Received 20 Sept, 2025; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version XX Month, XXXX.

Digital Object Identifier 10.1109/XXXX.2022.1234567

LeadKAN: A Low-Rank Kernelized Kolmogorov–Arnold Network for Efficient Nonlinear Representation of Electrocardiogram Signals

Mohammed Yusuf Ansari[1,2], MD Rabiul Islam[2], Mohammed Ishaq[4], Ibrahim Al-Muteb[3], Mohammed Yaqoob[4], Eduardo Feo-Flushing[1], Wajid Yousuf[6], Sarada Prasad Dakua[7], Marwa Qaraqe[5]

[1]Department of Computer Science, Carnegie Mellon University at Qatar, Doha, Qatar
[2]Department of Electrical and Computer Engineering, Texas A&M University, College Station, Tx, USA
[3]Department of Biomedical Engineering, Texas A&M University, College Station, Tx, USA
[4]University of Melbourne, Melbourne, Australia
[5]College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar
[6]Department of Petroleum Engineering, Texas A&M University, College Station, Tx, USA
[7]Department of Surgery/Clinical Advancements Department, Hamad General Hospital, Hamad Medical Corporation, Doha, Qatar

Corresponding author: Mohammed Yusuf Ansari (email: ma1@alumni.cmu.edu).

ABSTRACT Electrocardiography (ECG) is a widely used, non-invasive tool for assessing cardiac function, but conventional disease-centric models do not fully capture overall cardiovascular health. Recent work has introduced the concept of ECG age: a neural network–predicted age derived from ECG signals. Its difference from chronological age, known as delta age (ΔAge), has emerged as a surrogate marker of cardiovascular well-being. While deep learning approaches have shown promise for ECG age estimation, their computational complexity and lack of interpretability limit deployment in compute-constrained clinical environments. Kolmogorov–Arnold Networks (KANs) offer parameter efficiency and improved interpretability, yet existing variants remain compute-heavy, underexplored for regression tasks, and unable to disentangle contributions from individual ECG leads. To address these challenges, we propose LeadKAN, a lightweight and explainable KAN architecture for ECG age estimation. LeadKAN is built on LoRKAN layers, a novel layer design that replaces fully connected layers with low-rank bilinear mixing followed by an RBF-kernelized top, significantly reducing parameter count and computation. LeadKAN achieves ECG age estimation performance (MSE $\approx$ 112; MAE $\approx$ 8.25 years) comparable to state-of-the-art models, while requiring 16$\times$ fewer parameters and 45$\times$ fewer multiply–add operations. Additionally, lead-specific encoders enable attribution analysis, thereby enhancing clinical interpretability. These results position LeadKAN as an efficient and explainable framework for ECG age estimation, with strong potential for deployment in real-world, compute-limited settings.

INDEX TERMS Cardiovascular Well-being, Explainable Neural Network, Lead Importance Analysis, Lightweight Neural Networks, Low-rank Bilinear Mixing, Radial Basis Functions, Surrogate Metrics.

I. Introduction

Electrocardiography (ECG) is one of the most widely used non-invasive tools for assessing cardiac function. By capturing the heart's electrical activity through multiple leads, ECG enables the diagnosis of diverse conditions such as arrhythmias, myocardial infarction, and conduction abnormalities [1]. Traditionally, interpretation has relied on expert inspection, but manual analysis is subjective and often challenged by overlapping waveform patterns that can correspond to different pathologies. Recent advances in deep learning [2]–[4] have shown strong potential for automated ECG analysis, achieving high performance across single- and multi-disease classification tasks [5], [6]. While these approaches represent important progress, disease-centric models alone may not fully capture the broader state of cardiovascular health, particularly in patients with comorbidities or multiple risk factors.

Beyond diagnosing specific diseases, ECG signals can also capture subtle physiological changes associated with aging. Deep learning models trained to predict chronological age from ECG data have given rise to the concept *electrocardiogram age* (ECG age) [1], [7]. The difference between this predicted age and the actual chronological age (often referred to as ΔAge [7]) has emerged as a surrogate marker

of biological aging and overall cardiovascular well-being [8]. A higher ΔAge indicates reduced cardiovascular well-being, whereas a lower ΔAge suggests better overall cardiovascular health [9]. Despite its clinical promise, current deep learning approaches to ECG age estimation are computationally heavy [10], lack interpretability [11], and are poorly suited for deployment in resource-constrained environments.

Kolmogorov–Arnold Networks (KANs) have recently emerged as a promising alternative to traditional multilayer perceptrons (MLPs) for function approximation [12]. Unlike conventional MLPs, KANs learn univariate activation functions directly on the connections between neurons, rather than relying on node-based nonlinearities and weight optimization. This design enables the network to model complex functional relationships with higher parameter efficiency and improved interpretability. Despite these advances, existing KAN variants remain computationally intensive and parameter-heavy, particularly when scaling to high-dimensional inputs. These limitations restrict their deployment on resource-constrained devices, such as point-of-care medical equipment. This motivates the need for lightweight and compute-efficient KAN architectures. Furthermore, existing KANs do not disentangle the contributions of different input channels towards the final prediction, thereby limiting explainability. KANs have been applied to a variety of ECG-based diagnostic and forecasting tasks [13].

To address the limitations of existing KAN variants, this paper proposes LeadKAN, a lightweight, computationally efficient, and explainable architecture for ECG age estimation. Briefly, the main contributions are as follows:

- We introduce LeadKAN, a lightweight architecture built on Lowrank KAN (LoRKAN), a novel layer design that combines low-rank bilinear mixing with an RBF-kernelized top, providing an efficient alternative to parameter- and compute-heavy fully connected layers.
- We enhance interpretability by designing thirteen dedicated encoders, twelve for individual ECG leads and one for global features, enabling attribution analysis of lead-level contributions.
- We provide qualitative and quantitative evaluations, including principal component analysis of lead embeddings and a masking-based lead importance study, to assess representational quality and model explainability.
- We demonstrate the clinical utility of the ECG age metric (ΔAge) by comparing model performance on healthy and diseased cohorts, showing its potential as a surrogate measure of cardiovascular well-being.

The remainder of the paper is organized as follows. Section II provides a formulation of KANs and a brief overview of FASTKAN. Section III presents the methodology, including an overview of the LeadKAN architecture, a detailed description of the LoRKAN layers, and its parameter complexity. Section IV describes the experimental setup, including the datasets, implementation details of LeadKAN, and the evaluation metrics used to quantify performance in the ECG age estimation task, as well as network compute cost. Section V presents the results and discussion, comparing the ECG age estimation performance of LeadKAN against baseline models, along with qualitative analyses and assessment of the clinical utility of ECG age. Finally, Section VI concludes the paper.

