

CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES

https://cajmtcs.casjournal.org/index.php/CAJMTCS

Volume: 07 Issue: 01 | January 2026 ISSN: 2660-5309

Article

Data-Driven Models for Approximating Solutions to Friedholm's Type II Integrative Equations: A New Hybrid Framework with Comprehensive Theoretical Analysis

Susan Ridha Shakor[1], Israa Mustafa Omer[2]

1,2. Department of Mathematics, College of Education for Pure Science, University of Kirkuk, Iraq
Correspondence: suzanridha@uokirkuk.edu.iq[1], israamustafa@uokirkuk.edu.iq[2]

Abstract: This research paper presents an original framework called:" Integrally Informed Neural Network (IINN)", specially designed to solve Friedholm's integral equations of type II including linear and nonlinear states, integrative systems, and the uniquely weak nuclei. The framework is based on the fundamental architectural feature of integrating the Friedholm effect directly into the neural network architecture by an structural attention mechanism coupled with a wavelength embedding layer based on B-spline functions, rather than simply representing it within the error function as in traditional physically informed neural network models, allowing the network to implicitly learn the inverse effect $(I-\lambda K)^{-1}$ and immediate circulation to any new coercive function without regrouping. Theoretical analysis is based on three complementary pillars: Banach's fixed point theory to ensure the existence and unity of the solution, the spectral theory of monolithic effects for convergence analysis, and Barron's comprehensive approximation theory for deriving rigid error limits dismantled into three independent sources that are error Rounding, squaring error, and optimization error, as well as proof of stability under noise. The study conducts a standardized comparison between IINN and seven classical reference methods including B-spline waves, Har waves, heterogeneous repetition, Taylor expansion, plus mass functions, Adomian analysis, and homotopic disorder, as well as the modern effects learning models DeepONet and the Fourier neural effect. Numerical results across four categories of problems with real-world geometric applications verify an improvement in accuracy by one to two orders compared to the best competing methods, with higher stability Under Noise is supported by statistical analysis with *Student t-tests* and *confidence ranges*.

Keywords: Friedholm's Integrative Equations, Integrally Informed Neural Network, Banach's Constant Point Theory, Effect Learning, B-Spline Waves, Spectroscopy, Stability Under Noise

Citation: Shakor, S. R., Omer, I. M. Data-Driven Models for Approximating Solutions to Friedholm's Type II Integrative Equations: A New Hybrid Framework with Comprehensive Theoretical Analysis. Central Asian Journal of Mathematical Theory and Computer Sciences 2026, 7(1), 375-391.

Received: 23rd Nov 2025
Revised: 15th Dec 2025
Accepted: 10th Jan 2026
Published: 30th Jan 2026

Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/)

1. Introduction

The General Context and the Importance of the Issue

Friedholm's integral equations of the second type occupy a central place in the mathematical modeling of physical, geometrical, and biological phenomena, and their general formula is given in the form$y(x) = f(x) + \lambda \int_a^b K(x,t)y(t)dt, x \in [a,b]$, of K(x,t) represents the nucleus of the integration that describes the nature of the non-local correlation between the points of the field, and f(x) represents a known forced function, while the coefficient λ determinesThese equations arise in a wide range of scientific applications, including thermal radiation transfer in the scattered absorption medium,

inverse models in medical imaging, modeling of biological systems, stress analysis in composite materials, and some formulations in quantum mechanics. The main challenge is to achieve a delicate balance between numerical accuracy, computational efficiency, and generalizability, which makes this type of equation one of the most sensitive issues in modern numerical analysis [1].

Diagnosing the Research Gap

The numerical methods for solving Friedholm's type II equations are divided into three main approaches: starting with functional expansion methods such as B-spline waves, Harr waves, Legender waves, and Taylor expansion, which reduce the integrative equation to a finite algebraic system, but their accuracy deteriorates when dealing with irregular or weakly unique nuclei, and their computational cost rises sharply with high precision requirements [2], and then iterative and sequential methods such as heterogeneous recursion method, Adomian analysis, and homotopy disorder method that construct approximate sequential solutions without the need for network assembly, but whose convergence depends on regularity conditions that cannot be guaranteed in extreme nonlinear equations as well as the intrinsic locality of Taylor's expansion that deteriorates significantly away from the expansion point [3], Finally, deep learning and effect learning models such as Physically Informed Neural Networks, DeepONet, and Fourier Neural Effect, which represent a qualitative leap towards learning solution functions, but PINNs treat the integrative effect as a constraint in the error function rather than as an architectural element, while DeepONet and FNO learn the effect from data alone without investing in the known algebraic structure, which necessitates greater training data and provides weaker theoretical guarantees [4] This review reveals three specific gaps: the absence of integrative effector integration within the model's architecture, the absence of a rigorous mathematical framework linking data-driven models to functional effects theory and monolithic effect spectroscopy, and the lack of standardized normative assessments with accurate stability analysis under noise [5].

