TU Delft

Delft University of Technology

Document Version
Final published version

Licence
Dutch Copyright Act (Article 25fa)

Citation (APA)
Battiloro, C., Cavallo, A., & Isufi, E. (2026). Covariance Filters and Neural Networks Over Hilbert Spaces. In *Proceedings of the ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 421-425). IEEE. https://doi.org/10.1109/ICASSP55912.2026.11463820

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3315-6701-9/26/$31.00 ©2026 IEEE | DOI: 10.1109/ICASSP55912.2026.11463820

COVARIANCE FILTERS AND NEURAL NETWORKS OVER HILBERT SPACES

Claudio Battiloro[1], *Andrea Cavallo*[2] *and Elvin Isufi*[2]

[1]Harvard University [2]Delft University of Technology

ABSTRACT

Covariance Neural Networks (VNNs) perform graph convolutions on the empirical covariance matrix of signals defined over finite-dimensional Hilbert spaces, motivated by robustness and transferability properties. Yet, little is known about how these arguments extend to infinite-dimensional Hilbert spaces. In this work, we take a first step by introducing a novel convolutional learning framework for signals defined over infinite-dimensional Hilbert spaces, centered on the (empirical) covariance operator. We constructively define Hilbert coVariance Filters (HVFs) and design Hilbert coVariance Networks (HVNs) as stacks of HVF filterbanks with nonlinear activations. We propose a principled discretization procedure, and we prove that empirical HVFs can recover the Functional PCA (FPCA) of the filtered signals. We then describe the versatility of our framework with examples ranging from multivariate real-valued functions to reproducing kernel Hilbert spaces. Finally, we validate HVNs on both synthetic and real-world time-series classification tasks, showing robust performance compared to MLP and FPCA-based classifiers.

1. INTRODUCTION

Covariance-based processing underpins several signal processing and machine learning frameworks, such as data whitening, exposing principal directions, and capturing feature interdependence. Its benefits have been shown in a wide plethora of domains, from brain neuroscience to finance [1, 2, 3, 4, 5]. The covariance matrix spectrum implicitly captures the structure of a dataset via the *principal components*, and said structure can be exploited via the PCA transform. While PCA learning is broadly adopted [6], its practical implementation suffers from the inherent stochastic perturbations in the empirical covariance matrix estimated from finite data, which gets amplified in high-dimension/low-sample regimes or with tightly clustered eigenvalues [7, 8]. To address this, coVariance Neural Networks (VNNs) [9] perform graph convolutions using the empirical covariance matrix as graph shift operator, successfully merging the representational power of PCA and the stability/transferability properties of Graph Neural Networks (GNNs) [10]. However, VNNs operate only on finite-dimensional Hilbert spaces and little is known about how to generalize their benefits in cases where signals are defined over infinite-dimensional Hilbert spaces. This is a common setting in, for example, time-series processing [11] or quantum machine learning [12].

Related Works. Covariance-based data representation has been extensively studied from a graph signal processing perspective. For example, the works in [13, 14] formalized stationarity for graph signals and related the graph Fourier transform with the spectrum of the data covariances. This link has been also exploited in [15, 16] for graph-based dimensionality reduction. More recently, [9] introduced VNNs, which use the empirical covariance matrix as a graph shift operator, and showed that VNNs are more expressive and stable than PCA. The transferability properties of VNNs have been studied in [17]. VNNs have been successfully applied to brain data in [18]. A variant of VNNs for streaming spatiotemporal signals has been proposed in [19], while VNNs with explicit bias mitigation have been proposed in [20]. The work in [21] recently proposed to replace the empirical covariance matrix in VNNs with the empirical covariance density operators to improve user's control on the stability-discriminability trade-off of the network, and enhance robustness to noise. However, these works consider data only on finite-dimensional Hilbert spaces.

Contribution. We introduce a novel convolutional learning framework for random signals defined over general, possibly infinite-dimensional, Hilbert spaces, centered on the covariance operator of such space. We leverage the eigensystem of the covariance operator as a frequency domain to define a Hilbert coVariance Fourier Transform, and then we propose spectral Hilbert coVariance Filters (HVF) to manipulate the signal representation in such spectral domain. We show how one can seamlessly develop Hilbert coVariance Networks (HVNs) as a layered architecture of HVFs and nonlinear activations in the same spirit as GNNs [10]. We propose a principled discretization procedure to make HVFs and HVNs implementable, and we prove that empirical HVFs can recover the Functional PCA (FPCA) [22] of the filtered signals. Finally, we numerically validate HVNs on both synthetic and real-world classification tasks of time-series, corroborating our findings and showing robustness benefits w.r.t. FPCA and covariance-unaware learning such as MLPs.

