Diagnostic ECG Classification Using Machine Learning with 3D Vectorcardiographic Data

Lucenildo S Cerqueira[1], Jurandir Nadal[2]

[1]Substitute Lecturer, Universidade Federal do Rio de Janeiro (UFRJ), Brazil
[2]Professor, Biomedical Engineering Program, UFRJ, Brazil

Abstract

This study evaluates computational intelligence methods for classifying 3D vectorcardiographic (3D-VCG) data into normal (NORM), myocardial infarction (MI), and ST-T change (STTC) categories. Unlike conventional ECG, 3D-VCG captures spatial cardiac depolarization, offering enhanced diagnostic insights. Using the PTB-XL Database, 12-lead ECGs were transformed into XYZ leads via the Kors matrix. A balanced dataset of 7,146 patients (2,382 per class) was used to train and test three models: a multi-layer perceptron (MLP), a convolutional neural network (CNN), and a long short-term memory (LSTM). Each model processed 350 samples per lead. MLP achieved the highest test accuracy (99.80%), followed by LSTM (89.00%) and CNN (83.26%). A 10-fold cross-validation on LSTM yielded an average accuracy of 94.0 ± 1.8%, though it did not surpass MLP. These results show the potential of 3D-VCG and machine learning for automated MI and STTC detection.

1. Introduction

The electrocardiogram (ECG) is a non-invasive tool for assessing the overall patient's heart condition and is the first-line test for diagnosing cardiac disease (CVD) [1]. The 3D vectorcardiogram (VCG) signal adds value to ECG analysis by providing additional information and enabling the calculation of parameters that cannot be derived from separate ECG leads [2]. The VCG represents the sum of all instantaneous electrical vectors generated by myocardial cells in the heart and provides a multidirectional view of cardiac electrical activity in space and time [3].

This tool enables the development of various markers, such as the assessment of ventricular repolarization heterogeneity, which results from intercellular differences in depolarization times and action potential morphology [4]. Furthermore, a recent study showed that QT dispersion is mainly determined by T-loop morphology, as reflected in T-loop amplitude and width. In contrast, an older study reported a widened QRS-T angle in patients with left ventricular failure [5]. The QRS-T angle reflects the deviations between ventricular depolarizations. Spatial and frontal QRS-T angles are two different ways to measure the QRS-T angle [5].

However, these previously developed VCG morphology descriptors are insufficient to fully characterize the complex three-dimensional morphology of the VCG loop [5]. Analysis of the QRS loops in VCG morphology can help define the abnormal electrophysiological substrate in patients with life-threatening ventricular arrhythmias [5]. The morphology of the VCG loop can be characterized by the direction and amplitude of the initial instantaneous [5] and maximum peak and average spatial vectors of the loop [5]. In addition, deep learning-based artificial intelligence (AI) algorithms have recently achieved cutting-edge performance in multiple domains [6]. An advantage of deep learning is the automatic learning of features and relationships from specific data, without domain knowledge [6].

This study investigated deep learning-based AI algorithms to detect myocardial infarction (MI) and ischemic ST-T changes (STTC) using 3D VCG loop morphology and compared their performance with that of conventional methods reported in the literature. The approach consists of an MLP with three layers of nodes and backpropagation for training, a classical supervised learning method [7]. The convolutional neural network (CNN) was another model studied. CNN is a nonlinear statistical model that attempts to identify optimal linear combinations of the input variables and then model the result as a nonlinear function of these covariates [7]. This deep learning framework can distinguish data that is not linearly separable. Another model was the long short-term memory (LSTM) [7-8], which has been used for ECG signal classification [7-8] and has emerged as a relevant approach in recent deep-learning studies.

The objective of this study is to compare the performance of MLP, CNN, and LSTM classifiers for separating the NORM, MI, and STTC groups using XYZ ECG coordinates as input features. Thus, the study investigates the ability of XYZ ECG coordinates to discriminate between normal patients and those with ischemic disease or myocardial infarction using deep

 ISSN: 2325-887X DOI: 10.22489/CinC.2025.185

learning models.

