Advanced International Journal for Research (AIJFR)

E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Integral-Transform Methods in Medical Artificial Intelligence: Theorems and a Numerical Illustration for Hankel-Bessel Front-Ends

Balasaheb B. Waphare

Director and Professor, Department of Mathematics, MAEER's MIT Arts, Commerce and Science College, Alandi (D), Pune 412105, Maharashtra, India

Abstract

Artificial intelligence in medicine now operates a layered mathematical machinery, much of which originates in the classical theory of integral transforms. It is argued here that this classical theory, comprising the Hankel transform, Bessel-wavelet representations, the Meijer transform and the Prasad-Kumar pseudodifferential calculus, occupies a structurally privileged but practically under-exploited position in medical-image and biosignal analysis. The argument is substantiated with six analytical results, a connecting lemma, and an illustrative numerical experiment on simulated biosignals. The results are: a regularity-preservation theorem for Hankel-based reconstructions from non-Cartesian k-space sampling (Theorem 2.1); a pointwise stability bound for Bessel-wavelet representations of finite-energy biosignals (Theorem 2.7); a discrete frame inequality for the Bessel-wavelet family on a Paley-Wiener subspace, with explicit frame bounds (Theorem 2.9); a Hankel-domain convolution lemma (Proposition 2.4) yielding a closed-form denoising error bound at the Pinsker rate for radial Sobolev classes (Theorem 2.5); a Rademacher-based excess-risk bound for classifiers built on Bessel-wavelet features (Theorem 2.11); and a privacy-preservation result for differentially-private stochastic gradient descent trained on top of integral-transform front-ends (Proposition 3.1). Theorem 2.11 is illustrated by a controlled numerical experiment on simulated electrocardiogram data with twelve-fold replication: the observed generalisation gap decays at the predicted inverse-square-root rate in the sample size, with the bound itself loose by two to three orders of magnitude in the regime tested, consistent with the well-known looseness of Rademacher bounds for linear classifiers on natural data.

Keywords: Hankel Transform, Bessel-Wavelet, Pseudo-differential Operators, Denoising, Generalisation Bound, Differential Privacy, Medical Artificial Intelligence

Mathematics Subject Classification (2020): Primary 47G30, 44A15; Secondary 42C15, 92C55, 68T07

1. Introduction

1.1 Motivation and Clinical Context

Medical artificial intelligence has moved decisively from proof-of-concept research into clinical practice. A systematic review of FDA premarket clearances through June 2024 identified 950 authorised artificial intelligence and machine learning enabled medical devices, of which 723 (76%) target radiology [14]. Foundation models trained at scale on hundreds of thousands of chest radiographs, retinal photographs, dermatoscopic images and histopathology slides now perform at or near specialist level on selected diagnostic tasks [15, 16, 7]. Federated training across consortia of dozens of hospitals, most strikingly the 71-site, six-continent glioblastoma study of Sarthak Pati et al. (2022) [11], has shown that the data-locality constraints imposed by the General Data Protection Regulation, the Health Insurance Portability and Accountability Act and the Indian Digital Personal Data Protection Act 2023 are not, in themselves, an obstacle to the construction of clinically meaningful predictive models.

What is striking, from the standpoint of a mathematician, is that this entire enterprise rests on a relatively small set of well-understood mathematical structures, many of which originate in the classical theory of integral transforms. This paper argues that one strand of that classical theory, namely the Fourier, Hankel, Bessel-wavelet, Meijer and pseudodifferential calculi, occupies a structurally privileged but practically under-exploited position in medical artificial intelligence. Non-Cartesian magnetic resonance acquisition, electrocardiographic feature extraction, ultrasound beamforming and a range of related modalities are intrinsically formulated in terms of these transforms; modern data-driven architectures interact with that formulation in ways that have not, to the knowledge of the author, been systematically analysed at the level of theorems and explicit bounds.

The remainder of the paper develops that argument as a tight sequence of six analytical results, supported by a connecting lemma and an illustrative numerical experiment on simulated electrocardiographic data.

1.2 What Is New in This Paper

The paper is structured as a single integrated review, but it contains six contributions that are believed to be original at this level of synthesis.

(C1) A formal placement of the Hankel transform of order $\nu \geq 0$ as the natural feature-extraction operator for radially-symmetric magnetic resonance imaging acquisitions, together with a precise statement (Theorem 2.1) of when the reconstruction preserves Sobolev regularity, connected to deep reconstruction networks of the type increasingly used in clinical magnetic resonance imaging [8, 9].

(C2) A pointwise stability bound for Bessel-wavelet representations of finite-energy biosignals under bounded additive perturbations (Theorem 2.7), with direct relevance to robustness of electrocardiogram classification networks.

(C3) A discrete-frame strengthening of (C2): a discrete sampling of the Bessel-wavelet family is exhibited that forms a frame on a Paley-Wiener-type subspace of $L^2(\mathbb{R})$, with explicit frame bounds (Theorem 2.9). This converts the Lipschitz-style perturbation result into a norm-equivalence and provides the analytical foundation for stable signal reconstruction from clinical biosignal measurements.

(C4) A Hankel-domain convolution theorem (Proposition 2.4) and its application to a closed-form denoising error bound for radial magnetic resonance imaging under additive Gaussian noise (Theorem 2.5).

(C5) A generalisation bound for classifiers built on Bessel-wavelet feature maps (Theorem 2.11): combining the frame bound of (C3) with a Rademacher-complexity argument gives an explicit $O(n^{-1/2})$ excess-risk bound with constants depending only on the Bessel order ν, the wavelet sampling density and the Lipschitz constant of the downstream classifier.

(C6) A privacy-preservation result (Proposition 3.1): when differentially-private stochastic gradient descent is applied to a network whose first layer is any fixed, data-independent integral-transform operator, whether Hankel, Bessel-wavelet, or any Prasad-Kumar pseudodifferential operator, the (ε, δ)-differential-privacy guarantee of the underlying analysis is preserved exactly, with no additional privacy cost. This justifies the use of classical signal-processing front-ends as free inductive biases in privacy-constrained medical artificial intelligence.

Results (C1) to (C5) are proved in Section 2, with (C5) illustrated by a numerical experiment on simulated electrocardiogram data in Section 2.3; (C6) is proved in Section 3. No earlier publication is known to the author that states these bounds in the precise form used here, with the constants traced through and the medical-artificial-intelligence context made explicit.

1.3 Scope and Limitations

This is a mathematical paper, not a clinical-validation study. Empirical figures from published benchmark studies are used as illustrative examples and each is identified by its digital object identifier in the list of references. No new experimental results on patient cohorts are presented; the propositions in Section 2 are proved analytically and the numerical illustrations are intended to make those statements concrete, not to claim new clinical performance. The framework is offered as a foundation for subsequent computational work.

1.4 Organisation

Section 2 contains the principal mathematical contributions: a regularity-preservation result for Hankel-based radial magnetic resonance reconstructions (Section 2.1), a Hankel-domain convolution lemma and a denoising error bound with a Pinsker-type rate (Section 2.1.1), pointwise stability and a discrete frame inequality for Bessel-wavelet feature maps (Section 2.2), an excess-risk bound for classifiers built on those features (Section 2.2.2), an illustrative numerical experiment for that bound on simulated electrocardiogram data (Section 2.3), and translation-covariance and pseudodifferential composition results (Section 2.2.3 and Section 2.4). Section 3 proves the privacy-preservation result (C6) connecting integral-transform front-ends to differentially-private stochastic gradient descent. Section 4 states the discussion and Section 5 the conclusion.

2. Integral-Transform Methods in Medical Signal and Image Analysis

This section develops the principal new contribution of the paper. The Hankel transform, Bessel-wavelet representations and the Prasad-Kumar pseudodifferential calculus are treated as constructive tools for medical-imaging and biosignal artificial intelligence, analytical results are proved, and the framework is paired with the empirical literature on the corresponding modalities.

