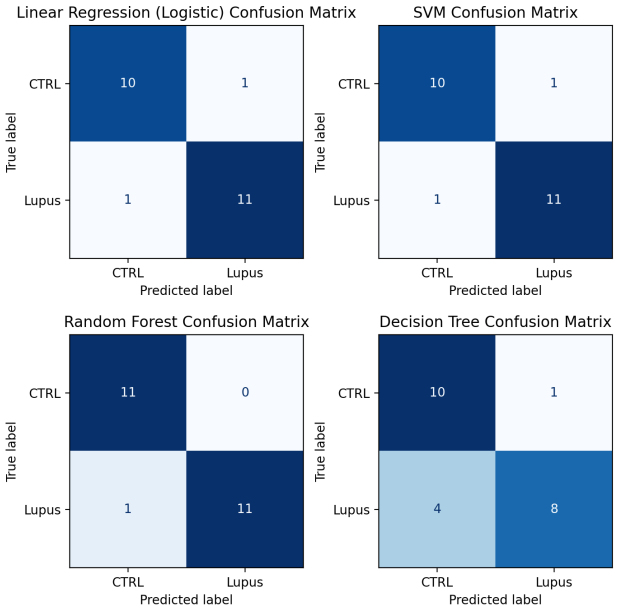


Figure 3: Confusion matrices for (A) Logistic Regression, (B) Support Vector Machine (SVM), (C) Random Forest, and (D) Decision Tree.

high accuracy, sensitivity, and specificity. Notably, the RF model reached perfect specificity, correctly identifying all control subjects while maintaining strong sensitivity, indicating zero false positives within the control group. In contrast, the Decision Tree model misclassified a substantial number of lupus cases as normal, suggesting that its predictive capacity is insufficient for this task

The main limitation is that the dataset is relatively small, with only 64 lupus patients and 64 healthy controls, which limits the generalizability of our conclusions. The sample size is insufficient to definitively confirm the proposed hypotheses or to establish these electrocardiographic features as reliable biomarkers for cardiovascular involvement in lupus. Therefore, while the initial findings are promising, further studies with larger, more diverse populations are required to validate these markers and to assess their clinical applicability.

5. Conclusion

This study highlights significant differences in both ECG and reconstructed VCG parameters between patients with systemic lupus erythematosus and healthy controls. In particular, features associated with T-wave morphology, propagation dynamics, and ventricular vector loop characteristics demonstrated strong discriminative power. These alterations suggest that SLE may affect ventricular repolarization, potentially due to autonomic dysregulation. In particular, the increased T wave loop speed and shortened T loop length observed in lupus patients may reflect abnormal repolarization pathways.

Machine learning models trained on these features—especially LR, SVM, and RF—were able to classify lupus and control subjects with high accuracy (0.91 and 0.96), demonstrating the potential of automated ECG/VCG analysis in early cardiovascular screening for lupus patients.

Future work will involve expanding the dataset and exploring additional clinical parameters that may help in refining these findings.

References

[1] Tsokos G, et al. Sle. N Engl 2011;.
[2] Miner J, et al. Cardiac manifestations of s. lupus erythematosus. Rheum Dis Clin N Am 2014;.
[3] Ippolito A, et al. Autoantibodies in sle: historical vs current seropositivity assessment. Lupus 2011;20(3):250–255.
[4] Hu Z, et al. Prevalence and associated factors of ecg abnormalities in sle: a ml study. Arthritis Care Res 2021;.
[5] Ramírez E, et al. Novel synchronization method for vcg reconstruction from ecg printouts: a comprehensive validation approach. Biomed Signal PC 2024;91:106027.
[6] Stabenau H, Waks J, et al. Braveheart: Open-source software for ecg and vcg analysis. Comput Methods Programs Biomed 2023;.

Address for correspondence:
Alejandro Pérez-González; ITACA Institute, Universitat Politècnica de València, Valencia, Spain; alejandropg002@gmail.com