Original Contributions to this Research

This paper introduces the IINN framework that integrates the Friedholm effect into the neural network architecture via an integrative attention mechanism coupled with the inclusion of a Hilbert-Schmidt justified wave in order to enable the inverse effect $(I-\lambda K)^{-1}$ to be approximated in the forward computation of the network and the immediate generalization of any new coercive function without retraining. noise and a spectral result showing the convergence of attention weights towards the spectral nucleus of the Friedholm effect [6], as well as a standardized methodological comparison that includes seven reference methods and two modern methods of learning effects with accurate statistical analysis with Studentt's t-tests and investigations on real geometric applications that reveal an improvement in accuracy by one to two degrees compared to the best competitors, confirming that the integration of the mathematical structure of the integrative effect into the neural network and not in the error function alone is the pathway The most effective and promising solution to this class of equations in scientific and engineering applications [7].

2. Theoretical Framework

Hilbert Spaces Theory and Monolithic Effects

The Hilbert space $\mathbf{L^2[a,b]}$ **is** the natural framework for the study of Friedholm's equations, as it is complete with respect to the internal multiplication and the isotope derived from it, to which the nucleus, the function and the solution belong under normal conditions, and the Friedholm$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \int \boldsymbol{f}(\boldsymbol{x})\boldsymbol{g}(\boldsymbol{x})\boldsymbol{dx} \|\boldsymbol{f}\| = \sqrt{\langle \boldsymbol{f}, \boldsymbol{f} \rangle} \boldsymbol{Kf}\ \boldsymbol{y}$ ∗**effect K** : $\boldsymbol{L}^2[\boldsymbol{a}, \boldsymbol{b}] \rightarrow \boldsymbol{L}^2[\boldsymbol{a}, \boldsymbol{b}]$ is defined as the relationship that describes the non-local correlation between the points of the field, and the stacking theory proves that if it is, then the effect is compact on by proving that it is an isotope limit of a sequence of orders of effects

terminated through the expansion of the nucleus$(\mathcal{K}\boldsymbol{y})(\boldsymbol{x}) = \int \boldsymbol{K}(\boldsymbol{x},\boldsymbol{t})\boldsymbol{y}(\boldsymbol{t})\boldsymbol{dt}\ \boldsymbol{K} \in \boldsymbol{L}^2([\boldsymbol{a},\boldsymbol{b}]\times[\boldsymbol{a},\boldsymbol{b}])\mathcal{K}\ \boldsymbol{L}^2[\boldsymbol{a},\boldsymbol{b}]$

$$K(x,t) = \Sigma\ k_mn\ e_m(x)e_n(t)$$

The use of the Cauchy–Schwartz oscillation to verify the convergence of this sequence, and the clustering of K has the intrinsic result that the spectrum σ(K) is either terminated or converged towards zero, ensuring that there is a single solution to the equation when it is not an intrinsic value of the effect, and this result is not abstract but directly justifies the spectral inclusion strategy inλ the IINN, where the solution y* belongs to the space generated by the eigen functions ofK that are efficiently approximated by bases B-spline [8].

Banach's Fixed Point Theory and its Application

Let (X,d) be a complete metric space, and let T:X→X be a contraction application with a constant, i.e.$0 \le \kappa < 1$

$$d(Tx,Ty) \le \kappa d(x,y)\ \text{لكل}\ x,y \in X.$$

It is a single element so that$x* \in X$

$$Tx* = x*$$
$$x_{n+1} = Tx_n$$

It converges to x∗ starting at any starting point, and at a linear convergence rate that depends on the constant $x0 \in X\kappa$.