2. PRELIMINARIES AND MOTIVATION

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is the sample space (set of all possible outcomes), $\mathcal{F}$ is a sigma-algebra, i.e., a collection of measurable subsets of Ω treated as events, and $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a probability measure, i.e., a function assigning probabilities to the elements in $\mathcal{F}$. Consider also a separable Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and norm $\|\cdot\|_{\mathcal{H}}$. In this paper, we work with random variables $X : \Omega \to \mathcal{H}$ that are square integrable, i.e. they have finite

WORK PARTIALLY FUNDED BY THE TU DELFT AI LABS PROGRAMME, THE NWO OTP GRASPA PROPOSAL #19497, THE NWO VENI PROPOSAL 222.032. EMAIL TO: CBATTILORO@HSPH.HARVARD.EDU

Authorized licensed use limited to: TU Delft Library. Downloaded on May 12,2026 at 06:40:35 UTC from IEEE Xplore. Restrictions apply.

expected squared norm $\mathbb{E}\,\|X\|_{\mathcal{H}}^2 < \infty$. We denote the mean $\mu = \mathbb{E}[X] \in \mathcal{H}$. We refer to the elements of $\mathcal{H}$ as signals.

Covariance operator. The covariance operator $C : \mathcal{H} \to \mathcal{H}$ is the linear map that encodes all directional (mixed) second moments of X. The covariance operator acts on $v \in \mathcal{H}$ as

$$Cv \;=\; \mathbb{E}\Big[\; \underbrace{\langle X-\mu,\, v\rangle_{\mathcal{H}}}_{\text{scalar coefficient}}\,(X-\mu)\Big]. \tag{1}$$

Note that the operator C is compact and trace-class [23], and hence there exist nonnegative eigenvalues $\{\lambda_\ell\}_{\ell\geq 1}$ and orthonormal eigenfunctions $\{\phi_\ell\}_{\ell\geq 1}$ such that

$$Cv \;=\; \sum_{\ell\geq 1} \lambda_\ell\; \langle v,\, \phi_\ell\rangle_{\mathcal{H}}\; \phi_\ell. \tag{2}$$

Hilbert Covariance Fourier Transform (HVFT). A spectral representation of signals is conventionally tied to a Fourier transform to analyze them. Given the eigen-system $\{(\lambda_\ell, \phi_\ell)\}_\ell$, we interpret the (ordered) eigenvalues as frequencies, and we define the Hilbert coVariance Fourier Transform (HVFT) of a signal $x \in \mathcal{H}$ as the projection

$$\widetilde{x}[\ell] \;\triangleq\; \langle x,\, \phi_\ell\rangle_{\mathcal{H}}\,, \qquad \ell \geq 1. \tag{3}$$

Crucially, the HVFT of a signal $x \in \mathcal{H}$ is equivalent to its Functional PCA (FPCA) transform [22] up to recentering. Although HVF/FPCA transforms are general means to obtain a different representation of signals, they also keep the usual variational characterization, i.e., taking the first r eigenfunctions $\{\phi_\ell\}_\ell$ is equivalent to selecting the r directions maximizing signal variance under orthonormality, or minimizing reconstruction error. In the finite-dimensional case ($\mathcal{H} \cong \mathbb{R}^m$), FPCA and PCA (transforms) are equivalent.

Problem Motivation. As explained, the HVFT transform provides a principled spectral domain for (random) signals in Hilbert spaces. Their spectral representations can be directly leveraged in processing/learning algorithms as information-rich features [24]. However, as we elaborate more in Sec. 4 and empirically show in Sec. 5, practically implementing FPCA requires a finite number of discretized signals. This often leads to numerical instabilities and poor transferability performance, i.e., two finite batches of discretized samples coming from the same underlying Hilbert space or from similar Hilbert spaces could potentially results in significantly different spectral representations, making FPCA-only learning extremely difficult. For these reasons, there is a clear need for a proper notion of (convolutional) filtering [25] and neural networks centered on the covariance operator, to successfully merge the representational power of FPCA with the stability/transferability properties of neural networks [10, 9].