II. Preliminaries & Related Work

This section provides the fundamental formulation of KANs and links it to the development of FASTKAN architecture. Additionally it provides brief literature overview of kANs and its application for ECG data.

A. Overview of KAN

KANs [12] are founded on the classical *Kolmogorov–Arnold representation theorem*, which establishes that any continuous multivariate function can be expressed as a sum of compositions of univariate functions. Formally, for a multivariate function $f : \mathbb{R}^n \to \mathbb{R}$, the theorem states:

$$f(x_1, x_2, \ldots, x_n) = \sum_{q=1}^{2n+1} \varphi_q \left(\sum_{p=1}^{n} \tau_{p,q}(x_p) \right), \qquad (1)$$

where $\tau_{p,q}(\cdot)$ represents the inner univariate transformations, $\varphi_q(\cdot)$ denotes the outer composition functions, and $x_1, x_2, \ldots, x_n$ ($x_i \in \mathbb{R}$) are the input features.

The original KAN framework leverages this theorem by approximating these univariate mappings using B-spline basis functions, which offer both universal approximation capability and localized updates for stable training. One particular layer in a KAN can be compactly represented in matrix form, where each entry corresponds to a learnable univariate function $\tau_{i,j}(\cdot)$ and row corresponds to a single composition of univariate functions:

$$\mathbf{x}_{out} = \begin{pmatrix} \tau_{1,1}(\cdot) & \tau_{1,2}(\cdot) & \cdots & \tau_{1,n}(\cdot) \\ \tau_{2,1}(\cdot) & \tau_{2,2}(\cdot) & \cdots & \tau_{2,n}(\cdot) \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{k,1}(\cdot) & \tau_{k,2}(\cdot) & \cdots & \tau_{k,n}(\cdot) \end{pmatrix} \mathbf{x}_{in}. \qquad (2)$$

Unlike a fully connected layer in a perceptron, which rely on fixed nonlinear activations applied after linear weighting, KANs directly learn the univariate transformations and aggregate them to generate the output.

B. Reducing KANs to RBF Networks

A major limitation of the original KAN implementation lies in its computational inefficiency on modern GPU architectures. Forward and backward propagation are significantly slower compared to the standard linear transformation followed by nonlinear activation in MLPs. This inefficiency arises primarily from the use of B-splines and updating their basis/grid during training.

To address this issue, Li *et al.* [14] proposed FASTKAN by replacing B-spline basis functions with radial basis functions (RBFs) as the activation basis. The key motivation is

that RBF kernels can approximate n-order B-splines with high accuracy, while being computationally efficient on GPUs. Specifically, if the input features are normalized to a bounded range, a set of RBF centers can be distributed across that range to smoothly localize the features. The RBF activations can be expressed as

$$f(x) = \sum_{i=1}^{K} w_i \exp\left(-\frac{\|x - c_i\|^2}{2h^2}\right), \quad c_i, h \text{ fixed}, \qquad (3)$$

where $x \in \mathbb{R}$ denotes a single input feature, w_i are the learnable coefficients corresponding to RBF kernels, c_i represent the RBF centers, and h is the kernel width parameter. The activations are obtained by evaluating the response of each RBF kernel, and the resulting values across all inputs in the layer are finally combined through a linear transformation. It should be noted that the K RBF kernels are shared across the input features. While FASTKAN improves the training and inference efficiency of KANs, it introduces a large linear transformation matrix at the top.

C. KANs for ECG

Liu et al. [12] modeled KANs by employing B-spline functions to represent these univariate transformations. While effective, the approach was not optimized for GPU execution, and updating the B-splines on CPUs incurred significant computational cost, which limited scalability for deeper or wider architectures. EFFKAN [15] addressed this limitation by expressing all functions as linear combinations of a fixed set of B-spline basis functions. This reformulation allowed the computations to be performed as matrix multiplications, reducing memory usage and improving efficiency. Subsequently, Li et al. [14] proposed FastKAN for accelerated inference by approximating B-splines with RBFs.

KANs have also been applied to a variety of ECG analysis. For instance, Zeleke et al. [13] employed a federated learning setup with KANs to investigate the trade-off between algorithmic performance and data privacy in arrhythmia classification using the MIT-BIH dataset [16]. Similarly, Zhao et al. [17] applied KANs to raw ECG signals, which were first processed using piecewise aggregate approximation to reduce signal frequency before performing one-dimensional feature extraction. Beyond ECG analysis, KANs have been utilized for related biomedical applications. For instance, Bataineh et al. [18] employed KANs to predict carotid intima-media thickness from tabular clinical data.

III. Methodology

This section describes the LeadKAN architecture for ECG data. Specifically, it provides an overview of the architecture and the innovations made with the LoRKAN layer to minimize parameter count and compute of the network.

A. LeadKAN Overview

The key motivation behind the proposed *LeadKAN* architecture is to develop a lightweight, efficient, and explainable KAN tailored for ECG data. In addition, LeadKAN is explicitly designed for regression tasks while maintaining a degree of explainability by capturing the dependence of predictions on individual ECG leads. To this end, the architecture introduces a dedicated encoder for each of the 12 leads as well as one encoder for global ECG features, resulting in a total of 13 parallel encoders. The input ECG data are partitioned accordingly, with each partition directed to a specific encoder. Each encoder is built from a single low-rank KAN layers (LoRKAN) that produces 8-dimensional lead-level embedding or global ECG feature embedding. These embeddings are concatenated to form an integrated representation, after which inter-lead interactions are modeled by another LoRKAN layer with output dimension 64. The resulting holistic embedding captures both local and global patterns across leads and is ultimately passed through a final LoRKAN layer with a single scalar output for ECG age estimation.