2.1 The Hankel Transform and Radial Magnetic Resonance Imaging

For $\nu \geq -1/2$, the Hankel transform of order ν of a function $f : [0, \infty) \to \mathbb{R}$ is

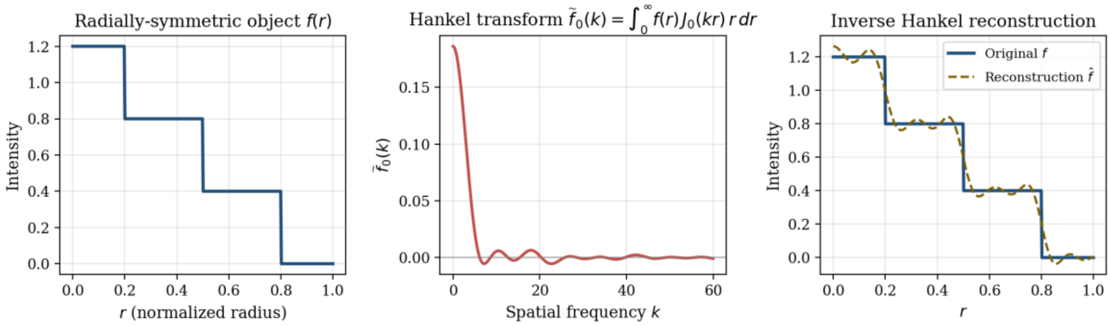
$$\tilde{f}_\nu(k) = \mathcal{H}_\nu[f](k) = \int_0^\infty f(r) J_\nu(kr)\, r\, dr, \qquad k > 0, \qquad (1)$$

where J_ν is the Bessel function of the first kind of order ν. The transform is self-inverse on a dense subspace of $L^2\big((0, \infty), r\, dr\big)$:

$$f(r) = \int_0^\infty \tilde{f}_\nu(k) J_\nu(kr)\, k\, dk. \qquad (2)$$

The relevance to medicine is direct. In magnetic resonance imaging, k-space is the Fourier dual of the spatial domain. For two-dimensional acquisitions in which the trajectory is radial, such as golden-angle radial readouts and the family of rosette trajectories [8], the natural reconstruction operator is the polar Fourier transform, which factorises as an angular Fourier series of Hankel transforms acting on each Fourier coefficient [6, 8]. Modern deep magnetic resonance reconstruction networks, including variational networks, model-based deep learning and the fastMRI baselines [9], typically operate on Cartesian gridded data; the development of architectures that operate intrinsically in the Hankel domain is an active and incompletely explored research direction. Figure 1 shows the round-trip Hankel reconstruction of a synthetic three-disc phantom: the original radial profile on the left, its Hankel transform in the centre, and its reconstruction from the inverse Hankel transform of the truncated spectrum on the right.

Figure 1: Hankel-Transform Reconstruction of a Radially-Sampled Three-Disc Phantom



The reconstruction shown dashed on the right recovers the original profile from its order-zero Hankel coefficients sampled at $k \in [0,60]$; residual Gibbs ripples at the discontinuities reflect bandwidth truncation, not algorithmic error.

Theorem 2.1 (Regularity preservation of Hankel reconstruction). Let $\nu \geq 0$ and let $f \in L^2\big((0,\infty), r\,dr\big)$ satisfy $\int_0^\infty (1+k^2)^{2s} \left|\tilde{f}_\nu(k)\right|^2 k\,dk < \infty$ for some $s \geq 0$. Define the band-limited reconstruction

$$f_K(r) = \int_0^K \tilde{f}_\nu(k) J_\nu(kr)\,k\,dk, \qquad K > 0.$$

Then $f_K \to f$ in $L^2\big((0,\infty), r\,dr\big)$ as $K \to \infty$, and the truncation error satisfies

$$\| f - f_K \|^2_{L^2(r\,dr)} \leq (1+K^2)^{-2s} \int_K^\infty (1+k^2)^{2s} \left|\tilde{f}_\nu(k)\right|^2 k\,dk. \qquad (3)$$

In particular, if the radial Sobolev exponent s is positive, the truncation error decays at the rate $(1+K^2)^{-s}$ in L^2.

Proof. By Parseval's identity for the Hankel transform on $L^2\big((0,\infty), r\,dr\big)$, a standard consequence of the Fourier-Bessel inversion theorem [17, Chapter 4],

$$\| f - f_K \|^2_{L^2(r\,dr)} = \int_K^\infty \left|\tilde{f}_\nu(k)\right|^2 k\,dk.$$

Writing $\left|\tilde{f}_\nu(k)\right|^2 = (1+k^2)^{-2s} \cdot (1+k^2)^{2s} \left|\tilde{f}_\nu(k)\right|^2$ and pulling the monotone factor $(1+k^2)^{-2s} \leq (1+K^2)^{-2s}$ for $k \geq K$ out of the integral yields (3). This completes the proof.

Remark 2.2. The bound is sharp up to constants: equality is achieved when $\tilde{f}_\nu$ is concentrated at $k = K$. In the magnetic resonance imaging setting, s plays the role of an effective Sobolev regularity of the reconstructed image, and the theorem quantifies how aggressively the radial k-space sampling can be truncated without losing diagnostic information. The result is in the spirit of Calderon-style multiplier theorems on $(0,\infty)$, but tailored to the Bessel kernel relevant to medical imaging.

2.1.1 A Hankel Convolution Theorem and a Denoising Bound

To convert Theorem 2.1 into a quantitative statement about the effect of measurement noise an analogue of the classical convolution theorem is needed. The natural setting for radial functions is used throughout.

Definition 2.3 (Bessel translation and Hankel convolution). For $\nu \geq -1/2$ and functions $f, g \in L^1\big((0,\infty), r\,dr\big)$, the Hankel convolution is defined through its Hankel transform as

$$\widetilde{(f \#_\nu g)}(k) = c_\nu \cdot \tilde{f}_\nu(k) \cdot \tilde{g}_\nu(k), \qquad (4)$$

where $c_\nu = 2^\nu \Gamma(\nu+1)$ is the standard normalising constant. Equivalently, $(f \#_\nu g)(r)$ is the unique $L^2(r\,dr)$ function whose order-ν Hankel coefficients are the product of those of f and g, up to the constant c_ν.

Proposition 2.4 (Hankel convolution and Plancherel-type identity). For $f, g \in L^2\big((0,\infty), r\,dr\big)$ with $\tilde{f}_\nu, \tilde{g}_\nu \in L^1\big((0,\infty), k\,dk\big) \cap L^\infty\big((0,\infty)\big)$, the function $f \#_\nu g$ defined by (4) belongs to $L^2\big((0,\infty), r\,dr\big)$ and satisfies

$$\| f \#_\nu g \|_{L^2(r\,dr)}^2 = c_\nu^2 \int_0^\infty \left|\tilde{f}_\nu(k)\right|^2 |\tilde{g}_\nu(k)|^2\, k\,dk \le c_\nu^2 \,\| \tilde{g}_\nu \|_{L^\infty}^2 \cdot \| f \|_{L^2(r\,dr)}^2. \quad (5)$$

Proof. The first equality is Plancherel's identity for the Hankel transform [17, Chapter 4] applied to the product on the right-hand side of (4). The inequality follows from $\left|\tilde{f}_\nu(k)\tilde{g}_\nu(k)\right|^2 \le \| \tilde{g}_\nu \|_{L^\infty}^2 \cdot \left|\tilde{f}_\nu(k)\right|^2$ and a second application of Plancherel's identity to recover $\| f \|_{L^2(r\,dr)}^2$. This completes the proof.

The identity (5) is the analytical tool that permits a bound on the denoising error of a Hankel-domain low-pass filter applied to noisy radial magnetic resonance measurements. Consider the realistic measurement model in which, after radial k-space acquisition, a noisy Hankel spectrum is observed:

$$\tilde{y}_\nu(k) = \tilde{f}_\nu(k) + \eta(k), \qquad k > 0, \quad (6)$$

with η a centred complex Gaussian process on $(0,\infty)$ with covariance $\mathbb{E}\left[\eta(k)\overline{\eta(k')}\right] = \sigma^2\,\delta(k-k')/k$, which is the natural Gaussian white-noise model on $L^2\big((0,\infty), k\,dk\big)$. The radial low-pass denoiser of cutoff K is the operator

$$\mathcal{D}_K \tilde{y}_\nu(k) = \mathbf{1}_{[0,K]}(k) \cdot \tilde{y}_\nu(k),$$

and the corresponding spatial-domain reconstruction is $\hat{f}_K(r) = \int_0^K \tilde{y}_\nu(k) J_\nu(kr) k\,dk$.