3. HILBERT COVARIANCE FILTERS AND NEURAL NETWORKS

In this section, we introduce spectral and spatial (polynomial) Hilbert coVariance Filters by building on the spectral domain induced by the covariance operator C. Then, we leverage these filters to design Hilbert coVariance Networks.

Definition 1 (Spectral Hilbert coVariance Filter)**.** *Consider a function $h : [0, \|C\|_{\mathcal{H}}] \to \mathbb{R}$ bounded and Borel (e.g., continuous, monotone). A spectral Hilbert coVariance Filter (HVF) with frequency response h, denoted with $\mathbf{h}$, is defined as*

$$g = \mathbf{h}(C)\, x \;\triangleq\; \sum_{\ell=1}^{\infty} h(\lambda_\ell)\; \langle x,\, \phi_\ell\rangle_{\mathcal{H}}\; \phi_\ell + h(0)\, x_\perp, \tag{4}$$

where $x \in \mathcal{H}$ is the input and $x_\perp$ is the orthogonal projection of x onto $\ker(C)$.

By direct computation, it is trivial to check that filtering is pointwise in the HVFT domain, i.e.,

$$\widetilde{g}[\ell] \;=\; h(\lambda_\ell)\, \widetilde{x}[\ell], \qquad \ell \geq 1. \tag{5}$$

This operation, where the frequency response $h(\lambda_\ell)$ amplifies or attenuates the different functional principal components, is in complete analogy with the convolutions/spectral filtering in several other domains, e.g., time [25], graphs [26], or manifolds [27]. By learning or designing $h(\lambda_\ell)$, we are able to put different weights on the different HVFT coefficients. As such, taking only the r variance-maximizing directions, i.e., employing the FPCA transform as a dimensionality-reduction method, is then equivalent to an ideal high pass HVF. General spectral filters are maximally expressive, but they require explicit computation of eigenfunctions and eigenvalues. For this reason, spatial (polynomial) filters are often preferred in practice, as they are (i) scalable and stable–they avoid eigendecomposition and are eigenvector-agnostic, and (ii) still expressive–polynomials are strong approximators.

Definition 2 (Spatial Hilbert coVariance Filters)**.** *Given filter order $J \geq 0$ and a set of real valued parameters $w_0, \ldots, w_J$, the corresponding spatial HVF is given by*

$$\mathbf{h}(C) \;\triangleq\; \sum_{j=0}^{J} w_j\, C^j. \tag{6}$$

A spatial HVF is entirely defined by the parameters w_j. Indeed, using (3), it is trivial to show that the frequency response of a spatial HVF has the form $h(\lambda) = \sum_{k=0}^{J} w_k\, \lambda^k$. Intuitively, C can be thought of as a shift operator (like the time delay [25] or a graph shift [26]). Each power C^j repeatedly spreads/smooths the signal through j-step covariance couplings (akin to j-hop diffusion), and the weights w_j mix these scales. We can now define Hilbert coVariance Networks using banks of HVFs and nonlinear activations.

Definition 3 (Hilbert coVariance Networks)**.** *A Hilbert coVariance Network (HVN) is a layered architecture composed of banks of HVFs and nonlinear activations (operators). The $t-$th layer considers F_t input signals $\left\{x_t^i\right\}_{i=1}^{F_t}$ and F_{t+1} output signals $\{x_{t+1}^u\}_{u=1}^{F_{t+1}}$. Denoting the nonlinear activation with $\sigma : \mathcal{H} \to \mathcal{H}$, the propagation rule reads as*

$$x_{t+1}^u = \sigma\left(\sum_{i=1}^{F_t} \mathbf{h}(C)_t^{u,i} x_t^i\right), \quad u = 1, \ldots, F_{t+1}. \tag{7}$$

A HVN with input signals $\left\{x^i\right\}_{i=1}^{F_0}$ is then built as the stack of T layers defined as in (7), where $x_0^i = x^i$. To globally represent the HVN, we collect all the filters in $\mathcal{W} = \left\{h_t^{u,i}\right\}_{t,u,i}$ and describe the HVN outputs as a mapping $\{x_T^u\}_{u=1}^{F_T} = \mathbf{\Psi}\left(\mathcal{W}, C, \left\{x^i\right\}_{i=1}^{F_0}\right)$ to enhance that it is parameterized by filters $\mathcal{H}$ and the covariance operator C. An HVN can be spectral or spatial, depending on the type of filters employed.

Authorized licensed use limited to: TU Delft Library. Downloaded on May 12,2026 at 06:40:35 UTC from IEEE Xplore. Restrictions apply.

