# Influence of the Training Set Composition on the Estimation Performance of Linear ECG-Lead Transformations

Daniel Guldenring[1], Dewar D Finlay[2], Raymond R Bond[2],
Alan Kennedy[2], Peter Doggart[2], Ghalib Janjua[3], James McLaughlin[2]

[1]HS Kempten, Kempten, Germany
[2]Ulster University, Belfast, United Kingdom
[3]Robert Gordon University, Aberdeen, United Kingdom

## Abstract

*Linear ECG-lead transformations (LELTs) are used to estimate unrecorded target leads by applying a number of recorded basis leads to a LELT matrix. Such LELT matrices are commonly developed using training datasets that are composed of ECGs that belong to different diagnostic classes (DCs).*

*The aim of our research was to assess the influence of the training set composition on the estimation performance of LELTs that estimate target leads V1, V3, V4 and V6 from basis leads I, II, V2 and V5 of the 12-lead ECG. Our assessment was performed using ECGs from the three DCs left ventricular hypertrophy, right bundle branch block and normal (ECGs without abnormalities).*

*Training sets with different DC compositions were used for the development of LELTs matrices. These matrices were used to estimate the target leads of different test sets. The estimation performance of the developed matrices was quantified using root mean squared error values calculated between derived and recorded target leads.*

*Our findings indicate that unbalanced training sets can lead to LELTs that show large estimation performance variability across different DCs. Balanced training sets were found to produce LELTs that performed well across multiple DCs. We recommend balanced training sets for the development of LELTs.*

## 1. Introduction

Linear electrocardiographic (ECG) lead transformations (LELTs) are used to estimate or derive unrecorded target leads by applying a number of recorded basis leads to a LELT matrix [1, 2]. LELTs are a well-established concept in computerized electrocardiography. Two main application areas of LELTs have been established.

One of these established application areas is the estimation of leads that are not commonly recorded in clinical practice. The estimation of these leads is typically performed using ECGs that are recorded in a standard format. One example for this application area is the estimation of the Frank vectorcardiogram (VCG) [3, 4] from the Mason-Likar [1] 12-lead ECGs. This estimation is performed as the Frank VCG is thought to offer prognostic and diagnostic information that is in addition to the information contained in the 12-lead ECG. A further example of this application area is the estimation of device specific leads. This approach allows for the utilization of existing standard 12-lead ECG databases for the performance assessment of ECG devices that make use of non-standard leads [5].

The other established application area of LELTs are reduced lead systems. These systems aim to estimate the unrecorded leads of the 12-lead ECG [6] or the Frank VCG [2] from a reduced number of monitoring compatible electrodes. Such LELTs are used in continuous ECG monitoring applications where the information provided by the estimated lead sets is of interest, but the direct recording of the estimated lead set would be challenging due to electrode locations that are not suitable for continuous ECG monitoring.

LELT matrices are commonly developed using training sets and multivariate linear regression analysis [3, 6]. Training sets typically contain ECG data from different subjects and contain ECGs that belong to different diagnostic classes. The composition of the diagnostic classes in the training sets used for the development of LELTs is thought to have an influence on the estimation performance of the developed LELT matrices. However, a systematic assessment of the extend of this influence has, to the best of our knowledge, not been reported in literature.

The aim of our research was to assess what influence the composition of the diagnostic classes in a training set has on the estimation performance of LELTs. This assessment was conducted on a set of LELTs that has previously been studied [6] and been adopted into clinical practice [7].

## 2. Material and methods

### 2.1. Study population

Our study population was composed of 229 subjects with no abnormalities in their ECGs (normal), 232 subjects with left ventricular hypertrophy (LVH) and 250 subjects with right bundle branch block (RBBB). Random sampling without replacement was used to generate 200 different instances for each of the different train and test set compositions outlined in Table 1. This was performed using ECG data of the diagnostic classes normal, LVH and RBBB obtained from the subjects in the study population.

Table 1. Composition of the different train and test sets.

| name | type[a] | composition | | |
|---|---|---|---|---|
| | | #norm[b] | #LVH[c] | #RBBB[d] |
| TRnorm | train | 171 | 0 | 0 |
| TRlvh | train | 0 | 171 | 0 |
| TRrbbb | train | 0 | 0 | 171 |
| TRmix | train | 57 | 57 | 57 |
| TEnorm | test | 55 | 0 | 0 |
| TElvh | test | 0 | 55 | 0 |
| TErbbb | test | 0 | 0 | 55 |

*Notes. [a]train indicates a dataset composition that was used for the generation of LELT matrices and test indicates a dataset composition that was used for the performance assessment of LELT matrices; [b]number of ECGs of normal subjects in the dataset; [c]number of ECGs of LVH subjects in the dataset; [d]number of ECGs of RBBB subjects in the dataset.*

### 2.2. ECG data

Standard 12-lead ECG data of different sources was used in this research.