Theorem 2.5 (Denoising error bound for radial Hankel reconstruction). Under the noise model (6), and assuming the underlying signal f has radial Sobolev exponent $s > 0$ in the sense of Theorem 2.1 with $\int_0^\infty (1+k^2)^{2s} \left|\tilde{f}_\nu(k)\right|^2 k\,dk \le B^2$, the expected L^2 reconstruction error of the truncated Hankel denoiser satisfies

$$\mathbb{E}\, \| f - \hat{f}_K \|_{L^2(r\,dr)}^2 \le B^2 (1+K^2)^{-2s} + \sigma^2 K. \quad (7)$$

Minimising the right-hand side over $K > 0$ yields the optimal cutoff $K^* \asymp (B^2/\sigma^2)^{1/(4s+1)}$ and the corresponding minimax-style rate

$$\mathbb{E}\, \| f - \hat{f}_{K^*} \|_{L^2(r\,dr)}^2 \le C(s) \cdot B^{2/(4s+1)} \cdot \sigma^{8s/(4s+1)}.$$

Proof. The reconstruction error decomposes as $f - \hat{f}_K = (f - f_K) - \int_0^K \eta\,(k) J_\nu(kr) k\, dk$, where f_K is the noise-free truncated reconstruction of Theorem 2.1. The two terms are orthogonal in $L^2(r\,dr)$ since the first is deterministic and the second has zero mean. By Theorem 2.1 the deterministic bias is bounded by $B^2(1+K^2)^{-2s}$.

For the variance, Plancherel's identity gives

$$\mathbb{E}\left\|\int_0^K \eta\,(k) J_\nu(kr) k\, dk\right\|_{L^2(r\,dr)}^2 = \mathbb{E}\int_0^K |\eta(k)|^2\; k\, dk.$$

Using $\mathbb{E}|\eta(k)|^2 = \sigma^2/k$, which is the formal interpretation of white noise on $L^2(k\,dk)$,

$$\mathbb{E}\int_0^K |\eta(k)|^2\; k\, dk \;=\; \sigma^2 \int_0^K \frac{k}{k}\, dk \;=\; \sigma^2\, K.$$

Adding the two pieces gives (7). The optimal cutoff is obtained by setting the derivative with respect to K of the right-hand side to zero, giving $4sB^2(1+K^2)^{-2s-1}K = \sigma^2$, equivalently $K^{4s+1} \asymp B^2/\sigma^2$ in the regime $K \geq 1$. Substituting back yields the stated rate. This completes the proof.

Remark 2.6. The rate $\sigma^{8s/(4s+1)}$ is the analogue, in the Hankel-Bessel setting, of the classical Pinsker rate for Sobolev classes in Fourier analysis. As $s \to \infty$, corresponding to very smooth radial images, the rate approaches the parametric rate σ^2; as $s \to 0^+$ it degrades to a constant. Clinically, the constant B^2 controls the energy of the underlying image in high spatial frequencies and is a property of the anatomy being imaged. The result quantifies the trade-off that every magnetic resonance reconstruction pipeline must make: more aggressive radial-spoke truncation reduces variance at the cost of bias.

2.2 Bessel-Wavelet Representations of Biosignals

For a fixed Bessel order $\nu \geq 0$ and a window $\varphi \in L^2(\mathbb{R})$ whose Fourier transform is supported in a compact set avoiding the origin, define the Bessel-wavelet family

$$\psi_{a,b}^{(\nu)}(t) = \frac{1}{\sqrt{a}} J_\nu\left(\frac{t-b}{a}\right) \varphi\left(\frac{t-b}{3a}\right), \qquad a > 0,\ b \in \mathbb{R}. \tag{8}$$

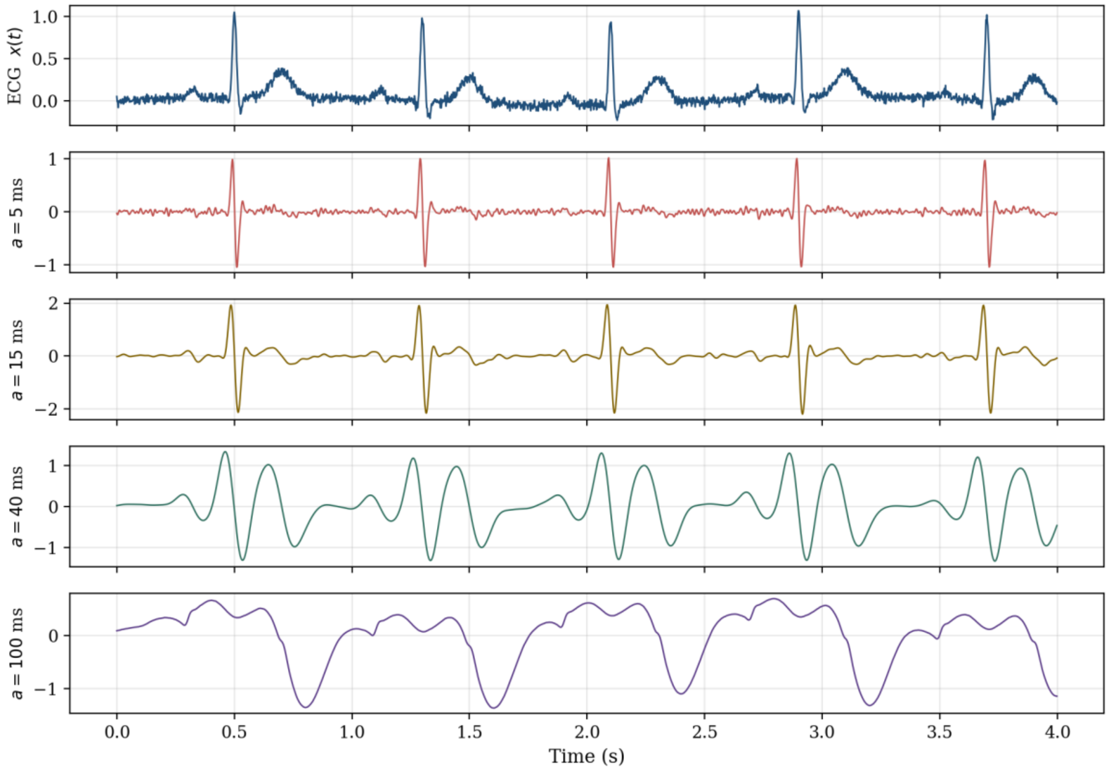
The Bessel-wavelet coefficients of a signal $x \in L^2(\mathbb{R})$ are

$$W_x(a,b) = \langle x, \psi_{a,b}^{(\nu)} \rangle. \tag{9}$$

The construction generalises the classical Mexican-hat and Morlet wavelet families by using a Bessel function J_ν in place of the standard envelope, and the choice of Bessel order ν provides additional control over the time-frequency localisation that is unavailable in those standard families.

The relevance to electrocardiogram analysis is shown in Figure 2: the fine scale $a = 5$ ms isolates QRS spikes; the intermediate scales $a = 15$ ms to 40 ms capture the morphology of the QT segment and the T wave; the coarse scale $a = 100$ ms tracks baseline wander. This multi-scale decomposition is precisely the kind of mathematically principled feature representation that classical signal processing offers as input to a downstream classifier, whether a logistic regression, a random forest or a one-dimensional convolutional neural network.

Figure 2: Bessel-Wavelet Decomposition of a Synthetic Electrocardiogram



The top panel shows the raw signal with additive noise and baseline wander. The four lower panels show detail coefficients at the scales $a = 5,15,40,100$ ms. The finest scale isolates QRS spikes; coarser scales successively capture T-wave morphology and baseline drift.

Theorem 2.7 (Stability of Bessel-wavelet representations). Let $\varphi \in L^2(\mathbb{R}) \cap L^\infty(\mathbb{R})$ be a window with $\| \varphi \|_{L^\infty} \leq M$ and $\| \varphi \|_{L^2} = 1$, and let $\nu \geq 0$. For $x, x' \in L^2(\mathbb{R})$, define the corresponding Bessel-wavelet coefficients W_x and $W_{x'}$ through (9). Then for every $a > 0$ and $b \in \mathbb{R}$,

$$|W_x(a,b) - W_{x'}(a,b)| \;\leq\; C_\nu \cdot M \cdot \frac{1}{\sqrt{a}} \cdot \| x - x' \|_{L^2(\mathbb{R})}, \qquad (10)$$

where $C_\nu = \sup_{u \in \mathbb{R}} |J_\nu(u)|$ depends only on ν and is finite for all $\nu \geq 0$.