4. DISCRETIZATION VIA BOUNDED OPERATORS

Hilbert coVariance Filters and Neural Networks operate on signals in a (possibly infinite-dimensional) Hilbert space whose distribution—and thus the covariance operator C—is unknown, so they are not directly implementable. We make them practical in two steps: (i) working with the empirical covariance operator using i.i.d., and (ii) discretizing the samples using bounded discretization operators to enable finite-dimensional processing and learning.

Empirical covariance operator. Given i.i.d. samples $x_1, \dots, x_n \in \mathcal{H}$, let $\bar{x} = n^{-1} \sum_{i=1}^{n} x_i$. The empirical covariance operator $\hat{C}_n : \mathcal{H} \to \mathcal{H}$ is defined, for $v \in \mathcal{H}$, by

$$\hat{C}_n v \triangleq \frac{1}{n} \sum_{i=1}^{n} \langle x_i - \bar{x},\, v \rangle_{\mathcal{H}} \, (x_i - \bar{x}), \quad v \in \mathcal{H}. \tag{8}$$

By definition, $\hat{C}_n$ is bounded, self-adjoint, positive, and finite rank with $r := \operatorname{rank}(\hat{C}_n) \leq n-1$. Hence there exist eigenvalues $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_r > 0$ (with $r \leq n-1$) and orthonormal eigenfunctions $\{\hat{\phi}_1, \dots, \hat{\phi}_r\}$ such that

$$\hat{C}_n v = \sum_{\ell=1}^{r} \hat{\lambda}_\ell \left\langle v,\, \hat{\phi}_\ell \right\rangle_{\mathcal{H}} \hat{\phi}_\ell, \qquad v \in \mathcal{H}. \tag{9}$$

The 0 eigenvalue may have infinite multiplicity with null spaces given by $\{v \in \mathcal{H} : \left\langle v,\, \hat{\phi}_\ell \right\rangle_{\mathcal{H}} = 0 \ \forall\, \ell \leq r\}$. As a valuable expressivity analysis of spatial HVFs, we now show that their empirical counterpart, i.e., replacing C with $\hat{C}_n$ in (6), can provably recover the (empirical) FPCA transform of the samples. Let $\{\hat{\lambda}_1, \dots, \hat{\lambda}_r\}$ be the positive eigenvalues of $\hat{C}_n$ (with multiplicity) and $\{\phi_\ell\}_{\ell=1}^{r}$ the corresponding orthonormal family. For a distinct eigenvalue value α, let $I(\alpha) = \{\ell \leq r : \hat{\lambda}_\ell = \alpha\}$ and define the orthogonal projector $P_\alpha : \mathcal{H} \to \mathcal{H}$ by

$$P_\alpha x \triangleq \sum_{\ell \in I(\alpha)} \left\langle x,\, \hat{\phi}_\ell \right\rangle_{\mathcal{H}} \hat{\phi}_\ell, \qquad x \in \mathcal{H}. \tag{10}$$

Theorem 1. *Let $\Lambda = \{\alpha_1, \dots, \alpha_q\}$ (with $q \leq r$) be the set of* distinct *positive eigenvalues of $\hat{C}_n$. For each $\alpha \in \Lambda$ there exists a spatial HVF $\mathbf{h}_\alpha$ as in* (6) *such that*

$$\mathbf{h}_\alpha(\hat{C}_n)\, x = P_\alpha x, \qquad x \in \mathcal{H}. \tag{11}$$

Moreover, one can choose $\deg(\mathbf{h}_\alpha) \leq q$.

Proof. See Appendix A in [28].

Corollary 1. *Fix $\alpha \in \Lambda$ and any orthonormal basis $\{\hat{\phi}_\ell\}_{\ell \in I(\alpha)}$ of its eigenspace. For any $x \in \mathcal{H}$, there exists a polynomial HVF $\mathbf{h}_\alpha$ as in* (6) *such that*

$$\left\langle \mathbf{h}_\alpha(\hat{C}_n) x\,,\, \hat{\phi}_\ell \right\rangle_{\mathcal{H}} = \left\langle x,\, \hat{\phi}_\ell \right\rangle_{\mathcal{H}}, \qquad \forall\, \ell \in I(\alpha), \tag{12}$$

i.e., recovering the α-th (empirical) FPCA score of x.