The standard 12-lead ECG data of 229 normal and 232 LVH subjects was extracted from body surface potential maps (BSPMs). Each BSPM contained electrocardiographic data of 120 BSPM leads. A representative average QRS-T complex was calculated for each of the 120 BSPM leads. Three of the 120 leads were recorded from electrodes placed on the right and left wrist and the left ankle (VR, VL and VF respectively). Electrodes situated at 81 anterior and 36 posterior locations were used to record 117 thoracic leads. A comprehensive description of the recording procedure can be found in [8]. A Laplacian 3D interpolation procedure [9] was applied to the 117 thoracic BSPM leads. This was performed to obtain body surface potentials at the locations of the 352 Dalhousie torso [10] nodes. Body surface potentials from electrode locations that were not a direct subset of the 352 Dalhousie torso nodes were obtained using linear interpolation [11]. Average QRS-T complexes of the standard 12-lead ECG were extracted from the interpolated BSPM data. More precisely, body surface potentials on the right wrist, the left wrist, the left ankle and from the location of the six precordial electrodes were used for the determination of the standard 12-lead ECG.

The standard 12-lead ECG data of 250 subjects with RBBB was assembled using the data in [12] and [13]. First, the annotations provided in [12] were used to identify 10 second 12-lead ECGs from different subjects with complete and incomplete RBBB. Second, average beats that were created using the Glasgow ECG program were obtained from [13]. Third, QRS-T complexes of the average beats where manually annotated and subsequently isolated.

### 2.3. Target and basis lead configuration

One established application area of LELTs are reduced lead systems. A reduced lead system that has been extensively studied [6] and has also been adopted into clinical practice [7] estimates the target leads V1, V3, V4 and V6 by applying a LELT matrix to the basis leads I, II, V2 and V5 of the standard 12-lead ECG. Given that this reduced lead system is a well-known application area of LELT matrices, we have chosen to assess the influence of the training set composition on the estimation performance of LELT matrices based on this basis and target lead configuration.

### 2.4. Development of the LELT matrices

The different instances of the training sets were used to generate 200 different LELT matrices for each training set composition. Individual LELT matrices were developed using the multivariate linear regression approach in (1).

$$ {}^{i}_{m}\boldsymbol{A} = \left( {}^{i}_{m}\boldsymbol{BL}^{\boldsymbol{T}} \cdot {}^{i}_{m}\boldsymbol{BL} \right)^{-1} \cdot {}^{i}_{m}\boldsymbol{BL}^{\boldsymbol{T}} \cdot {}^{i}_{m}\boldsymbol{TL}. \quad (1) $$

Where $[\cdot]^{\boldsymbol{T}}$ and $[\cdot]^{-1}$denote the transpose and the inverse of a matrix respectively, ${}^{i}_{m}\boldsymbol{A}$ refers to a $4 \times 4$ matrix of transformation coefficients that allows for the transformation of the basis leads into the target leads, $m \in \{TRnorm, TRlvh, TRrbbb, TRmix\}$ denotes the composition of the training set, $n$ refers to the number of sample values of the 171 QRS-T complexes in each lead of the training set, $i \in \{1, \ldots, 200\}$ denotes the instance of the training set that was used for the development of ${}^{i}_{m}\boldsymbol{A}$, ${}^{i}_{m}\boldsymbol{TL}$ refers to a $n \times 4$ matrix that contains $n$ sample values of the target leads and ${}^{i}_{m}\boldsymbol{BL}$ refers to a $n \times 4$ matrix that contains $n$ sample values of the basis leads.

### 2.5. Derivation of the target leads

Each ${}^{i}_{m}\boldsymbol{A}$ matrix was used to derive the target leads of three test sets. Each of these test sets was assembled in accordance to one of the test set compositions $k \in \{TEnorm, TElvh, TErbbb\}$ detailed in Table 1. The target lead derivation was performed using the approach in (2).

$$ {}^{i}_{m}\boldsymbol{dTL_k} = {}^{i}_{k}\boldsymbol{BL} \cdot {}^{i}_{m}\boldsymbol{A}. \quad (2) $$

Where ${}^{i}_{m}\boldsymbol{A}$, $m$ and $i$ are as defined in (1), ${}^{i}_{m}\boldsymbol{BL}$ is a $n \times 4$ matrix that contains the $n$ sample values of the QRS-T

complexes from the basis leads of one subject in the test dataset and $_{m}^{i}\boldsymbol{dTL_k}$ is $n \times 4$ matrix that contains the *n* sample values of the QRS-T complexes of the derived target leads.