Spectroscopy of the Friedholm effect

If the kernel K(x,t) is the same:

$$K(x,t) = K(t,x),$$

The integrative effect K on is self-accompanying and compressed, thus $L^2[a,b]$:

Subjective values areμ_n real,

eigen functions { } are ϕ_n perpendicular and form a complete rule in $L^2[a,b]$

Spectral expansion of the nucleus

$$K(x,t) = \sum_{n=1}^{\infty} m_n\ \varphi_n(x)\varphi_n(t).$$

Solution of the Integral Equation

$$y(x) - \lambda \int_a^b K(x,t)y(t)dt = f(x)$$

Gives by:

$$y^n(x) = \sum_{n=1}^{\infty} \frac{\langle f,\phi_n\rangle}{1-\lambda\mu_n}\phi n(x), \text{بشرط}\ \lambda\mu_n \neq 1.$$

Spectral radius:

$$r(k) = lim_{n\to\infty}\ \| Kn \|^{\frac{1}{n}} = sup_n\ |\ \mu_n\ |$$

Conditions of Existence and Oneness:

$$|\ \lambda\ | < \frac{1}{r(K)}.$$

Spectral Interpretation (IINN)

The inclusion layer approximates the eigen functions $\{\}\phi_n$ while the attention mechanism converges the weights so that only the solution condition provides a more accurate and comprehensive criterion than the usual classical conditions in the numerical analysis of integrative equations$\frac{1}{1-\lambda\mu_n}$ [9].

a. ***Universal approximation theory***

Hornik's comprehensive approximation theory asserts that multilayer neural networks (MLPs) with a non-polynomial activation function are able to approximate any continuous function on a compressed field with arbitrary precision. This is an existential consequence: it guarantees the possibility of approximation but does not give the error rate.

b. ***Barron 's theorem***

It adds an important quantitative dimension: if a function has an "ordered" Fourier representation (i.e., its spectrum is finite in a certain way), the approximation error

decreases by the O(1/N) rate where the number of nodes is NNN. Importantly, this rate is not dependent on dimension, which distinguishes neural networks from grid-based, which suffers from the dimensional curse.

This makes this theory a direct mathematical basis for approximation error analysis in models such as IINN [7].

c. ***Hornick's theorem (Universal Approximation):***

$$\forall f \in C([a,b]),\ \forall \varepsilon > 0,\ \exists\ N\Theta\ such\ that\ \sup_{x\in[a,b]} f(x) - N_\Theta(x) < \varepsilon.$$

d. ***Fourier condition***:

$$C_f = \int_{R^d} |\ \boldsymbol{\omega}\ ||\ \hat{\boldsymbol{f}}(\boldsymbol{\omega})\ |\ \boldsymbol{d\omega} < \infty$$

e. ***Barron's error limit:***

It adds an important quantitative dimension, as if the function has an organized Fourier representation, the approximation error decreases at the rate of O(1/N) where N is the number of nodes. Importantly, this rate is not dependent on dimension, which distinguishes neural networks from grid-based methods that suffer from the dimensional curse, making this theory a direct mathematical basis for analyzing the approximation error in models such as IINN [7].

$$\int_{Rd} (f(x) - N_\Theta(x))^2 dx \le \frac{C_f^2}{N}.$$

3. Formulation of the Issue and Definitions

Studied Equations

In this paper, four types of Friedholm equations of type II are studied over the period [a,b] [8]:

a. Linear Equation:

$$y(x) = f(x) + \lambda \int_a^b K(x,t)y(t)dt.$$

b. Hammerstein's nonlinear equation:

$$y(x) = f(x) + \int_{\boldsymbol{a}}^{\boldsymbol{b}} K(x,t)F(t,y(t))dt,$$

where:

$$F\colon [a,b] \times R \to R$$

Non-linear.

c. A system of linear integral equations:

$$y_i(x) = f_i(x) + \sum_{j=1}^{n} \int_a^b abKi,j(x,t)\ yj(t)\ dt, i = 1,2,\dots,n.$$

d. Equation with a weakly unique nucleus:

$$y(x) = f(x) + \int_a^b \frac{K(x,t)}{|\ x-t\ |^a} y(t)dt, 0 < \alpha < 1.$$

Mathematical Hypotheses

Maintain the following mathematical conditions throughout the paper to ensure the existence and unity of the solution:

- **(F1)** The kernel K is quadratic integrated:

$$K \in L^2([a,b] \times [a,b])$$

- **(F2)** The quadratic integral force function f is:

$$f \in L^2[a,b]$$

- **(F3)The coefficient** of λ **lies outside the spectrum**:

$|\lambda| < \frac{1)}{r(K}$

where r(K) is the radius of the Friedholm effector spectrum.