Definition 4 (Discretization Operator)**.** *Let $\{f_j\}_{j=1}^{m}$, with $f_j : \mathcal{H} \to \mathbb{R}$ be bounded linear functionals on $\mathcal{H}$. By the Riesz representation theorem, there exist $s_j \in \mathcal{H}$ with $f_j(x) = \langle x,\, s_j \rangle_{\mathcal{H}}$, $x \in \mathcal{H}$. The discretization operator $S_m : \mathcal{H} \to \mathbb{R}^m$ acts on $x \in \mathcal{H}$ as*

$$S_m x = [f_1(x), \dots, f_m(x)]^\top = [\langle x,\, s_1 \rangle_{\mathcal{H}}, \dots, \langle x,\, s_m \rangle_{\mathcal{H}}]^\top. \tag{13}$$

The Hilbert adjoint $S_m^* : \mathbb{R}^m \to \mathcal{H}$ satisfies $\langle x,\, S_m^* a \rangle_{\mathcal{H}} = \langle S_m x, a \rangle_{\mathbb{R}^m}$ and it acts on $a \in \mathbb{R}^m$ as

$$S_m^* a = \sum_{j=1}^{m} a_j\, s_j, \qquad a = [a_1, \dots, a_m]^\top. \tag{14}$$

Given $x_1, \dots, x_n \in \mathcal{H}$ we denote $x_i^{(m)} := S_m x_i \in \mathbb{R}^m$ and $\overline{x}^{(m)} = \frac{1}{n} \sum_i x_i^{(m)} = S_m \bar{x}$. Therefore, the empirical covariance matrix of the discretized signals is given by

$$\hat{\mathbf{C}}_n^{(m)} = \frac{1}{n} \sum_{i=1}^{n} \left(x_i^{(m)} - \overline{x}^{(m)}\right)\left(x_i^{(m)} - \overline{x}^{(m)}\right)^\top \in \mathbb{R}^{m \times m}. \tag{15}$$

As a sanity check for the empirical covariance matrix, the following result shows that the discretized version of the empirical covariance operator is the empirical covariance matrix. In other words, discretizing the samples and then computing the empirical covariance matrix is equivalent to computing the empirical covariance operator and then discretizing it.

Proposition 1. *Let S_m, $\hat{C}_n$, and $\hat{\mathbf{C}}_n^{(m)}$ be a discretization operator as in* (13)*, the empirical covariance operator as in* (8)*, and the empirical covariance matrix as in* (15)*, respectively. The following holds:*

$$\hat{\mathbf{C}}_n^{(m)} = S_m\, \hat{C}_n\, S_m^*. \tag{16}$$

Proof. See Appendix A in [28].

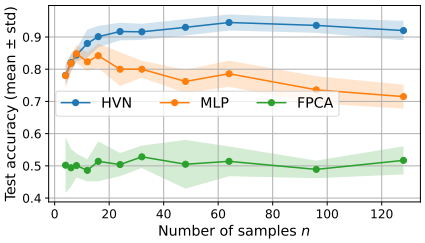
Discrete HVFs, and HVNs. Once signals are discretized and the empirical covariance matrix is computed, discrete HVFs and HVNs are simply given by (6) and (7) with the covariance operator replaced by the empirical covariance matrix and the signals replaced by their discretized version, i.e. $\{x_T^{(m),u}\}_{u=1}^{F_T} = \mathbf{\Psi}(\mathcal{W}, \hat{\mathbf{C}}_n^{(m)}, \{x_i^{(m)}\}_{i=1}^{n})$. Discretized HVNs are principled instances of coVariance Neural Network (VNN) [9] and, as a consequence, of Graph (convolutional) Neural Networks (GNNs) [10]. To better enhance this similarity, we rewrite the layer in (7) in matrix form by introducing the matrices $\mathbf{X}_t^{(m)} = \{x_{t,i}^{(m)}\}_{i=1}^{F_t} \in \mathbb{R}^{m \times F_t}$, and $\mathbf{W}_{t,j} = \{w_{t,j}^{i,q}\}_{q=1,i=1}^{F_l, F_{t+1}} \in \mathbb{R}^{F_l \times F_{t+1}}$, as

$$\mathbf{X}_{t+1}^{(m)} = \sigma\left(\sum_{j=0}^{J} \left(\hat{\mathbf{C}}_n^{(m)}\right)^j \mathbf{X}_t^{(m)} \mathbf{W}_{t,j} \right) \in \mathbb{R}^{m \times F_{t+1}}, \tag{17}$$

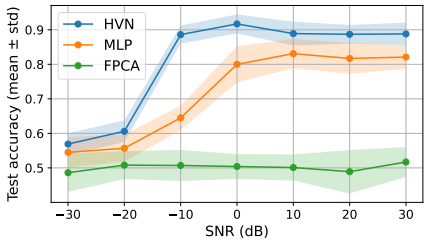
where the filter weights $\{\mathbf{W}_{t,j}\}_{t,j}$ are learnable parameters, and $\mathbf{X}_0^{(m)} = \{x_i^{(m)}\}_{i=1}^{n} \in \mathbb{R}^{m \times n}$.