2. Materials and Methods

2.1. Dataset

The PTB-XL dataset is publicly available in the Physionet repository [9-10]. The research procedures were conducted in accordance with the Helsinki Declaration [11]. The dataset comprises 21,837 clinical 12-lead 10-second ECG records from 18,885 patients, 52% male and 48% female. The data were organized hierarchically into five coarse superclasses (NORM: normal ECG, CD: conduction disturbance, MI: myocardial infarction, HYP: hypertrophy, and STTC: ST-T changes) [11]. Only the classes NORM, MI, and STTC were investigated, with 2,382 patients randomly selected per group (7,146 subjects total). The PTB-XL has a rich set of ECG annotations and additional metadata, making the dataset an ideal resource for training and evaluating machine learning algorithms [11]. Training and testing subsamples were balanced, with no significant differences in clinical or ECG characteristics.

2.2. Pre-Processing

Twelve-lead ECG signals were low-pass filtered with a 2nd-order Butterworth filter at 35 Hz. R-wave detection was performed on digitized ECG signals using the Pan & Tompkins algorithm [12]. A software implemented in Python 3.12 [13] was developed to perform ECG and 3D vectorcardiographic analysis.

2.3. Data Analysis

The Kors matrix was used to transform 12-lead ECG into XYZ ECG coordinates for all ECG signals [14]. After that, each XYZ ECG coordinate was averaged, considering the R-wave as a reference. The XYZ coordinates (X, Y, Z, superclass) were used for analysis. The data were organized in a matrix (2,501,100 x 4), with each segment of 350 samples being from a different patient. For each group, 2,382 ECG signals were randomly selected. After, the dataset was split into two non-overlapping sets: training (70%; n = 5,002) and testing (30%; n = 2,144).

For the MLP, the data were organized in a matrix (7,146 x 1,050) with (X, Y, Z, superclass) side by side. The ECG data were also organized in a matrix (7,146 x 350 x 3) to match the expected [samples, timesteps, features] structure required by the LSTM and CNN algorithms for 3D classification. The MLP, LSTM, and CNN models were created using the Keras framework on top of TensorFlow 2.1.

All three of these models begin with an embedding layer. The LSTM network consisted of two bidirectional LSTM layers followed by two fully connected layers [15]. The CNN network consisted of a convolutional layer, an average pooling layer, a convolutional layer, a global average pooling layer, and two fully connected layers [15]. The number of layers and epochs were empirically found on a single training set with the original distribution of *NORM, MI, and STTC groups*. The models were trained multiple times at each step using the aforementioned temporary training sets with varying sizes and prevalences. A test run empirically determined the number of epochs, yielding 20 for the MLP/LSTM/CNN models.

2.4. Performance Evaluation

Each model's performance was evaluated by assessing sensitivity and specificity. The predictive accuracies of the models were compared using the area under the receiver operating characteristic (ROC) curve (AUC). Since screening aims to detect all prevalent CVD cases, maximizing test sensitivity was prioritized. The CVD detection performance was confirmed in the independent dataset by measuring AUC and assessing the sensitivity and specificity of the selected threshold at the earlier step.

The sliding window approach and LSTM models were implemented using high-level libraries, including Keras, NumPy, Pandas, and Scikit-learn [13], and executed in Google Colab notebooks — a cloud-based platform built on Jupyter notebooks. This environment is integrated with Google Drive, providing free access to computational resources and easing model development and execution.

The age and VCG parameters were compared using a one-way ANOVA with a 95% confidence interval.

3. Results

The results from the comparison of VCG parameters in the three study groups are summarized in Table 1. There was no significant difference in age between the groups. However, significant differences were obtained for VCG parameters. An example of 3D VCG for NORM, MI, and STTC signals is presented in Figure 1. The spatial QRS-T angle, SVG, and EL-SVG presented larger values in patients with MI and STTC than in the NORM class ($p < 0.0001$).

A comparison of the prediction models' performance is shown in Table 2. The MLP model achieved the highest predictive accuracy across both the training and test samples. The CNN model showed a moderate predictive AUC, which was significantly lower than that of the MLP model with VCG input. The LSTM showed high predictive AUC only in the training set, which was considerably worse than that of the MLP model with VCG input.

One more test was performed on the LSTM model, using the K-fold approach with 10 folds. This approach improved performance, achieving an average accuracy of

94.0 ± 1.8%. The other models show intermediate accuracy values for a similar fit.