- **(F4)** For nonlinear state only — operator F continuous Lipschitz:

$$|\ \lambda\ |\cdot L_F \cdot \|\ K\ \| < 1|$$

where the Lipschitz constant L_F operator F with the second variable.

Under these conditions combined it is guaranteed that there is a single solution under the Friedholmeist alternative and Banach's fixed point theory $L^2[a,b]y^* \in$.

Standardized Evaluation Measures

Four standardized measures were adopted to evaluate the performance of comparative methods:

- The first scale: in the mean of the error square.MSE $= \frac{1}{N}\sum_{i=1}^{N}(y^*(x_i)) - \tilde{y}(x_i))^2$ ،
- The second criterion: in the maximum error analogue $\|E\|_\infty = \max_{1\le i\le N}|y^*(x_i) - \tilde{y}(x_i)|$ ،
- The third measure of the error analogue in space $\|E\|\{L^2\} = (\int_a^b |y^*(x) - \tilde{y}(x)|^2 dx)^2$
- Fourth Scale: Convergence Rate Calculated by Formula $Rate = -log_2(\frac{\|E_h\|}{\|E_{2h}\}\|})$

It allows the efficiency of each method to be measured when the network is condensed, and the selection of these metrics together provides a comprehensive and balanced picture of the performance of each method in terms of overall accuracy, local behavior, and rate of improvement with the network [7].

4. Classical numerical methods

The B-Spline Wave Method

a. Mathematical Structure

The B-Spline method is based on approximating the unknown function y(x) using a linear combination of local and soft base functions with a built-in support, which allows the integral equation to be converted into a finite algebraic system. These functions are highly capable of representing regular solutions with high accuracy while maintaining a semi-sparse matrix structure that relatively reduces the computational cost.

For regular node sequence:

$$0 = x_{-m+1} = \cdots = x_0 < x_1 < \cdots < x_n = 1,$$

M-order B-Spline functions are recursively known as the relation:

$$B_{m,j}(x) = \frac{x - x_j}{x_{j+m-1} - x_j} B_{m-1,j}(x) + \frac{x_{j+m} - x}{x_{j+m} - x_{j+1}} B_{m-1,j+1}(x),$$

With starting base:

$$B_{1,j}(x) = 1_{[x_j, x_{j+1})}(x)$$

These functions achieve the property of a built-in support:

$$supp(B_{m,j}) = [x_j, x_{j+m}],$$

It is an essential property that leads to the disintegration of the algebraic structure of the resulting system [2].

b. Fragmentation and the algebraic system

The solution is approximated to the following form:

$$y(x) \approx C\hat{}T\,\Psi(x)$$

Then, by substituting in the integral equation, we get the system:

$$C^T(p - \Theta) = C_1^T$$

where:

$$P = \int_a^b \Psi(x)\Psi^T(x)dx,$$

And: $\Theta_{i.j} = \int_a^b \int_a^b K(x,t)\,\Psi_i(t)\Psi_j(x)dtdx.$

The calculation cost is:

$$O(4^m)$$

To build the system, and:

$$O(8^m)$$

To solve it, where M level is shredding. This indicates a sharp increase in cost with an increase in the required accuracy, and it should be noted that this system is disintegrated thanks to the compact prop property, which relatively mitigates the burden of this computational escalation [10].

c. Theoretical Convergence Analysis

If the exact solution:

$$y^* \in H^s[a,b],$$

The B-spline rank was achieved:

$$m \geq s$$

The error will be checked:

$$\| y^* - yM \| L^2 \leq C2^{-Mmin(s,m)} \| y^* \| H^s.$$

That is:

$$\| y^* - yM \| = O(h^{min(s,m)}), h = 2 - M.$$

This suggests that the method achieves high convergence rates for regular solutions, but its efficiency decreases in the case of poorly synthesized nuclei or solutions with irregular behavior.