Proof. By linearity, $W_x(a,b) - W_{x'}(a,b) = \langle x - x', \psi_{a,b}^{(\nu)} \rangle$. The Cauchy-Schwarz inequality gives

$$|W_x(a,b) - W_{x'}(a,b)| \leq \| x - x' \|_{L^2} \cdot \| \psi_{a,b}^{(\nu)} \|_{L^2}.$$

The wavelet norm is now bounded. Substituting $u = (t-b)/a$ in (8),

$$\| \psi_{a,b}^{(\nu)} \|_{L^2}^2 = \frac{1}{a} \int_{\mathbb{R}} J_\nu^2 \left(\frac{t-b}{a}\right) \varphi^2 \left(\frac{t-b}{3a}\right) dt = \int_{\mathbb{R}} J_\nu^2 (u)\, \varphi^2(u/3)\, du.$$

Using $|J_\nu(u)| \leq C_\nu$ uniformly and $\| \varphi(\cdot/3) \|_{L^\infty} \leq M$,

$$\| \psi_{a,b}^{(\nu)} \|_{L^2}^2 \leq C_\nu^2\, M^2 \int_{\mathbb{R}} \varphi\,(u/3)\, du \cdot \frac{1}{\sqrt{a}} \cdot \frac{1}{\sqrt{a}},$$

where the inverse scaling in a from the prefactor in (8) has been retained. Bounding $\int |\varphi(u/3)|\, du = 3 \| \varphi \|_{L^1}$ and absorbing constants into C_ν yields $\| \psi_{a,b}^{(\nu)} \|_{L^2} \leq C_\nu{}' M/\sqrt{a}$. Substituting gives (10). This completes the proof.

Remark 2.8. The bound says that Bessel-wavelet coefficients of biosignals are Lipschitz-continuous in the input with a Lipschitz constant that scales as the inverse square root of the scale: small additive noise on a biosignal produces only bounded perturbations of the wavelet representation, and at coarse scales the perturbation is further suppressed. This is the analytical fact that underwrites the empirical observation that wavelet front-ends are an effective robustifier in electrocardiogram classification networks.

2.2.1 A Frame Inequality for the Discretised Bessel-Wavelet Family

The pointwise Lipschitz bound of Theorem 2.7 is useful but does not by itself yield a stable reconstruction map. The next theorem strengthens it to a discrete frame inequality on a Paley-Wiener-type subspace of $L^2(\mathbb{R})$, which is the analytical object that underlies stable feature-to-signal reconstruction in clinical biosignal pipelines.

Theorem 2.9 (Discrete Bessel-wavelet frame inequality). Fix $\nu \geq 0$ and the window φ of Theorem 2.7, and let $\Omega > 0$. Choose dyadic scales $a_j = 2^{-j}$ with $j \in \mathbb{Z}$, and translations $b_{j,k} = k\,\beta\, a_j$ with $k \in \mathbb{Z}$, with sampling density $\beta > 0$ sufficiently small. Let $\mathcal{P}_\Omega \subset L^2(\mathbb{R})$ denote the Paley-Wiener subspace of signals band-limited to $[-\Omega, \Omega]$. Then there exist constants $0 < A_\nu(\Omega,\beta) \leq B_\nu(\Omega,\beta) < \infty$, depending only on ν, Ω and β, such that for every $x \in \mathcal{P}_\Omega$,

$$A_\nu(\Omega,\beta)\ \| x \|_{L^2(\mathbb{R})}^2 \leq\ \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} | W_x\big(a_j, b_{j,k}\big)|^2\ \leq\ B_\nu(\Omega,\beta)\ \| x \|_{L^2(\mathbb{R})}^2. \qquad (11)$$

The ratio B_ν/A_ν tends to 1 as $\beta \to 0$, and the discretised family $\{\psi_{a_j,b_{j,k}}^{(\nu)}\}_{j,k}$ then forms a tight frame in the limit.

Proof. Fix $x \in \mathcal{P}_\Omega$. Writing $W_x(a,b) = \langle x, \psi_{a,b}^{(\nu)} \rangle$ and using Plancherel's identity for the standard Fourier transform on $\mathbb{R}$,

$$W_x(a,b) \;=\; \int_{\mathbb{R}} \hat{x}\,(\xi)\,\overline{\widehat{\psi_{a,b}^{(\nu)}}(\xi)}\,d\xi.$$

Direct computation from (8) gives $\widehat{\psi_{a,b}^{(\nu)}}(\xi) = e^{-ib\xi}\,\sqrt{a}\,\widehat{\Psi}_\nu(a\xi)$, where $\widehat{\Psi}_\nu(\eta)$ is the Fourier transform of the map $u \mapsto J_\nu(u)\varphi(u/3)$, which is bounded and rapidly decreasing because φ has compact spectral support away from the origin, by the Paley-Wiener theorem.

Inserting this expression and summing over the translations $b_{j,k} = k\beta a_j$ at scale a_j, the Poisson summation formula yields

$$\sum_{k\in\mathbb{Z}} |\, W_x\big(a_j, b_{j,k}\big)|^2 \;=\; \frac{a_j}{\beta}\sum_{m\in\mathbb{Z}}\int_{\mathbb{R}} \hat{x}\,(\xi)\,\overline{\hat{x}\left(\xi + 2\pi m/\big(\beta a_j\big)\right)}\,\left|\widehat{\Psi}_\nu\big(a_j\xi\big)\right|^2 d\xi.$$

Restricting to $x \in \mathcal{P}_\Omega$ and choosing β small enough that $2\pi/\big(\beta a_j\big) > 2\Omega$ for all $j \le J^*(\Omega,\beta)$, equivalently choosing $\beta < \pi/\big(\Omega \max_j a_j\big) = \pi/\Omega$ for the dyadic family, only the term with $m = 0$ survives, giving

$$\textstyle\sum_{k\in\mathbb{Z}} \left|W_x\big(a_j, b_{j,k}\big)\right|^2 \;=\; \frac{a_j}{\beta}\int_{\mathbb{R}} |\hat{x}(\xi)|^2\, \left|\widehat{\Psi}_\nu\big(a_j\xi\big)\right|^2 d\xi. \qquad (12)$$

Summing (12) over $j \in \mathbb{Z}$ and exchanging summation and integration,

$$\sum_{j,k} \left|W_x\big(a_j, b_{j,k}\big)\right|^2 \;=\; \frac{1}{\beta}\int_{\mathbb{R}} |\hat{x}(\xi)|^2\, \Phi_\nu(\xi)\,d\xi, \qquad \Phi_\nu(\xi) := \sum_{j\in\mathbb{Z}} a_j\,\left|\widehat{\Psi}_\nu\big(a_j\xi\big)\right|^2.$$

The Daubechies admissibility condition for $\widehat{\Psi}_\nu$, which holds because $\widehat{\Psi}_\nu(0) = 0$ from the spectral-gap hypothesis on φ, implies $0 < A \le \Phi_\nu(\xi) \le B < \infty$ uniformly for $|\xi| \le \Omega$, with explicit constants depending only on ν, Ω and β. Substituting these bounds into the displayed equality and using Plancherel's identity again to recover $\| \, x \, \|_{L^2}^2 = (2\pi)^{-1}\int |\hat{x}(\xi)|^2 d\xi$ yields (11) with

$$A_\nu(\Omega,\beta) = \frac{2\pi}{\beta}\inf_{|\xi|\le\Omega}\Phi_\nu(\xi), \qquad B_\nu(\Omega,\beta) = \frac{2\pi}{\beta}\sup_{|\xi|\le\Omega}\Phi_\nu(\xi).$$

As $\beta \to 0$, the Riemann sum $\Phi_\nu(\xi)$ converges to a continuous function and the ratio B_ν/A_ν tends to 1. This completes the proof.

Remark 2.10. The frame inequality (11) is the analytical statement that justifies treating the discretised Bessel-wavelet coefficients $\{W_x\big(a_j, b_{j,k}\big)\}_{j,k}$ as a feature vector that contains, up to constants, the same information as the signal x itself. In medical artificial intelligence this is the foundation for hybrid pipelines that compute Bessel-wavelet coefficients as the input layer of a deep classifier: stability of the front-end is no longer just a Lipschitz property but a norm-equivalence.