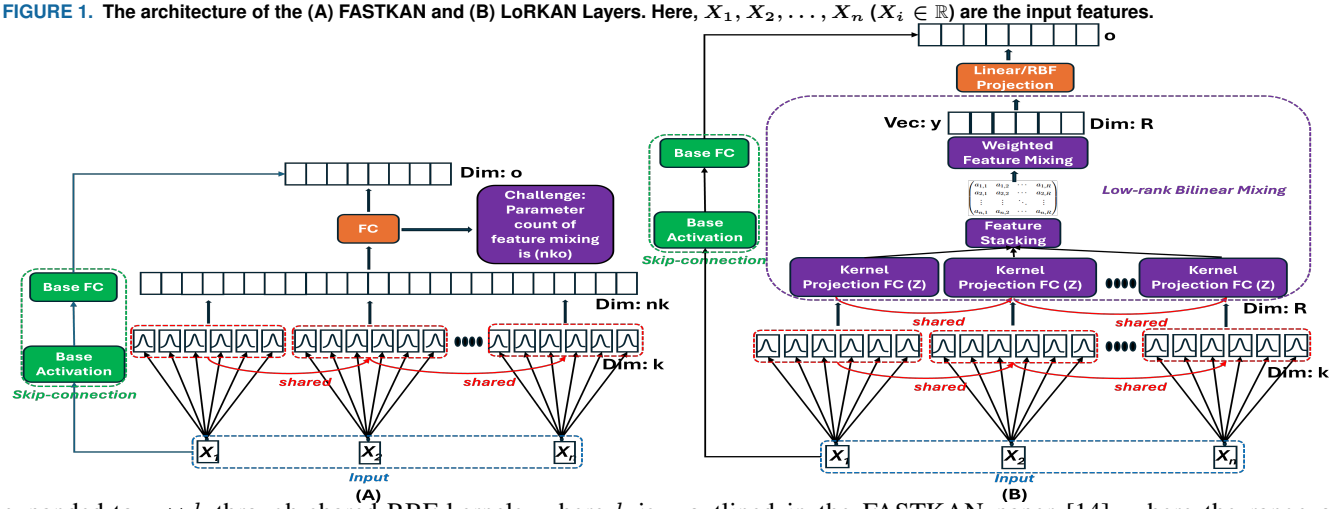
Algorithm 1 LoRKAN Layer

Stage 1: RBF Expansion (FASTKAN-style)
Initialize $X \in \mathbb{R}^{n \times k}$
for $i = 1$ to n **do**
  for $\ell = 1$ to k **do**
    $X_{i,\ell} \leftarrow \exp\Big(-\frac{(x_i - c_\ell)^2}{2h_{\text{rbf}}^2}\Big)$
  end for
end for
Stage 2: Kernel Projection (low-rank mixing across kernels)
$Z \leftarrow XP$ $\{P \in \mathbb{R}^{k \times R}$ *trainable*$\}$
Stage 3: Inter-feature Mixing (column-wise dot products)
for $j = 1$ to R **do**
  $y_j \leftarrow w_{:,j}^{\top} z_{:,j}$ $\{W \in \mathbb{R}^{n \times R}$ *trainable*$\}$
end for
$\mathbf{y} \leftarrow [y_1, \ldots, y_R]^{\top}$ $\{\mathbf{y} \in \mathbb{R}^R\}$
Stage 4: Output Head
if linear top **then**
  $o \leftarrow Ay + b$ $\{A \in \mathbb{R}^{o \times R},\ b \in \mathbb{R}^o$ *trainable*$\}$
  $\{A$ is linear transformation and b is bias$\}$
else
  for $m = 1$ to M **do**
    $d_m \leftarrow \|y - s_m\|_2$ $\{\{s_m\} \subset \mathbb{R}^R$ *trainable supports*$\}$
    $\phi_m \leftarrow \exp\Big(-\frac{d_m^2}{2h^2}\Big)$
  end for
  $\phi \leftarrow [\phi_1, \ldots, \phi_M]^{\top}$
  $o \leftarrow U\phi + c$ $\{U \in \mathbb{R}^{o \times M},\ c \in \mathbb{R}^o$ *trainable*$\}$
  $\{U$ is linear transformation and c is bias$\}$
end if
return o

B. LoRKAN Layer

The proposed LoRKAN layer (Figure 1 (B)) builds upon the FASTKAN (Figure 1 (A)) formulation introduced by Li *et al.* [14]. In FASTKAN, n input features of the layer are

FIGURE 1. The architecture of the (A) FASTKAN and (B) LoRKAN Layers. Here, $X_1, X_2, \ldots, X_n$ ($X_i \in \mathbb{R}$) are the input features.



expanded to $n \times k$ through shared RBF kernels, where k is the number of kernels. This representation is then mapped to an output of dimension o using a fully connected (FC) layer, introducing nko parameters. Such a transformation is parameter-intensive, computationally heavy, and scales poorly when either the number of input features or the number of kernels grows large. To overcome this limitation, the LoRKAN layer replaces the FC layer with a low-rank bilinear mixing strategy, followed by either an additional RBF kernel mapping or a linear top. Algorithm 1 shows the high-level steps of the LoRKAN layer.

1) Low Rank Bilinear Mixing

A key assumption for the successful application of low-rank bilinear mixing is that the interaction between the n input features can be captured in low-dim R-dimensional space. Once the input is expanded into an $n \times k$ representation, the vector can be viewed as a matrix with n rows and k columns. To efficiently capture interactions across the kernel dimension, a shared linear transformation, referred to as the kernel projection, is applied across the k kernel values of each input feature. Formally, the transformation is expressed as:

$$\mathbf{Z} = \mathbf{X}\mathbf{P}, \tag{4}$$

where $\mathbf{X} \in \mathbb{R}^{n \times k}$ is the expanded input representation, $\mathbf{P} \in \mathbb{R}^{k \times R}$ is the learnable kernel projection matrix, and $\mathbf{Z} \in \mathbb{R}^{n \times R}$ is the transformed output.

When $R = k$, the operation acts as a feature mixing step in the kernel space without changing dimensionality. When $R < k$, the operation reduces the kernel dimension, serving as a form of dimensionality reduction that retains essential features for the task. After kernel dimensionality reduction, the embeddings no longer preserve the precise localization of a feature within the kernel grid but instead encode broader structural characteristics, such as whether the feature lies toward the left or right of the grid or whether it takes positive or negative values. For selecting the parameters for the RBF kernel of LeadKAN, we follow the approach outlined in the FASTKAN paper [14], where the range of the feature values is normalized to the interval [-2, 2], with 8 equally spaced kernel centers (k=8) to adequately cover this range. The rank parameter R was empirically determined via an ablation study (supplementary material), where R=8 provided the best balance between prediction accuracy and parameter efficiency, with larger values yielding diminishing performance gains. The width of each kernel (h) is determined by dividing the grid range (i.e., 4) by the number of RBF centers minus one (i.e., 7), reducing overlap between different kernels.