Use-Cases for (Discrete) HVNs. To showcase the versatility of our framework, in Appendix B in [28], we illustrate three practical $(\mathcal{H}, S)$ pairs together with admissible nonlinear activations $\sigma : \mathcal{H} \to \mathcal{H}$. In particular, we describe how HVNs can be applied to functional data (e.g. time series, as we also show in Sec. 5), infinite sequences, and data defined over Reproducing Kernel Hilbert Spaces.

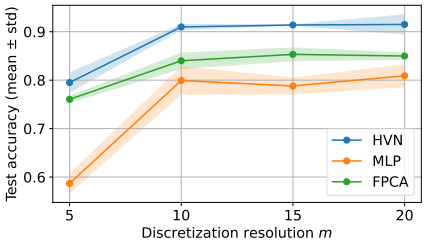
Authorized licensed use limited to: TU Delft Library. Downloaded on May 12,2026 at 06:40:35 UTC from IEEE Xplore. Restrictions apply.



(a) Test accuracy versus number of samples.



(b) Test accuracy versus SNR (in dB).



(c) Test accuracy on ECG dataset.

5. NUMERICAL RESULTS

We validate HVNs on a synthetic and a real task. In the synthetic setting, we perform binary classification of multivariate time-series bags. The time-series are evaluated on the unit interval and have d channels, i.e., they live in $\mathcal{H} = L^2([0,1], \beta; \mathbb{R}^d)$, where β is the Lebesgue measure. We observe them through the discretization operator S_p as defined in Ex. 1 of Appendix B in [28], i.e., we perform channelwise bin–averaging on a partition of $[0,1]$ into p intervals, yielding $m = dp$ discrete components per sample. For a fixed class $y \in \{0,1\}$, the i-th sample x_i is drawn from a zero–mean Gaussian process with matrix–valued kernel $K^{(y)}(t,s) = k_t(t,s)\Sigma_c^{(y)} \in \mathbb{R}^{d\times d}$, $0 \le t,s \le 1$. Equivalently, the covariance operator from (1) is given by

$$(C^{(y)}v)(t) = \int_0^1 K^{(y)}(t,s)v(s)ds, \qquad v \in \mathcal{H}. \tag{18}$$

HVNs leverage the whole covariance structure of the data. This means that, if the process is stationary, i.e., $k_t(t,s) = k_t(|t-s|)$, they are able to learn both from second order moments (the diagonal entries of the compressed covariance matrices) that are dependent only on a single channel at a certain time lag, and mixed second order moments (the off diagonal entries of the compressed covariance matrices) that are jointly dependent on all the channels at a certain time lag. For this reason, to structure a classification task whose discriminant elements can be captured by HVNs and not necessarily by models that only act component-wise, e.g., a linear model or an MLP, we deliberately share the temporal kernel across classes and make the label depend only on cross–channel correlation, i.e.:

$$k_t(t,s)^{(0)} = k_t(t,s)^{(1)} = \exp\Big(-\frac{(t-s)^2}{2\phi^2}\Big),$$
$$\text{and} \qquad \Sigma_c^{(0)} = I_d, \qquad \Sigma_c^{(1)} = \big[\rho^{|i-j|}\big]_{i,j=1}^{d}, \tag{19}$$