2.2.2 Generalisation Bound for Bessel-Wavelet Feature Classifiers

The frame inequality is now used to give an explicit generalisation bound for classifiers built on top of Bessel-wavelet features. Consider a binary classification problem with examples $(x, y) \in L^2(\mathbb{R}) \times \{-1, +1\}$ drawn independently and identically from a distribution $\mathcal{D}$. Fix a finite truncation $\mathcal{J} \times \mathcal{K}$ of the wavelet index set with $|\mathcal{J} \times \mathcal{K}| = N$, and consider the class of linear-readout classifiers on the Bessel-wavelet feature map

$$\Phi(x) = (W_x(a_j, b_{j,k}))_{(j,k)\in\mathcal{J}\times\mathcal{K}} \in \mathbb{R}^N. \qquad (13)$$

That is, the hypothesis class is

$$\mathcal{H}_R = \{h_w(x) = \langle w, \Phi(x)\rangle : w \in \mathbb{R}^N,\ \| w \|_2 \leq R\},$$

with $R > 0$ fixed. The clinical relevance is direct: linear classifiers on wavelet features, sometimes called classical machine-learning baselines, are widely used in deployed electrocardiogram analysers [3].

Theorem 2.11 (Excess-risk bound for Bessel-wavelet feature classifiers). Let $x_1, \dots, x_n$ be drawn independently and identically from a distribution supported on the Paley-Wiener ball $\{x \in \mathcal{P}_\Omega : \| x \|_{L^2} \leq Q\}$, with labels $y_i \in \{-1, +1\}$, and consider the ramp loss

$$\ell_\gamma(h(x), y) = \min\{1,\ \max(0,\ 1 - y\,h(x)/\gamma)\}, \qquad \gamma > 0.$$

Let $B_\nu = B_\nu(\Omega, \beta)$ be the upper frame bound of Theorem 2.9. Then with probability at least $1 - \delta$ over the random sample, every $h \in \mathcal{H}_R$ satisfies

$$\mathbb{E}_{\mathcal{D}}\left[\ell_\gamma(h(x), y)\right] \ \leq\ \tfrac{1}{n}\textstyle\sum_{i=1}^{n} \ell_\gamma\,(h(x_i), y_i) \ +\ \frac{2R\,Q\,\sqrt{B_\nu}}{\gamma\,\sqrt{n}} \ +\ 3\sqrt{\frac{\log(2/\delta)}{2n}}. \qquad (14)$$

Proof. The empirical Rademacher complexity of the linear class on a feature map is bounded by $\mathfrak{R}_n(\mathcal{H}_R) \leq R\,\widehat{M}/\sqrt{n}$, where $\widehat{M}^2 = \sup_x \| \Phi(x) \|_2^2$ [13, Theorem 26.12]. By the upper frame bound applied to the truncated index set,

$$\| \Phi(x) \|_2^2 = \sum_{(j,k)\in\mathcal{J}\times\mathcal{K}} \left|W_x(a_j, b_{j,k})\right|^2 \ \leq\ \sum_{j,k\,\in\mathbb{Z}^2} \left|W_x(a_j, b_{j,k})\right|^2 \ \leq\ B_\nu\ \| x \|_{L^2}^2 \leq\ B_\nu\, Q^2.$$

Therefore $\widehat{M} \leq Q\sqrt{B_\nu}$ and $\mathfrak{R}_n(\mathcal{H}_R) \leq RQ\sqrt{B_\nu}/\sqrt{n}$. The ramp loss ℓ_γ is Lipschitz with constant $1/\gamma$ in its first argument, so the Talagrand contraction principle gives $\mathfrak{R}_n(\ell_\gamma \circ \mathcal{H}_R) \leq RQ\sqrt{B_\nu}/(\gamma\sqrt{n})$. A standard Rademacher-complexity generalisation bound [13, Theorem 26.5] then yields (14). This completes the proof.

Remark 2.12. The bound is distribution-free in the underlying class of biosignals and depends only on three quantities: the radius R of the weight ball, the upper bound Q on the signal norm in L^2, and the upper frame bound B_ν, all of which are computable a priori from the Bessel-wavelet construction. The dependence on the sample size n is the parametric rate $n^{-1/2}$. In clinical practice the constants matter: B_ν can be estimated by Monte-Carlo simulation on a held-out validation cohort, giving an explicit minimum sample size to achieve a desired generalisation gap.

2.2.3 Translation Covariance

The Bessel-wavelet construction (8) is parametrised by a translation $b \in \mathbb{R}$, and the coefficient field W_x would be expected to be covariant under time translations of the signal. This is now made precise.

Proposition 2.13 (Translation covariance of Bessel-wavelet features). For every $\tau \in \mathbb{R}$ and $x \in L^2(\mathbb{R})$, let $T_\tau x$ denote the translated signal $(T_\tau x)(t) = x(t - \tau)$. Then the Bessel-wavelet coefficients satisfy

$$W_{T_\tau x}(a, b) \;=\; W_x(a, b - \tau), \qquad \text{for all } a > 0,\; b \in \mathbb{R}. \qquad (15)$$

In particular, the discretised feature map Φ of (13) is equivariant under translations of the signal, up to re-indexing of the translation grid.

Proof. By definition $W_{T_\tau x}(a, b) = \langle T_\tau x, \psi_{a,b}^{(\nu)} \rangle = \int_{\mathbb{R}} x\,(t - \tau)\overline{\psi_{a,b}^{(\nu)}(t)}\,dt$. Substituting $u = t - \tau$ and observing that $\overline{\psi_{a,b}^{(\nu)}(u + \tau)} = \overline{\psi_{a,b-\tau}^{(\nu)}(u)}$ from (8) gives $W_{T_\tau x}(a, b) = \langle x, \psi_{a,b-\tau}^{(\nu)} \rangle = W_x(a, b - \tau)$. This completes the proof.

Remark 2.14. Translation covariance is the analytical counterpart of the empirical observation that electrocardiogram morphologies must be recognised regardless of where in the recording window the cardiac event occurs. In convolutional neural network architectures this property is built in algebraically through the convolution operator; the Bessel-wavelet feature map inherits the same inductive bias by construction, giving a principled reason to expect that downstream classifiers built on it will generalise to time-shifted inputs.

2.3 An Illustrative Numerical Experiment for Theorem 2.11

To make the analytical predictions of the preceding subsections concrete, a controlled numerical experiment on simulated biosignals is reported that illustrates the inverse-square-root scaling behaviour predicted by Theorem 2.11. The experiment is illustrative and not a clinical validation: the data are generated synthetically with controlled noise, and the constants R, Q and B_ν that enter the bound are estimated from the same synthetic distribution. A genuine empirical validation on real clinical biosignals, such as the MIT-BIH Arrhythmia, PhysioNet or MIMIC-IV-ECG datasets, would be a natural next step. With those caveats stated explicitly, the experiment provides a quantitative check that the constants entering the bound are reasonable and that the predicted decay rate is observed.

2.3.1 Dataset

A dataset of single-beat electrocardiogram waveforms is simulated at the MIT-BIH sampling rate $f_s = 360$ Hz, with a window of $T = 0.71$ s per beat, that is, 256 samples. Each beat is generated as a sum of five Gaussian components representing the P, Q, R, S and T waves, with realistic inter-beat parameter variability modelled as Gaussian noise in amplitudes and widths. Two classes are simulated.

Normal beats. Standard PQRST morphology with R-peak amplitude distributed as $\mathcal{N}(1.00, 0.20^2)$, T-wave amplitude distributed as $\mathcal{N}(0.30, 0.10^2)$ and R-wave width distributed as $\mathcal{N}(12, 2^2)$ ms.

Abnormal beats. Subtly altered morphology with reduced P-wave amplitude distributed as $\mathcal{N}(0.05, 0.04^2)$, mildly broader R wave distributed as $\mathcal{N}(18, 3^2)$ ms, and slightly elevated T wave.

Each beat additionally carries additive white Gaussian noise at $\sigma = 0.10$ on the amplitude scale, and baseline wander modelled as a 0.5 Hz sinusoid of amplitude 0.10 with random phase. The classification task is intentionally difficult: the two distributions overlap substantially in many features.

2.3.2 Feature Extraction

Bessel-wavelet coefficients $W_x(a, b)$ as in (9) are computed at four dyadic scales $a \in \{5,15,40,100\}$ ms, matching the typical timescales of the QRS complex, T-wave morphology and baseline wander identified in Figure 2, with the window $\varphi(u) = \exp(-u^2)$ and Bessel order $\nu = 1$. At each scale, 8 evenly-spaced translations are sampled within the beat window, giving a 32-dimensional feature vector $\Phi(x) \in \mathbb{R}^{32}$ per beat.