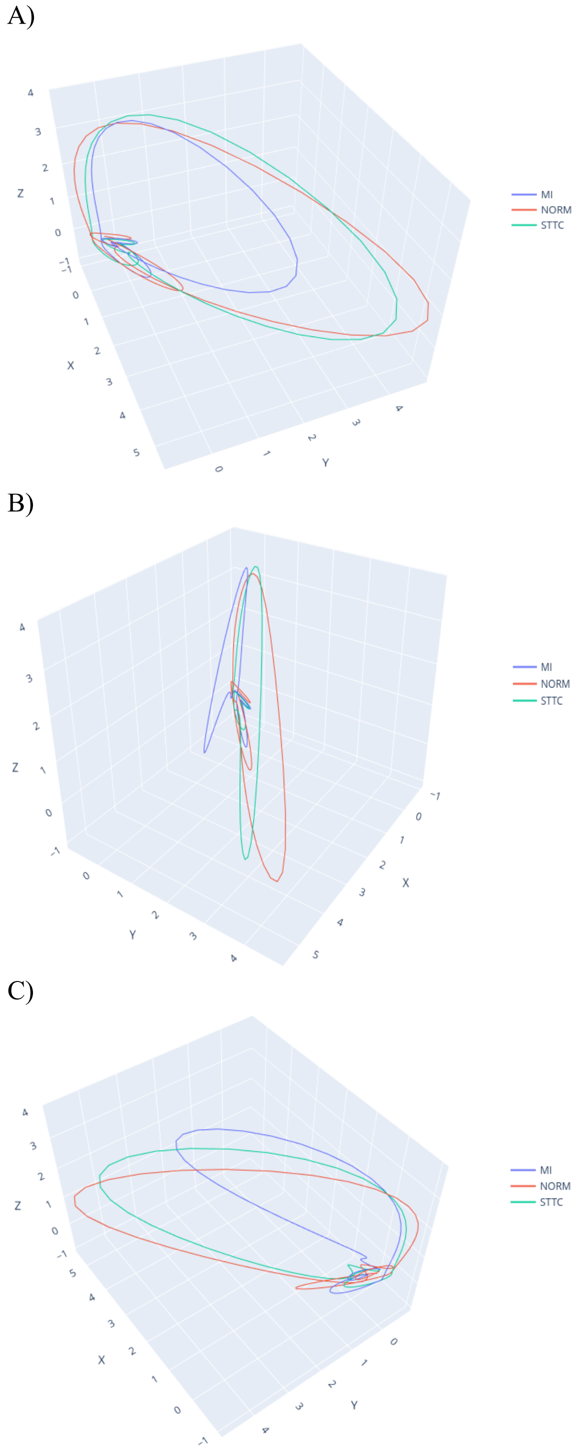

Figure 1. Average vectorcardiograms (VCGs) for patients with myocardial infarction (MI), ST-T change, and normal conditions, presented from different perspectives (A, B, and C).

Table 1. The average (standard deviation) for age, QRS-T angle, the magnitude of the spatial ventricular gradient (SVG), and the elevation angle of the SVG (EL-SVG)

	NORM	**MI**	**STTC**	**p-value**
Age	51.9 (17.2)	64.0 (12.7)	64.8 (14.4)	0.2341
QRS-T angle	55.9 (45.5)	86.4 (46.1)	89.1 (49.1) *	<0.0001
SVG	1226.6 (390.7)	1017.3 (351.2)	1083.1 (340.8) *	<0.0001
EL-SVG	66.6 (14.1)	72.2 (11.6)	65.1 (13.8) *	<0.0001

Table 2. Comparison of classifier performance indices

	MLP		**CNN**		**LSTM**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
AUC	0.99	0.99	0.96	0.83	0.99	0.89
Accuracy	0.99	0.97	0.89	0.69	0.97	0.77
Precision	0.99	0.96	0.89	0.69	0.97	0.77
Recall	0.99	0.99	0.89	0.68	0.97	0.77

4. Discussion

The purpose of the study was to compare the performance of the MLP, CNN, and LSTM classifiers in separating the NORM, MI, and STTC groups, using the XYZ ECG coordinates as input features. To the best of our knowledge, this study is the first to assess VCG morphology using a deep learning algorithm, thereby preventing any comparison with literature.

The classical MLP model produced the best results using three layers. An excellent performance was obtained with the 3D ECG samples without any segmentation on the ECG or fiducial point measurements. The results overcome earlier classification results for this type of ECG patterns [6-8][16].

However, CNN and LSTM models underperformed MLP, with comparable results. This cannot be used to claim that MPL is the best tool, since any classification model has a set of configuration parameters that could be further explored [16]. For example, rearranging the VCG data using a k-fold approach improved LSTM performance to values comparable to those of other methods in the literature [7][16].