Har Wave Method

a. Mathematical Idea

Har waves are one of the simplest and most efficient wave rules, as they rely on fixed deterministic functions that allow the representation of functions and the transformation of integrals into simple algebraic operations. This method is easy to apply and fast to calculate, but it suffers from limited accuracy due to the lack of continuity of base functions.

The parent function is defined:

$$h_1(x) = \begin{cases} 1, 0 \leq x < \frac{1}{2}, \\ -1, \frac{1}{2} \leq x < 1 \\ 0, otherwise \end{cases}$$

The rest of the functions are generated by extension and displacement:

$$h_n(x) = h_1(2^j x - k), n = 2^j + k$$

b. Orthogonality and Operational Matrix

Har Equations are characterized by the property of orthogonality:

$$\int_0^1 h_m(x) h_n(x) dx = 2^{-j}\, \delta mn,$$

This allows transactions to be calculated with high efficiency.

The Operational Matrix of Integration is also known as the relationship:

$$\int_0^x h_{(m)}(t) dt = P_{(M)} H_{(M)}(X),$$

where:

$$P(m) = \frac{1}{2m} \begin{bmatrix} 2mP\left(\frac{m}{2}\right) - H_{\frac{M}{2}} \\ H_{\frac{M}{2}}^{-1}\ 0 \end{bmatrix}$$

The computational complexity is as follows:

$$O(m^2).$$

The fundamental limitation is that the inconsistency of Har functions restricts their convergence rate at O(h) i.e. only one rank, which explains their reduced accuracy compared to B-spline when high accuracy is required [11].

The Method of Heterogeneous Repetition

Iteration method is based on converting the Friedholm integral equation into a partial differential by differentiation for the independent variable, with the aim of facilitating the construction of an iterative scheme of the solution. Next, what is known as a heterogeneous correction function is created, which is used to generate a sequence of approximations for the real solution. This function is based on adding a calculated correction to the current estimate of the solution, so that this correction reduces the error caused by the failure to achieve the original equation. The optimal Lagrangian multiplier is determined using Euler–Lagrange conditions, often This significantly simplifies the iterative formula. Iteration usually starts with a simple initial guess, such as the given function on the right-

hand side, and is gradually improved over successive iterations. The convergence of this method depends on the fulfillment of a particular compression condition associated with the Friedholm effect, a condition that corresponds to the basic hypotheses of Banach's fixed point theory, reflecting a deep correlation between functional analysis and iterative methods. Although this method is effective and relatively simple, its performance may be affected by the nature of the nucleus and the degree of nonlinearity of the equation, especially in cases where the convergence condition is not strictly met [12].

The heterogeneous recursion method is based on taking the derivative with respect tox in Friedholm's equation to get

$$y(x) = f(x) + \int_a^b K(x,t)\ y(t)dt$$

By deriving both sides with respect to x assuming the derivability of K

$$y'(x) = f'(x) + \lambda \int_a^b \frac{\partial K(x,t)}{\partial x}\ y(t)dt$$

Then build the heterogeneous correction function

$$y_{n+1}(x) = y_n(x) + \int_a^x \lambda(\xi)\,[y'_n(\xi) - f'(\xi) - \lambda \int_a^b K_x(\xi,t)\,y_n(t)dt]d\xi$$

where

$$\frac{\partial K(\xi,t)}{\partial \xi}$$

After determining the Lagrangian multiplier (mostly)$\lambda(\xi) = -1$

$$y_{n+1}(x) = y_n(x) - \int_a^x [y'(\xi) - f'(\xi) - \lambda \int_a^b K_x(\xi,t)\,y_n(t)dt]d\xi$$

Initial Guess:

$$y_0(x) = f(x)$$

Sufficient Affinity Condition:

$$|\lambda|\|\mathcal{K}\| < 1$$

The Method of Expanding the Taylor Sequence

This method is based on expanding around$y_j(t)$ x up to order m in the formula

$$y_j(t) = \sum_{r=0}^{m} \frac{[(t-x)^r}{r!}\ y_j^r(x) + R_m(t)$$

by compensation in the system after neglecting the limit of the rest to get R_m

$$y_i(x) - \sum_{j=1}^{n} \sum_{r=0}^{m} \frac{1}{r!} y_j^{(r)}(x) \int_a^b K_{ij}(x,t)(t-x)^r\,dt \approx f_i(x)$$

Error Estimation

$$\|R_m\|\infty \leq \frac{(b-a)^{m+1}}{(m+1)!}\|y^{(m+1)}\|\infty$$

It proves a rapid convergence of smooth solutions, but the fundamental limitation is the intrinsic localization of the method as the accuracy deteriorates significantly away from the point of expansion, making it unsuitable for capturing the non-local behavior that characterizes Friedholm's solutions and forms the core of its computational challenge [13].