where $\phi, \rho > 0$. In this task, a data point with class y consists of a bag of n i.i.d. samples drawn from $x_1, \dots, x_n \overset{\text{i.i.d.}}{\sim} \mathsf{GP}\big(0, K^{(y)}\big)$, discretized as noisy $x_i^{(m)} := S_m x_i + \epsilon \in \mathbb{R}^m$, where ϵ is AWGN, and centered by $\overline{x}^{(m)} = \frac{1}{n}\sum_{i=1}^n x_i^{(m)}$. We compare HVNs against two baselines acting on the same discretized bags: (i) MLP, obtained by simply setting $J = 1$ and $\hat{\mathbf{C}}_n^{(m)} = I_m$ in (7), and (ii) FPCA, consisting of computing the first O discrete FPCA scores (i.e., the PCA scores of the discrete $\{x_i^{(m)}\}_n$ in the bag). For all the tested models, the output of the last layer (for HVN and MLP) or the scores (for PCA) are fed to an MLP classifier to get the class probabilities. We set $\phi = 0.20$, and $\rho = 0.7$, train with ADAM optimizer to minimize cross–entropy loss, we normalize the compressed covariance matrices for training stability (for HVN), and we use GELU nonlinearity (for HVN and MLP), mean pooling, and a polynomial degree (for HVN) $J = 2$. For both HVN and MLP, we use 2 layers, 32 output signals in each HVN layer, and we adjust the MLP hidden dimensions to match the number of parameters of HVN. We fix $d = 4$ and $p = 32$, and we generate balanced train/test sets for different numbers of samples n and signal to noise ratios SNR_{dB}, defined, given a bag, as $10\log_{10}\Big((1/n)\sum_{i=1}^n \|x_i^{(m)}\|^2/\text{var}(\epsilon)\Big)$. In Figures 1a-1b we show the test accuracy vs n (for fixed $SNR_{dB} = 30$) and vs SNR_{dB} (for fixed $n = 24$), respectively. The MLP achieves mediocre accuracies (and is prone to overfitting) as it never mixes the m components of the samples, accessing only first–order information, whereas the label is encoded in cross–feature covariance. The FPCA baseline gives random accuracy as it rotates each bag to its variance–maximizing basis, leading to coordinate drift across samples (sign flips, permutations,...), which a single classifier must reconcile. In contrast, HVN shows strong performance thanks to the transferable nature of HVFs and their ability to capture discriminative second–order structure consistently across bags. We further evaluate HVNs on the ECG5000 dataset from the UCR repository [29], which contains 5000 ECG samples of length 140 and 5 classes denoting the heartbeat type. We again compare HVNs, MLP, and FPCA with bin averaging and varying discretization resolution m. The hyperparameters are set as in the synthetic experiment. In this case, the goal is to classify individual time-series and not (possibly streaming) bags; thus, both for HVNN and FPCA, we compute the empirical covariance matrix on the entire training set and then use it to classify the individual time-series. In this way, both HVNN and FPCA have a more robust estimate of the covariance operator. Indeed, Figure 1c shows that HVN and FPCA outperform MLP across all values of m, indicating the importance of covariance information. Moreover, HVN outperforms both MLP and FPCA by a margin, especially for increasing m, corroborating its capability to robustly incorporate covariance information into its end-to-end learning dynamics.

6. CONCLUSION

We proposed a novel convolutional learning framework for signals in (possibly infinite-dimensional) Hilbert spaces, built around the (empirical) covariance operator. The generality of our framework suggests several future directions. We will study the convergence properties of discrete HVFs/HVNs to their infinite-dimensional counterparts as the discretization resolution becomes infinitely fine-grained. Using these results, we will analyze the transferability properties of HVNs.

Authorized licensed use limited to: TU Delft Library. Downloaded on May 12,2026 at 06:40:35 UTC from IEEE Xplore. Restrictions apply.

7. REFERENCES

[1] Alaa Bessadok, Mohamed Ali Mahjoub, and Islem Rekik, "Graph neural networks in network neuroscience," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 5833–5848, 2022.

[2] José Vinícius de Miranda Cardoso, Jiaxi Ying, and Daniel Pérez Palomar, "Algorithms for learning graphs in financial markets," *CoRR*, vol. abs/2012.15410, 2020.

[3] Yuanrong Wang and Tomaso Aste, "Network filtering of spatial-temporal gnn for multivariate time-series prediction," in *Proceedings of the 3rd ACM International Conference on AI in Finance (ICAIF '22)*, New York, NY, USA, 2022, pp. 463–470, ACM.

[4] Yan Wang, Xin Wang, Hongmei Yang, et al., "Mhagnn: A novel framework for wearable sensor-based human activity recognition combining multi-head attention and graph neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–14, 2023.

[5] Tianzheng Liao, Jinjin Zhao, Yushi Liu, Kamen Ivanov, Jing Xiong, and Yan Yan, "Deep transfer learning with graph neural network for sensor-based human activity recognition," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2022, pp. 2445–2452.