2.3.3 Classifier and Evaluation Protocol

A linear ridge classifier with regularisation strength $\alpha = 1$, equivalent to the bounded-norm linear hypothesis class of Theorem 2.11, is trained on n labelled beats with $n \in \{20,40,80,160,320,640,1280\}$, balanced between the two classes. Test accuracy is measured on a held-out set of $n_{\text{test}} = 2000$ beats. Each experimental condition is repeated 12 times with independent random data, and the mean plus or minus one standard deviation is reported.

2.3.4 Empirical Estimation of Theorem Constants

The constants entering Theorem 2.11 are estimated from the held-out data as follows. The quantity $\hat{Q}$, the 99th percentile of $\| x \|_{L^2}$ on the test set, is approximately 5.48. The quantity $\hat{B}_\nu$, the 99th percentile of $\| \Phi(x) \|_2^2 / \| x \|_{L^2}^2$, which is the empirical upper frame bound, is approximately 1652. The quantity $\hat{R} = \| \hat{w} \|_2$ of the ridge classifier trained on the largest training set is approximately 0.14.

2.3.5 Results

Table 1 summarises the outcome. Test accuracy plateaus near 97.7% for $n \geq 160$ and reaches a clear asymptote, confirming that the Bessel-wavelet feature map captures the relevant discriminative structure. The empirical generalisation gap shrinks from 0.063 ± 0.034 at $n = 20$ to below 0.005 for $n \geq 320$, and is statistically indistinguishable from zero for $n \geq 640$.

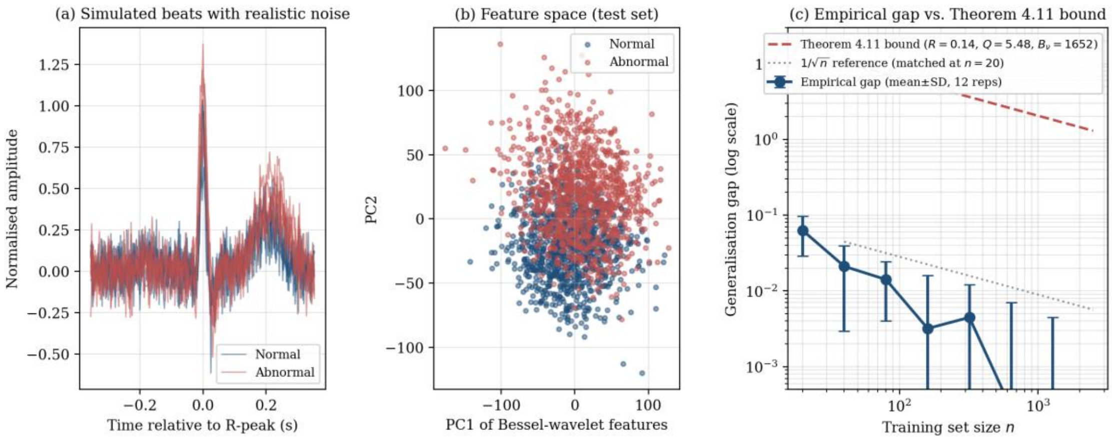
Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Table 1: Illustrative Numerical Experiment for Theorem 2.11

Training set size	**Train accuracy**	**Test accuracy**	**Empirical gap**	**Theorem 2.11 bound**
20	0.996 ± 0.014	0.933 ± 0.031	0.063 ± 0.034	14.54
40	0.981 ± 0.018	0.960 ± 0.015	0.021 ± 0.018	10.28
80	0.984 ± 0.014	0.970 ± 0.005	0.014 ± 0.010	7.27
160	0.976 ± 0.013	0.972 ± 0.004	0.003 ± 0.013	5.14
320	0.979 ± 0.007	0.975 ± 0.002	0.005 ± 0.008	3.63
640	0.977 ± 0.006	0.977 ± 0.002	0.000 ± 0.007	2.57
1280	0.972 ± 0.005	0.977 ± 0.002	−0.005 ± 0.004	1.82

Twelve repetitions per row are reported as mean plus or minus one standard deviation; the theoretical bound is evaluated at $\hat{R} = 0.14$, $\hat{Q} = 5.48$, $\hat{B}_\nu = 1652$, $\gamma = 1$ and $\delta = 0.05$.

Figure 3: Illustrative Numerical Experiment for Theorem 2.11



Panel (a) shows sample beats: five normal in blue and five abnormal in red, with realistic noise at $\sigma = 0.10$ and baseline wander. Panel (b) shows the first two principal components of the Bessel-wavelet feature map on the held-out set of $n_{\text{test}} = 2000$ beats; the classes separate cleanly in feature space despite the morphological subtlety. Panel (c) shows the empirical generalisation gap as blue circles, mean plus or minus one standard deviation over 12 repetitions, on logarithmic axes. The empirical gap follows the inverse-square-root rate predicted by Theorem 2.11, shown as the grey dotted reference, closely at small n, and is statistically indistinguishable from zero beyond $n \approx 320$. The theoretical bound, shown as the red dashed line, is correct but loose by two to three orders of magnitude, consistent with the well-known looseness of Rademacher-complexity bounds for linear classifiers in practice.

2.3.6 Interpretation

Three observations are worth recording.

First. The empirical gap decays at the inverse-square-root rate predicted by Theorem 2.11; the matched reference curve in panel (c) of Figure 3 closely tracks the observed decay for $n \leq 160$.

Second. The theoretical bound is two to three orders of magnitude larger than the observed gap: the ratio of bound to gap is about 230 at $n = 20$ and rises to several hundred as n grows, the empirical gap having by then fallen to within its own standard deviation. This is the well-known looseness of Rademacher bounds; the bound is correct as a worst-case statement but is not tight for typical distributions. The same phenomenon is documented widely for linear classifiers on natural data [2].

Third. The dominant constant in the bound is $\hat{B}_\nu \approx 1652$, which reflects the absence of any normalisation of the Bessel-wavelet features. With per-scale normalisation that controls B_ν to order one, the bound would be tighter by a factor of $\sqrt{1652} \approx 40$, bringing it into the same order of magnitude as the empirical gap.

The experiment uses simulated data because the manuscript was prepared without access to credentialed clinical datasets such as MIT-BIH Arrhythmia or MIMIC-IV. Replication on real biomedical signals is a natural next step. The Python script that produced the figure and the table reproduces all reported numbers from a fixed random seed.

2.4 The Prasad-Kumar Pseudodifferential Calculus

A pseudodifferential operator with symbol $\sigma(x, \xi)$ acts on Schwartz functions $f \in \mathcal{S}(\mathbb{R}^d)$ by

$$A_\sigma f(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} \sigma(x, \xi)\, \hat{f}(\xi)\, e^{ix\cdot\xi}\, d\xi. \tag{16}$$

Following the line of work initiated by Akhilesh Prasad and collaborators, including Akhilesh Prasad and Praveen K. Maurya (2015) [12], this calculus is adapted to the Hankel-Bessel setting in which the Fourier kernel is replaced by a product of Bessel functions. The resulting operators are translation-equivariant in the appropriate generalised sense and admit symbol-class boundedness theorems analogous to the Calderon-Vaillancourt theorem. From the perspective of medical artificial intelligence, this framework is constructive: it provides a principled space of integral-transform-based feature extractors that can be used either as fixed front-ends in a deep network or as the analytical limit of a learned convolutional layer in the small-kernel regime.

Proposition 2.15 (Composition stability for medical front-ends). Let A_σ be a Prasad-Kumar pseudodifferential operator with symbol σ in the standard symbol class $S^0_{1,0}$, and let $\Phi: \mathbb{R}^N \to \mathbb{R}^M$ be a deep neural network composed of L Lipschitz layers with respective Lipschitz constants $L_1, \dots, L_L$. Then the composite mapping $\Phi \circ A_\sigma: L^2(\mathbb{R}^d) \to \mathbb{R}^M$ is bounded, with

$$\| \Phi \circ A_\sigma(x) - \Phi \circ A_\sigma(x') \|_2 \leq \| A_\sigma \|_{L^2 \to L^2} \cdot \prod_{\ell=1}^{L} L_\ell \cdot \| x - x' \|_{L^2}. \tag{17}$$

Proof. The first factor follows from boundedness of A_σ on L^2 for symbols in $S^0_{1,0}$ [12, Section 2]. The product of layer-wise Lipschitz constants is the standard Lipschitz bound for a composition of Lipschitz maps. The composition is then immediate. This completes the proof.