After the kernel projection, the matrix $\mathbf{Z} \in \mathbb{R}^{n \times R}$ encodes n input features represented with R channels. To aggregate information across the n input features, a learnable feature mixing matrix $\mathbf{W} \in \mathbb{R}^{n \times R}$ of the same shape is introduced. Each column $\mathbf{w}_{:,j}$ of $\mathbf{W}$ serves as a weight vector for combining the corresponding column $\mathbf{z}_{:,j}$ of $\mathbf{Z}$. The inter-feature mixing is then obtained by a column-wise dot product,

$$y_j = \mathbf{w}_{:,j}^{\top} \mathbf{z}_{:,j}, \quad j = 1, \ldots, R, \tag{5}$$

resulting in a low-rank vector $\mathbf{y} \in \mathbb{R}^R$. This operation can be interpreted as a weighted averaging mechanism, where each input feature contributes according to its learned weight, thus enabling the model to emphasize the most relevant features when generating the vector $\mathbf{y}$.

2) RBF Kernel/Linear Top

The final stage of the LoRKAN layer involves mapping the low-rank representation $\mathbf{y} \in \mathbb{R}^R$ into an output vector $\mathbf{o} \in \mathbb{R}^o$. A straightforward approach is to apply a FC layer ($A \in \mathbb{R}^{o \times R}$, $b \in \mathbb{R}^o$ *trainable*), which has been shown to work effectively in architectures where several LoRKAN layers are stacked and the input to the network consists of all tabular features jointly. However, in LeadKAN, where the design employs dedicated encoders for each ECG lead, a simple FC mapping often proves insufficient. To better capture nonlinear interactions, we employ an RBF-kernelized projection. that enriches the representation by mapping it into a higher-dimensional feature [19].

Formally, let $\{\mathbf{s}_1, \ldots, \mathbf{s}_M\}$ denote a set of learnable support vectors in the R-dimensional space (i.e., $\{\mathbf{s}_i \in \mathbb{R}^R\}$). For each support $\mathbf{s}_m$, the Euclidean distance to the low-rank vector $\mathbf{y}$ is computed as

$$d_m = \|\mathbf{y} - \mathbf{s}_m\|_2, \quad m = 1, \ldots, M. \tag{6}$$

Each distance is then passed through an RBF kernel to generate an activation,

$$\phi_m(\mathbf{y}) = \exp\left(-\frac{d_m^2}{2h^2}\right), \tag{7}$$

where h is the kernel width parameter. The resulting activations $\phi_1(\mathbf{y}), \ldots, \phi_M(\mathbf{y})$ are concatenated to form an M-dimensional kernelized representation of low-rank vector $\mathbf{y} \in \mathbb{R}^R$. This kernelized feature vector ($\in \mathbb{R}^M$) is then transformed by a FC layer ($U \in \mathbb{R}^{o \times M}$, $c \in \mathbb{R}^o$ *trainable*) to produce the final output vector $\mathbf{o} \in \mathbb{R}^o$.

C. Parameter Complexity

The parameter count of the proposed LoRKAN formulation can be decomposed into three parts. First, the kernel projection ($\mathbf{P}$) requires $k \times R$ parameters, where k is the number of kernels and R the reduced rank. Second, the inter-feature mixing introduces an additional $n \times R$ parameters through the learnable weight matrix $\mathbf{W}$. Finally, the RBF-kernelized projection from the R-dimensional vector to the output $\mathbf{o} \in \mathbb{R}^o$ introduces $M \times o$ parameters, where M is the number of support vectors. The total parameter count of the low-rank design is therefore:

$$\mathcal{O}(kR + nR + Mo).$$

In contrast, the standard FASTKAN approach requires $\mathcal{O}(nko)$ parameters. Hence, low-rank bilinear mixing achieves a substantial reduction in parameter complexity, while retaining expressive capacity through kernelized non-linear projections.

D. LeadKAN Explainability

To assess the dependence of LeadKAN on individual ECG leads, we conducted a masking-based attribution analysis. Each encoder generates a vector embedding of size 8, and the output of one encoder was masked at a time by setting its embedding to zero. The model was re-evaluated on the full test set after masking each encoder's output. This process was carried out sequentially for each of the 12 lead-specific encoders and the global encoder. The contribution of each encoder was quantified by the Δ MSE, calculated as the difference between the model's performance with complete unmasked embeddings and the performance after masking the output of one specific encoder. Unlike the MSE loss function used for training the network, which optimizes the overall model parameters, Δ MSE measures the performance degradation caused by masking the output of one particular encoder. Δ MSE is defined as

$$\Delta\text{MSE}j = \text{MSE}(\mathcal{M}j(\mathbf{X})) - \text{MSE}(\mathbf{X}), \tag{8}$$

where $\mathbf{X}$ represents the unmasked input embeddings, $\mathcal{M}j(\mathbf{X})$ represents the embeddings with the j-th encoder masked, and $\Delta\text{MSE}j$ indicates the performance degradation due to masking. Figure 5 presents the ΔMSE values for the 12 ECG lead encoders and the global feature encoder, capturing performance degradation in age estimation when one encoder's output is masked (i.e., nullified). A higher Δ MSE indicates a greater degradation in performance, highlighting the significance of that particular encoder to the model's overall function. Conversely, lower Δ MSE values suggest a smaller impact on the model's performance.

IV. Experimental Setup

This section outlines the ECG dataset, LeadKAN implementation details, and evaluation metrics for ECG age estimation.

A. PTB-XL+ Dataset

The experiments in this work are conducted using the PTB-XL+ dataset [20], an extension of the PTB-XL corpus. PTB-XL [21] comprises 21,799 raw 12-lead ECG recordings collected from a German cohort, and PTB-XL+ augments this resource by providing clinically validated features extracted from the raw signals using the GE 12SL system (782 features). This conversion ensures that the dataset includes standardized ECG descriptors commonly used in clinical practice. In addition, PTB-XL+ includes metadata, such as age, gender, and folds for 10-fold cross-validation with strict patient-level separation, ensuring that all recordings from a single patient are assigned to the same fold and thereby preventing data leakage during evaluation. During preprocessing, we identified that only 8 out of the 782 features contained missing values; since this represented a negligible fraction of the feature set and the affected features were not uniquely informative, these features were removed, resulting in a final data dimensionality of 774 features.

B. Implementation Details

The LeadKAN framework was implemented in PyTorch, building upon the publicly available FASTKAN codebase. Training was performed using mean squared error (MSE) loss and the Adam optimizer with an initial learning rate of 0.01 and an L_2 regularization coefficient of 10^{-4}. A step-based learning rate scheduler was employed with a decay factor $\gamma = 0.05$ and a step size of 35 epochs. The model was trained under a 10-fold cross-validation strategy, with each LeadKAN trained for 75 epochs within each fold. All reported results represent the mean performance across 10 folds, with the standard deviation calculated over these folds to provide a robust and reliable evaluation. We have employed a range of tuned machine learning models as a baseline for the ECG feature-based age estimation. Additionally, we included AttiaNet [11] and ResNet1D [10] architectures for age estimation to ensure a fair comparison with raw ECG-based age estimation methods.