5. Data-driven models: from PINNs to IINN

Modern Effects Learning Models

a. DeepONet

DeepONet is based on the idea of a nonlinear effect learning that connects the function space to itself, so that the resulting function is not rounded directly, but rather learns a general transformation rule from the input function to the solution. This is done by separating the representation into two parts: a branch grid to extract the representation of the input function at specific points, and a trunk network representing the dependence on

the xxx spatial variable$L^2[a,b]$. This chapter allows the model to generalize the affective relationship to different families of functions without retraining [14].

$$G(f)(x) \approx \sum_{k=1}^{p} b_K(f(x_1), \dots, f(x_M))t_K(x) \ .$$

b. Fourier Neurotransmitter Effector (FNO)

FNO operates in frequency space rather than spatial space, where the state is converted to a Fourier field, and then a learned filter is applied to the frequency representation before returning to the original domain. This structure enables the model to capture non-local interactions with high efficiency, which makes it particularly suitable for integrative equations with spectral behavior [14].

$$v_{l+1}(x) = \sigma(W_l v_l(x) + F^{-1}(R_l \cdot F(vl))(x)).$$

Proposed Framework IINN

a. Motivation and Foundational Vision

The proposed IINN framework reformulates the solution problem as an inverse effect learning that relates the input function to the solution, rather than learning the solution directly. Since the solution of Friedholm's equations could be written as an inverse effector, embedding this structure within the architecture reduces the search space and enhances generalizability compared to models that rely solely on error function reduction [15].

$$y^* = (1 - \lambda K)^{-1} f$$

b. Detailed Architecture

IINN consists of four integrated components.

The first component: the wavelength layer

IINN is based on a multi-precision representation that aims to approximate the spectral structure of a solution via wavebases capable of representing local and global behavior at the same time. This representation connects spectroscopy and numerical analysis, providing the model with a rich representation prior to the learning process [16].

$$e(x) = [\varphi_{j0,K}(x)]_K \cup [\psi_{j,k}(x)]_{j,k}.$$

Component Two: The Mechanism of Integrative Attention

This layer simulates the integrative effect by integrating the mechanism of neural attention with numerical integration using Gauss–Legender square, making attention a direct numerical representation of the integrative effector structure rather than just a statistical correlation between points. [17]

$$IntAttn(x) = \sum_{k=1}^{N_q} w_K A(x, t_k) h(t_K), A(x,t) = softmax(\frac{Q(x)K(t)^T}{\sqrt{d_k}}).$$

The third component: the remaining ResNet block

This block is based on residual learning, where incremental corrections are learned on the current representation rather than rebuilding the solution entirely. This design improves the stability of the training and increases the efficiency of the affinity, especially when the influencer is close to the identity or when the coupling force is weak [18].

$$h^{(l+1)} = h^{(l)} + \sigma(W_2^{(l)} \sigma(W_1^{(l)} h^{(l)} + b_1^{(l)}) + b_2^{(l)}).$$

Fourth component: Output layer with explicit correction

This layer reconstructs the final solution by adding the given function f(x)f(x)f(x) directly to the grid output, reflecting the basic structure of Friedholm's equations, and making the learning task limited to estimating only the effect of the integrative effector and not reconstructing the solution entirely [4].

$$\hat{y}(x) = W_{out} h^{(L)} + b_{out} + f(x).$$

a. Training Algorithm

The IINN training algorithm is based on integrating the spectral structure of the effector into the forward scroll, so that the solution is not learned directly, but the remainder of the integrative equation is reduced. The process begins with the construction of a multi-precision waverepresentation that serves as the initial input to the network,

providing the model with a rich representation that reflects the spectral structure of the effector from the outset [15].