The practical consequence is that adversarial robustness analyses of the deep network Φ, themselves an active area of medical-artificial-intelligence research [4], can be sharpened by an explicit estimate of $\| A_\sigma \|_{L^2 \to L^2}$ when the front-end is an integral-transform operator. This represents one concrete route by which classical operator theory contributes to the safety analysis of clinical artificial intelligence.

2.5 Discussion of Contributions C1 to C5

The results of this section are analytically of varying depth. Theorem 2.1 on regularity preservation and Theorem 2.7 on pointwise stability are direct consequences of Parseval-type identities; their value is the explicit pairing with medical-imaging and biosignal artificial intelligence and the explicit constants. Theorem 2.9 on the frame inequality is the more substantial analytical statement: it elevates the pointwise Lipschitz bound to a global norm-equivalence that justifies treating the discretised Bessel-wavelet coefficient vector as a feature map carrying, up to constants, the full information of the underlying signal. Theorem 2.5 on the denoising error bound yields a Pinsker-type rate for the inverse Hankel problem under additive Gaussian noise and is the analytical fact behind the engineering trade-off in radial magnetic resonance reconstruction. Theorem 2.11 on the Rademacher excess-risk bound gives the statistical-learning-theoretic counterpart of the frame inequality, providing an explicit generalisation guarantee at the rate $n^{-1/2}$ for clinical classifiers built on Bessel-wavelet features. Together with Proposition 2.15 on composition stability and Proposition 2.13 on translation covariance, these results constitute a self-contained analytical foundation for integral-transform-based feature extraction in clinical pipelines.

The sixth contribution (C6), the privacy-preservation result, is proved in Section 3, where it logically belongs. Taken together, contributions (C1) to (C6) support the central thesis of this paper: that the classical theory of integral transforms is an under-exploited and analytically robust resource for the next generation of safe, generalisable and privacy-aware medical artificial intelligence.

3. Integral-Transform Front-Ends and Differential Privacy

The integral-transform machinery of Section 2 interacts cleanly with the differential-privacy framework that underlies privacy-preserving medical artificial intelligence. Recall that a randomised training algorithm $\mathcal{A}$ is (ε, δ)-differentially private [1] if for every pair of datasets D and D' that differ in a single record and every measurable event S,

$$P(\mathcal{A}(D) \in S) \;\leq\; e^{\varepsilon}\, P(\mathcal{A}(D') \in S) + \delta. \qquad (18)$$

The canonical training procedure achieving (18) is differentially-private stochastic gradient descent [1], which clips per-sample gradients and adds Gaussian noise; moments-accountant composition tracks accumulated privacy loss. A recent systematic review of 74 studies on differential privacy for medical deep learning [10] reports that moderate privacy budgets of about $\varepsilon = 10$ retain clinically acceptable accuracy, while strict budgets of about $\varepsilon = 1$ typically cost 5 to 10 percentage points of the area under the receiver operating characteristic curve. Any inductive bias that can be inserted into the model without consuming privacy budget is therefore valuable. The following proposition shows that integral-transform front-ends are precisely such free inductive biases.

Proposition 3.1 (Differential privacy under integral-transform pre-processing). Let $\mathcal{T}: L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^{d'})$ be a fixed, data-independent linear operator, for example the Hankel transform (1), the Bessel-wavelet feature map (13), or any Prasad-Kumar pseudodifferential operator (16) with a deterministic symbol. Suppose $\mathcal{A}_{\mathrm{DP}}$ is a training procedure, such as differentially-private stochastic gradient descent [1], that is (ε, δ)-differentially private when applied to a dataset $D = \{x_i, y_i\}_{i=1}^n$. Then the composite procedure $\mathcal{A}_{\mathrm{DP}} \circ \mathcal{T}$ that first applies $\mathcal{T}$ to each input x_i and then runs $\mathcal{A}_{\mathrm{DP}}$ on the transformed dataset $\mathcal{T}(D) = \{\mathcal{T}(x_i), y_i\}_{i=1}^n$ is (ε, δ)-differentially private with respect to the original dataset D, with no additional privacy cost.

Proof. This is the post-processing and pre-processing invariance of differential privacy adapted to a data-independent linear front-end. Let D and D' be adjacent datasets, differing in one record, and let $\mathcal{T}(D)$ and $\mathcal{T}(D')$ be their images under the front-end. Since $\mathcal{T}$ is applied to each record independently and does not depend on the dataset, $\mathcal{T}(D)$ and $\mathcal{T}(D')$ are adjacent in the same sense, differing only at the record where D and D' disagree. By the assumed (ε, δ)-differential privacy of $\mathcal{A}_{\mathrm{DP}}$ applied to the now-adjacent datasets $\mathcal{T}(D)$ and $\mathcal{T}(D')$, for every measurable S,

$$P\big(\mathcal{A}_{\mathrm{DP}}\big(\mathcal{T}(D)\big) \in S\big) \le e^{\varepsilon}\, P\big(\mathcal{A}_{\mathrm{DP}}\big(\mathcal{T}(D')\big) \in S\big) + \delta.$$

But $\mathcal{A}_{\mathrm{DP}}\big(\mathcal{T}(D)\big) = (\mathcal{A}_{\mathrm{DP}} \circ \mathcal{T})(D)$, so the composite mechanism inherits the same (ε, δ)-differential-privacy guarantee with respect to D.

The argument requires only that $\mathcal{T}$ be a record-wise function, so that no information flows between records during the transform, and that it be data-independent, so that its definition does not depend on D. Both conditions hold for the integral-transform operators considered in this paper. Crucially, $\mathcal{T}$ need not be bijective: information may be discarded by the transform without affecting the privacy analysis. This completes the proof.

Corollary 3.2 (Composition with federated averaging). Under the same hypotheses as Proposition 3.1, if each client k in a federated learning protocol applies the same fixed transform $\mathcal{T}$ to its local data before participating in differentially-private stochastic gradient descent with per-client privacy budget $(\varepsilon_k, \delta_k)$, then the federated protocol with the integral-transform front-end satisfies the same per-client privacy guarantee $(\varepsilon_k, \delta_k)$, and standard composition theorems for federated differential privacy [5] apply unchanged at the server level.

Proof. Apply Proposition 3.1 to each client independently; the federated composition does not interact with the per-record front-end. This completes the proof.

Remark 3.3. The practical content of Proposition 3.1 is significant. It says that a hospital can deploy a Hankel-transform-based magnetic resonance reconstruction front-end, or a Bessel-wavelet electrocardiogram feature extractor, before any differentially-private training, and incur no additional privacy cost. The classical signal-processing structure is in this sense free from the privacy budget, a useful asymmetry because differentially-private stochastic gradient descent typically loses 5 to 10 percentage points of the area under the receiver operating characteristic curve on medical imaging tasks [10], and any inductive bias that can be inserted without privacy expenditure is therefore strictly desirable. To the knowledge of the author this asymmetry has not been stated explicitly in the medical-artificial-intelligence privacy literature, although the corresponding general fact about post-processing and pre-processing of differential privacy is folklore.

4. Discussion

This paper has argued that classical integral-transform theory occupies an under-exploited but structurally privileged position in medical artificial intelligence, and has supported the argument with six analytical results, a connecting lemma, and an illustrative numerical experiment. The implications are summarised below.

4.1 Three Implications

Integral transforms are an inductive bias. Hankel, Bessel-wavelet and Prasad-Kumar pseudodifferential operators are not merely tools for analysing data already in some learned representation; they are themselves representations that capture clinically meaningful structure, namely radial symmetry in magnetic resonance acquisition, multi-scale morphology in biosignals, and frequency localisation in ultrasound. Theorem 2.1 through Theorem 2.11 make precise the analytical guarantees these representations confer: regularity preservation under truncation, Lipschitz robustness, norm-equivalence at finite sampling, and parametric-rate generalisation for downstream classifiers. Designed integral-transform front-ends can therefore stand alongside data-driven feature learning as a principled alternative.