## 2.6. Performance assessment

The influence of the training set composition on the estimation performance of the LELTs under investigation was assessed as detailed subsequently.

First, root mean square error (RMSE) values were calculated between the QRS-T complexes of the recorded and the derived target leads.

Second, the population RMSE and the population subject to subject variability were determined for each $_{m}^{i}\boldsymbol{A}$ using test data of the corresponding test set instance $i$. The population RMSE and population SSV were defined as the mean and the standard deviation of the RMSE values respectively. Values for the population RMSE and for the population SSV were determined over the 55 RMSE values per target lead and test set of each instance $i$. The outcome of this assessment were multiple $_{m}\boldsymbol{PRMSE_k}$ and $_{m}\boldsymbol{PSSV_k}$ matrices with $m \in \{TRnorm, TRlvh, TRrbbb, TRmix\}$ and $k \in \{TEnorm, TElvh, TErbbb\}$. Where $_{m}\boldsymbol{PRMSE_k}$ and $_{m}\boldsymbol{PSSV_k}$ are 200x4 matrices that contain the values of the population RMSE and the population SSV of the four target leads for each of the 200 instances $i$ respectively.

Third, the estimation performance associated with a training set composition $m$ was assessed for the different test set compositions $k$. The estimation performance was quantified as the mean population RMSE ($_{m}\overline{\boldsymbol{PRMSE}}_{\boldsymbol{k}}$) and the mean population SSV ($_{m}\overline{\boldsymbol{PSSV}}_{\boldsymbol{k}}$). Where $_{m}\overline{\boldsymbol{PRMSE}}_{\boldsymbol{k}}$ and $_{m}\overline{\boldsymbol{PSSV}}_{\boldsymbol{k}}$ are 1x4 vectors that were computed by calculating the mean over the 200 rows of the corresponding $_{m}\boldsymbol{PRMSE_k}$ and $_{m}\boldsymbol{PSSV_k}$ matrices.

Forth, the mean (95% confidence interval) of the differences $_{m}\boldsymbol{PRMSE_k}\text{-}_{mix}\boldsymbol{PRMSE_k}$ and $_{m}\boldsymbol{PSSV_k}\text{-}_{mix}\boldsymbol{PSSV_k}$ was calculated for all combinations $(m,k) = \{(TRnorm, TEnorm), (TRlvh, TElvh), (TRrbbb, TErbbb)\}$ in which $m$ and $k$ belong to the same diagnostic class. This was performed across the 200 differences associated with each target lead.

Fifth, the statistical significance of the differences in the mean population values $_{m}\overline{\boldsymbol{PRMSE}}_{\boldsymbol{k}} - _{mix}\overline{\boldsymbol{PRMSE}}_{\boldsymbol{k}}$ and $_{m}\overline{\boldsymbol{PSSV}}_{\boldsymbol{k}} - _{mix}\overline{\boldsymbol{PSSV}}_{\boldsymbol{k}}$ was assessed for all $m \in \{TRnorm, TRlvh, TRrbbb\}$ and $k \in \{TEnorm, TElvh, TErbbb\}$ using a two-tailed paired t-test (alpha level of significance .05).

## 3. Results

A summary of the findings from our analysis is provided in Table 2 and Table 3 for target leads V1 and V3 respectively. The findings for target leads V4 and V6 were similar to what can be observed in Table 2 and Table 3 and are omitted for brevity.

Table 2. Estimation performance [mean population RMSE; mean population SSV] associated with target lead V1 for the different train and test set compositions under investigation.

| Test | Train | | | | |
|---|---|---|---|---|---|
| | TRnorm | TRlvh | TRrbbb | TRmix | difference[a] |
| TEnorm | [102.5*; 45.4*] | [115.3*; 51.7*] | [132.1*; 56.7*] | [109.1; 46.0] | 6.5 [6.1; 7.0]<br>0.7 [0.3; 1.0] |
| TElvh | [111.1*; 91.9*] | [103.1*; 81.1*] | [136.1*; 85.9*] | [106.9; 82.5] | 3.8 [3.3; 4.2]<br>1.4 [0.8; 2.1] |
| TErbbb | [120.4*; 74.6*] | [117.2;* 72.5*] | [103.6*; 56.6*] | [109.8; 67.9] | 6.2 [5.6; 6.8]<br>11.3 [10.8; 11.8] |