In converting the 12-lead ECG to the VCG, much redundant information is removed. However, extracting XYZ coordinates mathematically may omit key features.

Even so, it enables us to achieve significant results in classifying heart diseases, such as MI and STTC [16].

The performance of VCG models was slightly worse in the test sample than in the training sample. Further studies are needed to compare the performance of machine learning models using the raw VCG signal as input versus the derived ECG metrics [16].

5. Conclusion

This study provides a practical framework for the automated detection of MI and STTC on short segments of XYZ ECG using machine learning models. Specifically, it can classify NORM, MI, and STTC with more than 94% accuracy and hence can be employed in clinical settings using the MLP and LSTM models. In future studies, the performance of the proposed model will be evaluated on different MI and STTC datasets. This may constitute an interesting line of future research, where the same approach should be extended to real-time data.

Acknowledgments

This study was partially supported by the Brazilian Agencies FAPERJ, CAPES, and CNPq.

References

[1] A. I. Zachi, D. M. Harmon, E. R. Behr, et al., "Application of artificial intelligence to the electrocardiogram." *Eur. Heart J.*, vol. 46, pp. 4717-4730, 2021.

[2] R. Jaros, R. Martinek, L. Danys, "Comparison of different electrocardiography with vectorcardiography transformations," *Sensors*, vol 14, Article. 3072, 2019.

[3] H. Yang, F. Leonelli, "Self-organizing visualization and pattern matching of vectorcardiographic QRS waveforms,". *Comput. Biol. Med.*, vol. 79, pp. 1-9, 2016.

[4] C. Voulgari, N. Tentolouris, "Assessment of the spatial QRS-T angle by vectorcardiography: Current data and perspectives." *Curr. Cardiol. Rev.*, vol. 4, pp. 251–262, 2009.

[5] J. D. Pollard, K. T. Haq, K. J. Lutz, et al., "Sex differences in vectorcardiogram of African Americans with and without cardiovascular disease: A cross-sectional study in the Jackson Heart Study cohort," *BMJ Open*, vol. 11, article e042899, 2021.

[6] C. Janiesch, P. Zschech, K. Heinrich, "Machine learning and deep learning," *Electronic Markets*, vol. 31, pp. 685–695, 2021.

[7] S. Savalia, V. Emamian, "Cardiac arrhythmia classification by multi-layer perceptron and convolution neural networks," *Bioengineering*, vol. 5, n. 2, article 35, 2018.

[8] A. Peimankar, S. Puthusserypady, "DENS-ECG: A deep learning approach for ECG signal delineation," *Expert Syst. Appl*., vol. 165, article 113911, 2021.

[9] P. Wagner, N. Strodthoff, R.-D. Bousseljot, et al., "PTB-XL, a large publicly available electrocardiography dataset," *PhysioNet*, https://doi.org/10.13026/kfzx-aw45, 2022.

[10] Goldberger, A., Amaral, L., Glass, L., et al. (2000). "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, n. 23, pp. e215–e220.

[11] P. Wagner, N. Strodthoff, R.-D. Bousseljot, et al., "PTB-XL: A large publicly available ECG dataset. *Scientific Data*. https://doi.org/10.1038/s41597-020-0495-6, 2020.

[12] J. Pan, W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng*., vol 32, pp. 230-236, 1985.

[13] *Python Language*, available at https://www.python.org/, accessed in April 2025.

[14] J. A. Kors, G. van Herpen, A. C. Sittig, J. H. van Bemmel "Reconstruction of the Frank vectorcardiogram from standard electrocardiographic leads: Diagnostic comparison of different methods," *Clinical Trial Eur. Heart. J*., vol. 12, pp. 1083-1092, 1990.

[15] F. Chollet, *Deep Learning with Python*, 1st ed., Shelter Island: Manning, 2017.

[16] B. U. Baloglu, M. Talo, O. Yildirim, et al., "Classification of myocardial infarction with multi-lead ECG signals and deep CNN," *Pattern Recognit. Lett*., vol. 122, pp. 23–30, 2019.

Address for correspondence:

Prof. Jurandir Nadal
Cardiovascular Engineering Lab.
Biomedical Engineering Program, COPPE
Universidade Federal do Rio de Janeiro
Rua Muniz de Aragão, 360 - Bloco 3 – Sala 310
Cidade Universitária
Rio de Janeiro - RJ, 21941-594
Brasil
E-mail: jn@peb.ufrj.br