$$e(x_i) = [\varphi_{j0,k}(x_i)]_k \cup [\psi_{j,k}(x_i)]_{j,k}.$$

The network parameters are then initialized using Xavier initialization to ensure training stability and prevent gradient fading or exploding, which is standard practice in deep network training [19].

$$\Theta = \{W_A, W_Q, W_K, W_1^{(l)}, W_2^{(l)}, W_{out}\}.$$

During forward scrolling, the wavelength is first projected into an appropriate representative space, and then the data is passed through residual layers that include the integrative attention mechanism that approximates the K effect using numerical squared. This combination of attention and integration enables the network to simulate the non-local structure of the effector directly within the computation [14].

$$c(x) = \sum_{K=1}^{N_Q} w_K A(x, t_K) h(t_K).$$

Representation is updated via ResNet blocks, where incremental corrections are learned rather than completely rebuilding the solution, which improves numerical stability and increases convergence efficiency [18].

$$h^{(l+1)} = h^{(l)} + \sigma(W_2^{(l)} \sigma(W_1^{(l)}[h^{(l)}; c(x)] + b_1^{(l)}) + b_1^{(l)}).$$

In the output layer, tribal knowledge is integrated by directly adding f(x)f(x)f(x), reducing the learning burden and shifting the task to estimating the impact of the influencer only [4].

$$\hat{y}(x) = W_{out} h^{(L)} + b_{out} + f(x).$$

The model is trained by reducing the residual integrative equation without the need for precise solution data, which is similar toPINNs but differs in that the effector here is built into the architecture [20].

$$L_{eq} = \frac{1}{N_r} = \left| \sum_{i=1}^{N_r} \hat{y}(x_i) - f(x_i) - \lambda \sum_{K=1}^{N_Q} w_k K(x_I, t_K) \hat{y}(tk) . \right|^2$$

Transactions are optimized using the Adam algorithm followed by L-BFGS for faster convergence and higher accuracy, while adopting an early stop standard based on transaction stability [21].

$$\Theta^{(k+1)} = \Theta^{(k)} - \eta \nabla_\Theta L_{total}.$$

b. Comparison with well-known models

The comparison shows that IINN features the integration of the influencer directly within the architecture, unlike PINNs that rely on the error function only, andDeepONet and FNO that learn the influencer from the data without explicitly including its mathematical structure. This integration enhances theoretical interpretation and improves generalization [15].

$$y^* = (I - \lambda K)^{-1} f.$$

The use of wavelength also gives IINN a spectral representation justified by the Hilbert-Schmidt theory, while other models rely on unconstrained implicit representations [16].

$$K(x, t) = \sum_N \mu_n \phi_n(x) \phi_n(t).$$

In addition, integrating in-network integration via the attention mechanism enables IINN to handle complex cores, including weakly unique cores, naturally without the need for external processing [14].

$$\int_a^b K(x, t) y(t) dt \approx \sum_{k=1}^{N_q} w_k K(x, t_K) y(t_K).$$

Finally, IINN is distinguished by its ability to instantly generalize to new functions f, similar to DeepONet, while maintaining a physical structure built into the model, making it combine the advantages of data-driven learning with rigorous mathematical analysis [4].

$$\hat{y}(x; f_{new}) \to y^*(x).$$

6. Rigorous Theoretical Analysis ofIINN

The Theory of Error Removal

Let the only solution to the Friedholm equation of type II be under the standard hypotheses $y*$ (F1)– (F1), and let the output of$\hat{y}\Theta$ the IINN model be after training. We define the three sources of error as follows:

$$\varepsilon_{approx}(N_h, L) = \frac{inf}{\Theta} \parallel y^* - \hat{y}\Theta \parallel L^2,$$

$$\varepsilon_{quad}(N_q) = sup_{x\in[a,b]} \int_a^b K(x,t)\hat{y}\Theta(t)dt - \sum_{k=1}^{N_q} w_k K(x,t_k)\hat{y}\Theta(t_k)\,,$$

$$\varepsilon_{opt}(\Theta^*, \Theta_{trained}) = \parallel \Theta^* - \Theta_{trained} \parallel F.$$

The following estimate of the total error is then achieved:

$$\parallel y^* - \tilde{y}\Theta_{trained} \parallel L^2 \le \frac{1}{1-\mid \lambda \mid\mid K \parallel}(\varepsilon_{approx} + \mid \lambda \mid (b-a)^{\frac{1}{2}}\varepsilon_{quad} + C_{Lip}\varepsilon_{opt}),$$

where Lipschitz's constant of the loss function is relative to the transactionsC_{Lip} .