C. Evaluation Metrics

To evaluate the performance of LeadKAN for ECG age estimation, we employed MSE and mean absolute error (MAE). Let y_i denote the chronological age of the i-th sample and $\hat{y}_i$ the corresponding ECG age. **MSE** is defined as $\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2$ and was selected because it corresponds directly to the training loss function. **MAE** is defined as $\frac{1}{N}\sum_{i=1}^{N}|y_i-\hat{y}_i|$ and provides a more interpretable measure, since its units are expressed in years, indicating the average deviation of the ECG age from the chronological age.

In addition to predictive accuracy, the efficiency of the network was assessed using **parameter count**, **disk utilization**, and the number of **multiplications and additions** required during inference. These metrics capture the lightweight nature of the proposed architecture and facilitate fair comparison with existing models.

V. Results & Discussion

This section presents the performance of machine learning and deep learning models for ECG age estimation. Also, it provides a qualitative analysis of the LeadKAN including its dependence on individual ECG leads, and discusses the utility of ECG age as a metric of cardiovascular wellbeing.

A. Age Estimation Performance

TABLE 1. ECG Age estimation performance of different machine learning regressors.

Machine Learning Model	Test MSE	SD MSE	Test MAE	SD MAE
Decision Tree Regressor	190.57	7.93	10.91	0.19
Elastic Net	137.13	6.11	9.21	0.16
KNeighbors Regressor	169.34	5.61	10.35	0.19
Extra Trees Regressor	143.98	4.89	9.55	0.14
Random Forest Regressor	140.37	**4.79**	9.39	0.14
Support Vector Regressor	118.15	5.50	8.48	0.18
LightGBM Regressor	**117.08**	4.10	**8.48**	**0.12**
XGBoost Regressor	123.40	3.57	8.66	0.13

TABLE 2. Performance of deep learning models for ECG age estimation.

Deep Learning Model	Variant	Test MSE	SD MSE	Test MAE	SD MAE
AttiaNet [11] (100Hz)	–	112.02	5.05	8.23	0.17
ResNet1D [10] (100Hz)	–	150.15	20.83	9.42	0.69
EFFKAN [15]	Conf1	118.37	3.80	8.51	0.09
EFFKAN [15]	Conf2	116.78	4.01	8.42	0.12
EFFKAN [15]	Conf3	116.22	8.39	8.74	**0.11**
FASTKAN [14]	Conf1	114.55	5.57	8.35	0.15
FASTKAN [14]	Conf2	111.97	8.25	8.74	**0.11**
FASTKAN [14]	Conf3	**112.12**	4.19	**8.24**	0.12
MLP Regressor	Conf3	126.20	5.71	8.82	0.17
LeadKAN (ours)	–	112.44	**3.54**	8.25	**0.11**

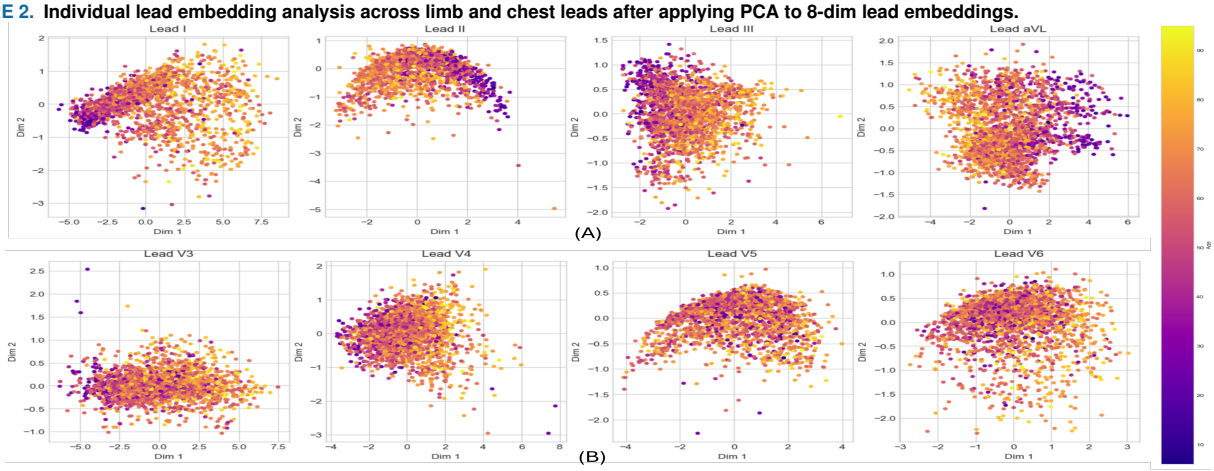
Table 1 summarizes the performance of several machine learning models for ECG age estimation task, offering a baseline to understand the task's complexity. The corresponding hyperparameter search space and the best configurations used for each model are provided in supplementary material to ensure reproducibility. As shown in Table 1, simple regressors such as decision trees and Kneighbors regressor exhibit high MSE, indicating poor ECG age estimation performance. In particular, the degradation of Kneighbors regressor can be attributed to the high dimensionality of the ECG feature space ($\sim$774 features), where the data points are sparsely distributed and distance-based similarity measures become ineffective [22]. In contrast, ensemble and bagging-based methods such as random forest and extra trees regressors achieve significantly better performance, with MSE values around 140 and MAE below 10 years. Elastic Net regression performs comparably to these ensemble approaches, suggesting that combining regularization approaches is effective for age estimation. The best-performing model is the LightGBM regressor, which achieves the lowest MSE and competitive MAE across all evaluated methods, demonstrating the effectiveness of gradient boosting with leaf-wise tree growth for capturing complex nonlinear relationships in ECG-derived features. These results highlight the limitations of distance-based and shallow learners for high-dimensional ECG data, while underscoring the advantages of kernel-based methods for ECG age estimation.

TABLE 3. Compute cost of deep learning models for ECG age estimation.