[a]*mean (95% confidence interval) of the differences* $_{m}\boldsymbol{PRMSE_k}\text{-}_{mix}\boldsymbol{PRMSE_k}$ *and* $_{m}\boldsymbol{PSSV_k}\text{-}_{mix}\boldsymbol{PSSV_k}$ *for* $(m,k) = \{(TRnorm, TEnorm), (TRlvh, TElvh), (TRrbbb, TErbbb)\}$. *Notes. All values are in µV. Asterisks indicate statistical significance of the differences* $_{m}\overline{\boldsymbol{PRMSE}}_{\boldsymbol{k}} - _{mix}\overline{\boldsymbol{PRMSE}}_{\boldsymbol{k}}$ *and* $_{m}\overline{\boldsymbol{PSSV}}_{\boldsymbol{k}} - _{mix}\overline{\boldsymbol{PSSV}}_{\boldsymbol{k}}$ *for* $m \in \{TRnorm, TRlvh, TRrbbb\}$ *and* $k \in \{TEnorm, TElvh, TErbbb\}$ *at the ≤ 0.05 level.*

Table 3. Estimation performance [mean population RMSE; mean population SSV] associated with target lead V3 for the different train and test set compositions under investigation.

| Test | Train | | | | |
|---|---|---|---|---|---|
| | TRnorm | TRlvh | TRrbbb | TRmix | difference[a] |
| TEnorm | [104.4*; 61.8*] | [171.1*; 85.1*] | [146.4*; 78.3*] | [128.0; 69.5] | 23.6 [22.2; 25.0]<br>7.7 [6.7; 8.8] |
| TElvh | [183.0*; 174.1*] | [144.3*; 108.6*] | [150.6*; 107.2*] | [153.8; 121.3] | 9.5 [8.5; 10.6]<br>12.6 [10.7; 14.5] |
| TErbbb | [125.7*; 82.6*] | [120.6;* 72.2*] | [117.7*; 73.6*] | [118.1; 74.5] | 0.4 [0.1; 0.7]<br>0.9 [0.5; 1.3] |

[a]*mean (95% confidence interval) of the differences* $_{m}\boldsymbol{PRMSE_k}\text{-}_{mix}\boldsymbol{PRMSE_k}$ *and* $_{m}\boldsymbol{PSSV_k}\text{-}_{mix}\boldsymbol{PSSV_k}$ *for* $(m,k) = \{(TRnorm, TEnorm), (TRlvh, TElvh), (TRrbbb, TErbbb)\}$. *Notes. All values are in µV. Asterisks indicate statistical significance of the differences* $_{m}\overline{\boldsymbol{PRMSE}}_{\boldsymbol{k}} - _{mix}\overline{\boldsymbol{PRMSE}}_{\boldsymbol{k}}$ *and* $_{m}\overline{\boldsymbol{PSSV}}_{\boldsymbol{k}} - _{mix}\overline{\boldsymbol{PSSV}}_{\boldsymbol{k}}$ *for* $m \in \{TRnorm, TRlvh, TRrbbb\}$ *and* $k \in \{TEnorm, TElvh, TErbbb\}$ *at the ≤ 0.05 level.*

## 4. Discussion

The findings in Table 2 and Table 3 show lowest $\overline{PRMSE}$ values (highest estimation performance) when the LELT matrices were developed and tested on ECGs belonging to the same diagnostic class. This suggests suggest that the estimation performance of the LELTs under investigation can be optimized for ECGs that belong to one of the assessed diagnostic classes.

However, from the findings in Table 2 and Table 3 it can be seen that the estimation performance ($\overline{PRMSE}$ and $\overline{PSSV}$) of LELT matrices that are trained on only one of the assessed diagnostic classes does decrease notably when such a matrix is applied to ECGs that belong to a different

diagnostic class. Considering that, in a clinical setting, ECGs belonging to different diagnostic classes will be applied to LELT matrices it is evident that the practical utility of selectively trained LELT matrices is limited.

Interestingly, the findings in Table 2 and Table 3 show that the utilization of the balanced training set composition $m = TRmix$ does lead to LELT matrices that have good levels of estimation performance across the different assessed diagnostic classes. More precisely, the estimation performance of such LELT matrices was found to be close to the performance achieved by LELT matrices that were designed for and used on a specific diagnostic class.