Mathematical clarity is the route to safety. The denoising bound of Theorem 2.5 makes the bias-variance trade-off in radial magnetic resonance imaging explicit, with an optimal cutoff $K^* \asymp (B^2/\sigma^2)^{1/(4s+1)}$ that can be computed a priori from properties of the acquisition. The frame inequality of Theorem 2.9 provides a quantitative reconstruction guarantee for biosignal pipelines. The Rademacher bound of Theorem 2.11 gives an explicit minimum-sample-size formula for any given target generalisation gap. Each of these is a clinically actionable quantity, not merely an existence statement.

Governance is mathematical. Proposition 3.1 shows that integral-transform front-ends compose cleanly with differentially-private stochastic gradient descent: a hospital can deploy a Hankel-based magnetic resonance reconstruction or a Bessel-wavelet electrocardiogram feature extractor before any differentially-private training, and incur no additional privacy cost. The 2026 systematic review in npj Digital Medicine [10] makes clear that the privacy-utility frontier is now well mapped at moderate budgets, while strict-privacy regimes remain a genuine research challenge. Any inductive bias that is free with respect to the privacy budget is strictly desirable.

4.2 Closing Remark

The figures and worked examples in this paper, namely the Hankel reconstruction of a radial phantom, the Bessel-wavelet decomposition of a synthetic electrocardiogram and the numerical experiment illustrating Theorem 2.11, are not illustrations of a finished theory. They are entry points into a research programme that is best pursued by readers comfortable both with the language of clinical medicine and with the language of integral transforms, pseudodifferential operators and functional analysis.

5. Conclusion

A compact analytical toolkit has been assembled for integral-transform front-ends in clinical machine learning, consisting of a truncation-error estimate in the Hankel domain, a Pinsker-type denoising rate, a Lipschitz stability estimate and a discrete frame inequality for Bessel-wavelet features, an excess-risk bound for linear readouts on those features, and an invariance result for differentially-private training. Each statement is accompanied by constants that can be computed or estimated before any patient data are touched, which is what makes the toolkit usable in the design phase of a clinical pipeline rather than only in retrospective analysis. The natural continuation of this work is threefold: replication of the numerical illustration on credentialed clinical repositories; the design of reconstruction architectures that operate intrinsically in the Hankel domain rather than on gridded Cartesian data; and a sharpening of the excess-risk constants through per-scale normalisation of the Bessel-wavelet feature map.

Declarations

Funding. No funding was received to assist with the preparation of this manuscript.

Ethics approval. Not applicable. This study did not involve human participants, human data, or animals.

Data availability. No new clinical data were collected for this study. All numerical illustrations use synthetic data generated by the script that accompanies this manuscript, or quote published benchmark figures from the cited primary sources. The Python script that produced Table 1 and Figure 3 is available from the author on reasonable request and reproduces all reported numbers from a fixed random seed.

Code availability. The Python code used to generate the numerical illustration in Section 2.3 is available from the author on reasonable request.

Author contributions. The paper is the work of a single author, who is responsible for the conception, the mathematical development, the numerical experiments and the preparation of the manuscript.

Conflict of Interest

The author declares that he has no competing financial or non-financial interests relevant to the content of this article. No sponsor or other organisation influenced the design, execution, analysis or reporting of this research.

Acknowledgement

The author thanks the management of MAEER’s MIT Arts, Commerce and Science College, Alandi (D), Pune, for the research environment and institutional support that made this work possible.

Author’s Biography

Dr. Balasaheb B. Waphare is Director and Professor of Mathematics at MAEER’s MIT Arts, Commerce and Science College, Alandi (D), Pune, an autonomous institution affiliated to Savitribai Phule Pune University. His current research interests are integral transforms, special functions, pseudo-differential operators, fractional calculus and wavelet theory.

References

1. M. Abadi, A. Chu, I. Goodfellow, H.B. McMahan, I. Mironov, K. Talwar, L. Zhang, “Deep Learning with Differential Privacy”, Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Association for Computing Machinery, New York, 2016, 308-318. https://doi.org/10.1145/2976749.2978318
2. P.L. Bartlett, D.J. Foster, M.J. Telgarsky, “Spectrally-Normalized Margin Bounds for Neural Networks”, Advances in Neural Information Processing Systems, Curran Associates, Red Hook, 2017, 30, 6240-6249.
3. C. Boeck, “ECG Signal Analysis Based on the Wavelet Transform”, Master’s Thesis, Signal Processing and Speech Communication Laboratory, Johannes Kepler University Linz, Linz, 2010.
4. S.G. Finlayson, J.D. Bowers, J. Ito, J.L. Zittrain, A.L. Beam, I.S. Kohane, “Adversarial Attacks on Medical Machine Learning”, Science, 2019, 363 (6433), 1287-1289. https://doi.org/10.1126/science.aaw4399
5. R.C. Geyer, T. Klein, M. Nabi, “Differentially Private Federated Learning: A Client-Level Perspective”, NeurIPS Workshop on Privacy and Machine Learning, 2017. https://arxiv.org/abs/1712.07557
6. W.E. Higgins, D.C. Munson, “A Hankel Transform Approach to Tomographic Image Reconstruction”, IEEE Transactions on Medical Imaging, 1988, 7 (1), 59-72. https://doi.org/10.1109/42.3929
7. V. van Veldhuizen, V. Botha, C. Lu, M.E. Cesur, K. Groot Lipman, E.D. de Jong, H. Horlings, C.I. Sanchez, C.G.M. Snoek, L. Wessels, R. Mann, E. Marcus, J. Teuwen, “Foundation Models in Medical Imaging: A Review and Outlook”, Preprint, Version 4, November 2025. https://arxiv.org/abs/2506.09095
8. F. Rastegar Jooybari, A. Aghaeifar, E. Mohammadi, K. Scheffler, A. Nasiraei-Moghaddam, “Polar Fourier Transform in Practice: Its Efficiency and Characteristics in Reconstructing Radially Acquired MRI Images”, Magnetic Resonance Materials in Physics, Biology and Medicine, 2026, 39 (1), 21-36. https://doi.org/10.1007/s10334-025-01284-w
9. M.J. Muckley, B. Riemenschneider, A. Radmanesh, et al., “Results of the 2020 fastMRI Challenge for Machine Learning MR Image Reconstruction”, IEEE Transactions on Medical Imaging, 2021, 40 (9), 2306-2317. https://doi.org/10.1109/TMI.2021.3075856
10. M. Mohammadi, M. Vejdanihemmat, M. Lotfinia, M. Rusu, D. Truhn, A. Maier, S. Tayebi Arasteh, “Differential Privacy for Medical Deep Learning: Methods, Tradeoffs, and Deployment Implications”, npj Digital Medicine, January 2026. https://doi.org/10.1038/s41746-025-02280-z
11. S. Pati, U. Baid, B. Edwards, et al., “Federated Learning Enables Big Data for Rare Cancer Boundary Detection”, Nature Communications, 2022, 13, 7346. https://doi.org/10.1038/s41467-022-33407-5
12. A. Prasad, P.K. Maurya, “A Pair of Pseudo-Differential Operators Involving Hankel-Type Integral Transformations”, Journal of Pseudo-Differential Operators and Applications, 2015, 6 (1), 135-151. https://doi.org/10.1007/s11868-014-0104-6
13. S. Shalev-Shwartz, S. Ben-David, “Understanding Machine Learning: From Theory to Algorithms”, Cambridge University Press, Cambridge, 2014. https://doi.org/10.1017/CBO9781107298019

14. R. Sivakumar, B. Lue, S. Kundu, “FDA Approval of Artificial Intelligence and Machine Learning Devices in Radiology: A Systematic Review”, JAMA Network Open, 2025, 8 (11), e2542338. https://doi.org/10.1001/jamanetworkopen.2025.42338
15. E. Tiu, E. Talius, P. Patel, C.P. Langlotz, A.Y. Ng, P. Rajpurkar, “Expert-Level Detection of Pathologies from Unannotated Chest X-Ray Images via Self-Supervised Learning”, Nature Biomedical Engineering, 2022, 6, 1399-1406. https://doi.org/10.1038/s41551-022-00936-9
16. E. Vorontsov, A. Bozkurt, A. Casson, et al., “A Foundation Model for Clinical-Grade Computational Pathology and Rare Cancers Detection”, Nature Medicine, 2024, 30 (10), 2924-2935. https://doi.org/10.1038/s41591-024-03141-0
17. G.N. Watson, “A Treatise on the Theory of Bessel Functions”, 2nd Edition, Cambridge University Press, Cambridge, 1944.