[6] Ian Jolliffe, *Principal Component Analysis*, Springer, New York, 2002.

[7] Debashis Paul, "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model," *Statistica Sinica*, vol. 17, no. 4, pp. 1617–1642, 2007.

[8] Jinho Baik, Gérard Ben Arous, and Sandrine Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," *The Annals of Probability*, vol. 33, no. 5, pp. 1643–1697, 2005.

[9] Saurabh Sihag, Gonzalo Mateos, Corey T. McMillan, and Alejandro Ribeiro, "covariance neural networks," in *Advances in Neural Information Processing Systems*, 2022, vol. 35, pp. 17003–17016.

[10] Fernando Gama, Joan Bruna, and Alejandro Ribeiro, "Stability properties of graph neural networks," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5680–5695, 2020.

[11] Emanuel Parzen, "Statistical inference on time series by hilbert space methods, i," Tech. Rep., 1959.

[12] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd, "Quantum machine learning," *Nature*, vol. 549, no. 7671, pp. 195–202, 2017.

[13] Antonio G. Marques, Santiago Segarra, Geert Leus, and Alejandro Ribeiro, "Stationary graph processes and spectral estimation," *IEEE Transactions on Signal Processing*, vol. 65, no. 22, pp. 5911–5926, 2017.

[14] Nathanaël Perraudin and Pierre Vandergheynst, "Stationary signal processing on graphs," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3462–3477, 2017.

[15] Li Rui, Hemanth Nejat, and Ngai-Man Cheung, "Dimensionality reduction of brain imaging data using graph signal processing," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 1329–1333.

[16] Nika Shahid, Nathanaël Perraudin, Vassilis Kalofolias, Gilles Puy, and Pierre Vandergheynst, "Fast robust pca on graphs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 740–756, 2016.

[17] Saurabh Sihag, Gonzalo Mateos, Corey T. McMillan, and Alejandro Ribeiro, "Transferability of covariance neural networks and application to interpretable brain age prediction using anatomical features," *arXiv preprint arXiv:2305.01807*, 2023.

[18] Saurabh Sihag, Gonzalo Mateos, Corey McMillan, and Alejandro Ribeiro, "Explainable brain age prediction using covariance neural networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2024, vol. 36.

[19] Andrea Cavallo, Mohammad Sabbaqi, and Elvin Isufi, "Spatiotemporal covariance neural networks," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2024, pp. 18–34.

[20] Andrea Cavallo, Madeline Navarro, Santiago Segarra, and Elvin Isufi, "Fair covariance neural networks," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5.

[21] Om Roy, Yashar Moshfeghi, and Keith Smith, "Covariance density neural networks," *arXiv preprint arXiv:2505.11139*, 2025.

[22] Han Lin Shang, "A survey of functional principal component analysis," *AStA Advances in Statistical Analysis*, vol. 98, no. 2, pp. 121–142, 2014.

[23] Halsey Lawrence Royden and Patrick Fitzpatrick, *Real analysis*, vol. 32, Macmillan New York, 1988.

[24] Ian T Jolliffe, "A note on the use of principal components in regression," *Journal of the Royal Statistical Society Series C: Applied Statistics*, vol. 31, no. 3, pp. 300–303, 1982.

[25] Martin Vetterli, Jelena Kovačević, and Vivek K Goyal, *Foundations of signal processing*, Cambridge University Press, 2014.

[26] Elvin Isufi, Fernando Gama, David I Shuman, and Santiago Segarra, "Graph filters for signal processing and machine learning on graphs," *IEEE Transactions on Signal Processing*, vol. 72, pp. 4745–4781, 2024.

[27] Claudio Battiloro, Zhiyang Wang, Hans Riess, Paolo Di Lorenzo, and Alejandro Ribeiro, "Tangent bundle convolutional learning: from manifolds to cellular sheaves and back," *IEEE Transactions on Signal Processing*, vol. 72, pp. 1892–1909, 2024.

[28] Claudio Battiloro, Andrea Cavallo, and Elvin Isufi, "covariance filters and neural networks over hilbert spaces," *arXiv preprint*, 2025.

[29] Hoang Anh Dau, Anthony Bagnall, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanamahatana, and Eamonn Keogh, "The ucr time series archive," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 6, pp. 1293–1305, November 2019.

Authorized licensed use limited to: TU Delft Library. Downloaded on May 12,2026 at 06:40:35 UTC from IEEE Xplore. Restrictions apply.