Deep Learning Model	Variant	Parameter Count	Disk Utilization (MB)	Mult-Adds
AttiaNet [11] (100Hz)	–	593,505	2.37	0.35×10^6
ResNet1D [10] (100Hz)	–	6,940,065	27.76	143.97×10^6
EFFKAN [15]	Conf1	62,000	0.28	0.06×10^3
EFFKAN [15]	Conf2	2,310,400	8.90	2.08×10^3
EFFKAN [15]	Conf3	496,000	1.79	0.45×10^3
FASTKAN [14]	Conf1	57,389	0.22	0.06×10^3
FASTKAN [14]	Conf2	2,082,085	7.90	2.08×10^3
FASTKAN [14]	Conf3	448,157	1.70	0.45×10^3
MLP Regressor	Conf3	49,793	0.20	0.05×10^3
LeadKAN (ours)	–	**28,205**	**0.16**	0.01×10^3

Table 2 summarizes the performance of deep learning architectures for ECG age estimation using raw ECG signals (100 Hz) or tabular feature representations, while the corresponding computational costs are reported in Table 3. Among the raw ECG data models, AttiaNet achieves an MSE of $\approx$112 with an MAE of 8.23 years, whereas ResNet-1D, despite being a larger architecture, performs poorly with an MSE of $\approx$150 and MAE of 9.23 years due to overfitting on low sampling rate raw ECG data. For KAN-based approaches, we evaluated three configurations of EfficientKAN (EFFKAN) and FASTKAN. Configurations are as follows: Config-1 [774, 8, 1], Config-2 [774, 256, 128, 1], and Config-3 [774, 64, 1], where the first and last values correspond to the input and output dimensions of the KAN, respectively. In EFFKAN, the shallow configuration with minimal hidden units (Config 1) yields an MSE of 118.37, while deeper and wider networks (Config 2) improve performance slightly to 116.78. Notably, Config 3, which remains shallow but increases the intermediate dimension to 64, attains the best EFFKAN result of 116.22, highlighting the advantage of wider but simpler architectures. FASTKAN consistently outperforms EFFKAN, with Config 1 achieving

FIGURE 2. Individual lead embedding analysis across limb and chest leads after applying PCA to 8-dim lead embeddings.



MSE of 114.55 and Config 3 improving it further to 112.12 while using fewer parameters than Config 2.

The proposed LeadKAN achieves comparable performance to FASTKAN Config 3, with an MSE of 112.44 and an MAE of 8.25 years, but with a drastically reduced computational footprint. Specifically, LeadKAN requires only 28,205 parameters ($\sim$0.16 MB) and $\sim 0.01 \times 10^3$ multiplication–addition operations, representing a $16\times$ reduction in parameter count and $45\times$ reduction in mult-adds compared to FASTKAN Config 3, while maintaining similar age estimation performance. It should be noted that the multiplication–addition operations are accounted for within the *nn.Linear* layers. The additional overhead from updating the B-spline grid in EFFKAN or from the kernelized RBF top in LeadKAN is relatively minor compared to the overall computational cost. In contrast, raw ECG models such as AttiaNet (595k parameters) and ResNet-1D ($\sim$7M parameters) incur substantially higher computational costs due to convolutional operations, with mult-add counts on the order of 10^6. Overall, these results demonstrate that LeadKAN achieves a favorable balance between ECG age estimation performance and efficiency, outperforming ECG raw-data architectures and offering a more compact/compute efficient alternative to existing KAN variants.

B. Qualitative Analysis

1) Individual Lead Embedding Analysis

LeadKAN incorporates 13 dedicated encoders, one for each ECG lead and one for global features, which enables examination of individual lead embeddings. To illustrate this, embeddings were extracted from four limb leads (I, II, III, aVL) and four chest leads (V3, V4, V5, V6) and projected into two dimensions using principal component analysis (PCA). Figure 2 presents the resulting visualization.

The limb leads display a clear separation of ECG instances with respect to age. Younger cases, shown in blue and purple, cluster toward one region of the PCA space, while older cases, shown in orange and yellow, occupy another. This separation suggests that the limb leads contain features that are highly informative for age estimation. In contrast, the chest leads exhibit limited separation. ECGs for younger patients are distributed throughout the PCA space, indicating that these leads provide less age-specific information.

This observation is consistent with the physiological basis of ECG measurement. Limb leads record the electrical activity of the heart in the frontal plane, capturing global conduction patterns that are sensitive to age-related changes in atrial and ventricular depolarization [23]. Chest leads record activity in the transverse plane, focusing on localized ventricular regions, which may vary less systematically with age. The greater separability observed in the limb-lead embeddings highlights their importance in improving the interpretability and performance of LeadKAN for ECG age estimation.

2) Global Embedding Analysis

After the 12-lead embeddings and the global feature embedding are generated in LeadKAN, they are concatenated and passed through a mixer layer. Figure 4 illustrates the distribution of the embeddings in two-dimensional PCA space before and after mixing. Prior to the mixer, the concatenated representation sparsely occupy a high-dim feature space, and ECG instances of different ages are broadly scattered, as seen in 2D PCA plot (Figure 4 (A)). After inter-lead mixing, the embeddings become more compact, and similar age instances are closely packed together, forming a smooth gradient from younger to older cases. Furthermore, the post-mixer feature distributions along the two principal axes approximate Gaussian shapes, indicating that the skewness present in the initial concatenated embeddings is reduced.

3) Lead Importance for Age Estimation

Figure 5 shows the lead attribution analysis of LeadKAN. The results indicate that the global feature encoder has the largest impact, with substantial increases in error when masked. These features include average P-wave, QRS complex, and T-wave durations and amplitudes, as well as heart rate variability indices, all of which are clinically known to

FIGURE 3. Clinical utility of ECG Age established by the delta age in LeadkAN trained on healthy instances.

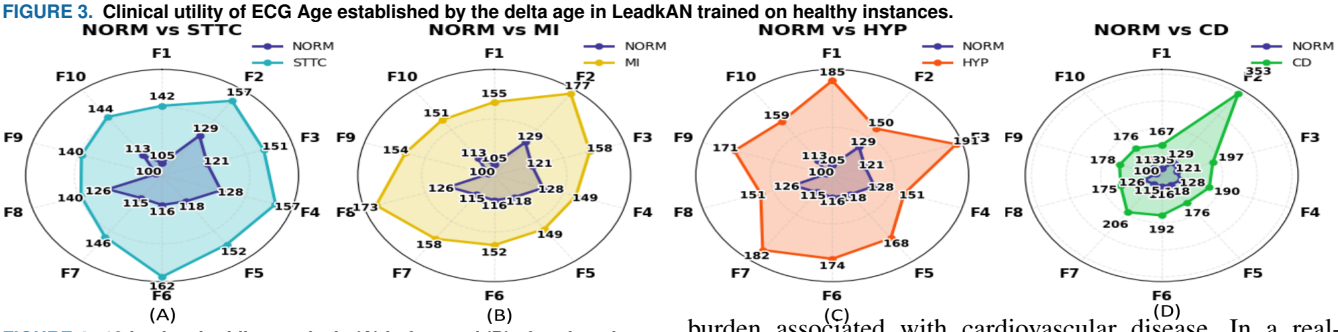


FIGURE 4. 12-lead embedding analysis (A) before and (B) after the mixer layer in LeadKAN.