The literature typically reports the estimation performance of LELT matrices over test sets containing ECGs that belong to multiple diagnostic classes. However, the findings in Table 3 show differences in the estimation performance across different diagnostic classes. Such as for example across the $_{mix}\overline{PRMSE}_k$ values for $k \in \{TEnorm, TElvh, TErbbb\}$. The composition of the utilized test set does therefore influence the observed estimation performance. A consequence of this is that a meaningful comparison of RMSE-based estimation performance values reported in the literature is only possible if they were obtained from similarly composed test sets.

This research has assessed the influence of the training set composition for the estimation performance of LELT matrices based on a particular configuration of target and basis leads. However, we speculate that similar findings can be obtained for LELT matrices based on other lead configurations.

A limitation of this research is that it has only assessed the influence of the training set composition on the waveform reconstruction error (RMSE between derived and recorded target leads). The extent to which the training set composition does influence diagnostic interpretation of reconstructed ECGs has not been assessed.

## 5. Conclusion

The composition of the training set was found to have an influence on the estimation performance of the LELTs under investigation. Our findings indicate that unbalanced training sets can lead to biased LELT matrices. The estimation performance of such biased LELT matrices was found to vary notably across different diagnostic classes. This variability in the estimation performance may not be tolerable in clinical settings where LELT matrices will be applied to ECGs belonging to different diagnostic classes. However, balanced training sets were found to produce LELT matrices that performed well over different diagnostic classes. We recommend the utilization of balanced training sets, that reflect the diagnostic classes of the target population, for the development of LELT matrices.

## References


[1] Guldenring D, Finlay DD, Strauss DG, et al. Transformation of the Mason-Likar 12-lead electrocardiogram to the Frank vectorcardiogram. Conf Proc IEEE Eng Med Biol Soc. 2012;2012:677-680.

[2] Guldenring D, Finlay DD, Bond RR, et al. The derivation of the spatial QRS-T angle and the spatial ventricular gradient using the Mason-Likar 12-lead electrocardiogram. J Electrocardiol. 2015;48(6):1045-1052.

[3] Kors A, van Herpen G, Sittig AC, et al. Reconstruction of the Frank vectorcardiogram from standard electrocardiographic leads: diagnostic comparison of different methods. Eur Heart J. 1990;11(12):1083–1092.

[4] Frank E. An accurate, clinically practical system for spatial vectorcardiography. Circulation 1956; 13(5):737-749.

[5] Guldenring D, Finlay, DD Bond RR, et al., Which part of the P-QRS-T is best when developing linear ECG-lead transformations for the performance assessment of patch based electrocardiographic devices?, J Electrocardiol. 2019 57(Supplement):101.

[6] Nelwan SP, Kors JA, Meij SH, et al., Reconstruction of the 12-lead electrocardiogram from reduced lead sets, J Electrocardiol. 2004:37(1):11-18.

[7] U.S. Department of Health and Human Services Food and Drug Administration, Center for Devices and Radiological Health (CDRH). (2003). 510(K) Summary: K030738, *Siemens Infinity MultiView Workstation Telemetry System Enhanced with TruST*$^{TM}$.

[8] Montague TJ, Smith ER, Cameron DA, Rautaharju PM, Klassen GA, Felmington CS, et al. Isointegral analysis of body surface maps: surface distribution and temporal variability in normal subjects. Circulation 1981; 63(5): 1166-1172.

[9] Oostendorp TF, van Oosterom A, Huiskamp G. Interpolation on a triangulated 3D surface. J Comput Phys 1989; 80(2):331-343.

[10] Horáček BM. Numerical Model of an Inhomogeneous Human Torso. Adv Cardiol 1974; 10:51-57.

[11] Schijvenaars BJA, Kors JA, van Herpen G, Kornreich F, van Bemmel JH. Interpolation of body surface potential maps. J Electrocardiol 1995; 28 Suppl 1: 104-109.

[12] Wagner P, Strodthoff N, Bousseljot RD, et al. PTB-XL, a large publicly available electrocardiography dataset. Scientific Data. 2020; 7(1):1-15.

[13] Nils Strodthoff, Mehari T, Nagel C, et al. PTB-XL+, a comprehensive electrocardiographic feature dataset. Scientific Data. 2023; 10(1):1-11.



Address for correspondence:

Daniel Guldenring
Room T117, Faculty of Electrical Engineering, HS Kempten,
Bahnhofstraße 61, 87435 Kempten, Germany
daniel.gueldenring@hs-kempten.de