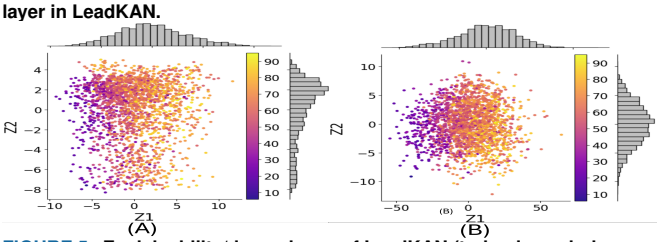
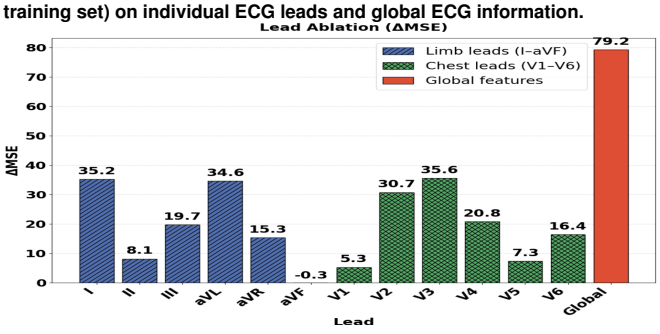


FIGURE 5. Explainability/dependence of LeadKAN (trained on whole training set) on individual ECG leads and global ECG information.



vary systematically with age. Among the individual leads, four limb leads and four chest leads show high importance, while Lead II and Lead aVF among the limb leads, and V1 and V5 among the chest leads, exhibit relatively low contribution. This reduced importance is because certain limb leads are linear combinations of others, and chest leads are placed in close proximity provide overlapping information [24]. These findings suggest that the lead set could potentially be reduced without significantly degrading performance, offering opportunities for simplified acquisition in practical ECG age estimation.

C. Clinical Utility of ECG Age

To examine the clinical relevance of ECG age, we trained LeadKAN exclusively on healthy subjects and evaluated its performance on distinct test subsets, including normal cases (NORM), ST/T changes (STTC), myocardial infarction (MI), hypertrophic cardiomyopathy (HYP), and conduction disturbances (CD). Figure 3 presents radar plots of the MSE across the ten folds comparing normal class with disease classes. For the normal class, the model achieves a low average MSE of approximately 110. In contrast, the disease classes exhibit markedly higher errors, indicating a larger deviation between predicted ECG age and chronological age. This deviation, denoted as ΔAge, reflects the excess disease burden associated with cardiovascular disease. In a real-world setting, higher magnitude of ΔAge can therefore serve as a surrogate marker of cardiovascular health and provide a clinically meaningful measure of well-being [8], [25]–[28].

A critical analysis of the proposed framework reveals several key insights. The superior performance of the LeadKAN-based architecture can be attributed to its shared low-rank projection and kernelized transformations. In particular, the model benefits from capturing both global and lead-specific patterns. Furthermore, a consistent performance gap is observed between healthy and disease test instances when the model is trained on healthy ECG data, indicating that pathological variations disrupt learned age-related representations. Finally, a key limitation of this study is the reliance on the PTB-XL+ datasets, which uniquely provide both raw signals and standardized GE 12SL features [29].

VI. Conclusion

In summary, LeadKAN is a lightweight and computationally efficient KAN designed for regression tasks, specifically for ECG age estimation. The architecture replaces the parameter-intensive fully connected layers of FASTKAN with low-rank bilinear mixing followed by an RBF-kernelized top layer, achieving comparable accuracy to FASTKAN while using 16× fewer parameters and 45× fewer multiply-add operations. Despite its reduced computational footprint, LeadKAN matches the performance of deep learning models trained on raw ECG signals. Lead-specific encoders in LeadKAN enhance interpretability by enabling attribution analysis of individual ECG leads. Training on healthy cohorts and evaluating on diseased cohorts further demonstrates the clinical utility of ECG age as a surrogate marker of cardiovascular health. A limitation of LoRKAN layer is that the low-rank bilinear mixing and RBF-kernelized top are slower in practice than a single large matrix multiplication, leading to relatively longer inference times on GPU when compared to FASTKAN. Future work will optimize these components to improve training and inference efficiency. Furthermore, clinical studies for ΔAge analysis through explicit age-dependent bias correction and population-specific calibration.

Acknowledgement

Portions of this research were conducted with computing resources provided by Texas A&M High Performance Research Computing.

REFERENCES

[1] M. Y. Ansari, M. Qaraqe, F. Charafeddine, E. Serpedin, R. Righetti, and K. Qaraqe, "Estimating age and gender from electrocardiogram signals: a comprehensive review of the past decade," *Artificial Intelligence in Medicine*, vol. 146, p. 102690, 2023.

[2] M. Yaqoob, M. Y. Ansari, M. Ishaq, U. Ashraf, S. Pavuluri, A. Rabbani, H. S. Rabbani, and T. D. Seers, "Fluidnet-lite: Lightweight convolutional neural network for pore-scale modeling of multiphase flow in heterogeneous porous media," *Advances in Water Resources*, vol. 200, p. 104952, 2025.

[3] H. Dahmani, M. Yaqoob, M. Y. Ansari, and E. F. Flushing, "Thermal homography in photovoltaic panels: Evaluating deep learning and feature-based methods," in *2025 IEEE Texas Power and Energy Conference (TPEC)*. IEEE, 2025, pp. 1–6.

[4] M. Yaqoob, M. Y. Ansari, M. Ishaq, I. S. A. J. Jayachandran, M. Hashim, and T. D. Seers, "Microcrystalnet: An efficient and explainable convolutional neural network for microcrystal classification using scanning electron microscope petrography," *IEEE Access*, 2025.

[5] I. Afsa, M. Y. Ansari, S. Paul, O. Halabi, E. Alataresh, J. Shah, A. Hamze, O. Aboumarzouk, A. Al-Ansari, and S. P. Dakua, "Development and validation of a class imbalance-resilient cardiac arrest prediction framework incorporating multiscale aggregation, ica and explainability," *IEEE Transactions on Biomedical Engineering*, 2024.

[6] M. Y. Ansari, M. Yaqoob, M. Ishaq, E. F. Flushing, I. A. c. Mangalote, S. P. Dakua, O. Aboumarzouk, R. Righetti, and M. Qaraqe, "A survey of transformers and large language models for ecg diagnosis: advances, challenges, and future directions," *Artificial Intelligence Review*, vol. 58, no. 9, p. 261, 2025.

[7] M. Y. Ansari, M. Qaraqe, R. Righetti, E. Serpedin, and K. Qaraqe, "Enhancing ecg-based heart age: impact of acquisition parameters and generalization strategies for varying signal morphologies and corruptions," *Frontiers in Cardiovascular Medicine*, vol. 11, p. 1424585, 2024.

[8] C.-H. Chang, C.-S. Lin, Y.-S. Luo, Y.-T. Lee, and C. Lin, "Electrocardiogram-based heart age estimation by a deep learning model provides more information on the incidence of cardiovascular disorders," *Frontiers in Cardiovascular Medicine*, vol. 9, p. 754909, 2022.

[9] J. Libiseller-Egger, J. E. Phelan, Z. I. Attia, E. D. Benavente, S. Campino, P. A. Friedman, F. Lopez-Jimenez, D. A. Leon, and T. G. Clark, "Deep learning-derived cardiovascular age shares a genetic basis with other cardiac phenotypes," *Scientific reports*, vol. 12, no. 1, p. 22625, 2022.

[10] E. M. Lima, A. H. Ribeiro, G. M. Paixão, M. H. Ribeiro, M. M. Pinto-Filho, P. R. Gomes, D. M. Oliveira, E. C. Sabino, B. B. Duncan, L. Giatti *et al.*, "Deep neural network-estimated electrocardiographic age as a mortality predictor," *Nature communications*, vol. 12, no. 1, p. 5117, 2021.

[11] Z. I. Attia, P. A. Friedman, P. A. Noseworthy, F. Lopez-Jimenez, D. J. Ladewig, G. Satam, P. A. Pellikka, T. M. Munger, S. J. Asirvatham, C. G. Scott *et al.*, "Age and sex estimation using artificial intelligence from standard 12-lead ecgs," *Circulation: Arrhythmia and Electrophysiology*, vol. 12, no. 9, p. e007284, 2019.

[12] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljacic, T. Y. Hou, and M. Tegmark, "KAN: Kolmogorov–arnold networks," in *The Thirteenth International Conference on Learning Representations*, 2025. [Online]. Available: https://openreview.net/forum?id=Ozo7qJ5vZi

[13] S. N. Zeleke and M. Bochicchio, "Federated kolmogorov-arnold networks for health data analysis: A study using ecg signal," in *2024 IEEE International Conference on Big Data (BigData)*. IEEE, 2024, pp. 8070–8077.

[14] Z. Li, "Kolmogorov-arnold networks are radial basis function networks," *arXiv preprint arXiv:2405.06721*, 2024.

[15] Blealtan, "EFFKAN: Efficient Kan (github repository)," https://github.com/Blealtan/efficient-kan, 2024, accessed: 2025-09-14.

[16] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE engineering in medicine and biology magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[17] Z. Zhao, L. Sui, Y. Song, B. Vucetic, and Z. Lin, "Kolmogorov-arnold-based network with lightweight feature fusion schema for single-lead electrocardiogram atrial fibrillation detection," in *2024 IEEE 24th International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE, 2024, pp. 1–6.

[18] A. Al Bataineh, B. Vamsi, M. El-Abd, and B. P. Doppala, "Kolmogorov–arnold networks for predicting carotid intima-media thickness in cardiovascular risk assessment," *Scientific Reports*, vol. 15, no. 1, p. 32108, 2025.

[19] B. Schölkopf, "The kernel trick for distances," *Advances in neural information processing systems*, vol. 13, 2000.

[20] N. Strodthoff, T. Mehari, C. Nagel, P. J. Aston, A. Sundar, C. Graff, J. K. Kanters, W. Haverkamp, O. Dössel, A. Loewe *et al.*, "Ptb-xl+, a comprehensive electrocardiographic feature dataset," *Scientific data*, vol. 10, no. 1, p. 279, 2023.

[21] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze, W. Samek, and T. Schaeffter, "Ptb-xl, a large publicly available electrocardiography dataset," *Scientific data*, vol. 7, no. 1, pp. 1–15, 2020.

[22] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional space," in *International conference on database theory*. Springer, 2001, pp. 420–434.

[23] J. Sundnes, G. T. Lines, X. Cai, B. F. Nielsen, K.-A. Mardal, and A. Tveito, *Computing the electrical activity in the heart*. Springer Science & Business Media, 2007, vol. 1.

[24] U. Jain, A. A. Butchy, M. T. Leasure, V. A. Covalesky, D. McCormick, and G. S. Mintz, "Redundancy and novelty between ecg leads based on linear correlation." in *BIOSIGNALS*, 2023, pp. 359–365.

[25] H. E. van der Wall, G.-J. Hassing, R.-J. Doll, G. J. van Westen, A. F. Cohen, J. L. Selder, M. Kemme, J. Burggraaf, and P. Gal, "Cardiac age detected by machine learning applied to the surface ecg of healthy subjects: Creation of a benchmark," *Journal of Electrocardiology*, vol. 72, pp. 49–55, 2022.

[26] F. Lopez-Jimenez, S. Kapa, P. A. Friedman, N. K. LeBrasseur, E. Klavetter, K. E. Mangold, and Z. I. Attia, "Assessing biological age: the potential of ecg evaluation using artificial intelligence: Jacc family series," *Clinical Electrophysiology*, vol. 10, no. 4, pp. 775–789, 2024.

[27] R. Leung, B. Wang, M. Gottbrecht, A. Doerr, N. Marya, A. Soni, D. D. McManus, and H. Lin, "Association between deep neural network-derived electrocardiographic-age and incident stroke," *Frontiers in Cardiovascular Medicine*, vol. 11, p. 1368094, 2024.

[28] S. Shelly, F. Lopez-Jimenez, A. Chacin-Suarez, M. Cohen-Shelly, J. R. Medina-Inojosa, S. Kapa, Z. Attia, A. A. Chahal, V. K. Somers, P. A. Friedman *et al.*, "Accelerated aging in lmna mutations detected by artificial intelligence ecg–derived age," in *Mayo Clinic Proceedings*, vol. 98, no. 4. Elsevier, 2023, pp. 522–532.

[29] M. Y. Ansari, R. Righetti, and M. Qaraqe, "Quantifying cardiovascular wellbeing through ecg age: Acquisition constraints and the case for handcrafted features," in *2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2025, pp. 1–7.