Proof of Knowledge: Knowing the Remainder

$$r_\Theta(x) = \hat{y}_\Theta(x) - f(x) - \lambda \int_a^b K(x,t)\hat{y}\Theta(t)dt.$$

Subtracting the original equation from the approximation:

$$y^*(x) - \tilde{y}\Theta(x) = \lambda \int_a^b K(x,t)(y^*(t) - \tilde{y}\Theta(t))dt - r\Theta(x).$$

I take a standard l^2

$$\parallel y^* - \tilde{y}\Theta \parallel L^2 \le \mid \lambda \mid\mid K \parallel\mid\mid y^* - \tilde{y}\Theta \parallel L^2 + \parallel r\Theta \parallel L^2.$$

Hence:

$$\parallel y^* - \tilde{y}\Theta \parallel L^2 \le \frac{\parallel r\Theta \parallel L^2}{1-\mid \lambda \mid\mid K \parallel}$$

‖rΘ‖ $L^2$is estimated as the sum of the rounding, squaring, and optimization errors, so we get

$$\parallel y^* - \hat{y}\Theta_{trained} \parallel L^2 \le \frac{1}{1-\mid \lambda \mid\mid K \parallel}(\varepsilon_{approx} + \mid \lambda \mid (b-a)^{\frac{1}{2}}\varepsilon_{quad} + C_{Lip}\varepsilon_{opt}),$$

Result:

To achieve a total error ε, it is enough to choose:

$$N_h = O\left(\varepsilon^{-\frac{2}{2s+1}}\right), N_q = \left(\varepsilon^{-1(2p)}\right),$$

where s is the order of the sweetness of the sweet p the rank of the numerical square

Scientific Importance:

The estimate shows $\parallel y^* - \hat{y}\Theta_{trained} \parallel L^2 \le \frac{1}{1-|\lambda|\|K\|}(\varepsilon_{approx} + \mid \lambda \mid (b-a)^{\frac{1}{2}}\varepsilon_{quad} + C_{Lip}\ \varepsilon_{opt})$, that the total error is separated into three independent components that can be controlled via

Network Capacity (Roundup)

Number of squared points (integrative precision)

Training Algorithm (Optimization)

While the worker reflects

$$(1 - -\mid \lambda \mid\mid K \parallel)^{-1}$$

The property of the problem itself (compression condition)

The Theory of Stability Under Noise

Theory if the data is contaminated so that:

$\parallel f - f^\delta \parallel L2 \le \eta$,

The approximate solution achieves:y^δ

$$\parallel y^* - y^\delta \parallel L^2 \le \frac{\eta}{1-\mid \lambda \mid\mid K \parallel} + \parallel y^* - \tilde{y}\Theta \parallel L^2$$

Proof.

From the sin of the influencer:

$$y^{\delta} - y^{*} = (I - \lambda K)^{-1}(f^{\delta} - f).$$

***Taking the criterion**:*

$$\| y^{\delta} - y^{*} \|_{L^2} \leq \| (I - \lambda K)^{-1} \| \eta \leq \frac{\eta}{1 - |\lambda| \|K\|}.$$

Then by adding the rounding line, we get:

$$\| y^{*} - y^{\delta} \|_{L^2} \leq \frac{\eta}{1 - | \lambda | \| K \|} + \| y^{*} - \tilde{y}\Theta \|_{L^2}$$

Spectral Interpretation of the Mechanism of Attention

Suggestion

that the weight of attention in the trained model converges towards a finite spectral representation of the Friedholm effect:

$$A(x, t; \Theta_A^{*}) \approx \sum_{n=1}^{N_{attn}} \propto_n \phi_n(\boldsymbol{x}) \boldsymbol{\phi_n}(t),$$

where the eigenfunctions of the K $\{\phi n\}$.

Interpretation

This means that:

- *IINN learns almost low-rank for the K influencer*
- *Then across the remaining structure approaches:*

$$(I - \lambda K)^{-1}$$

It is the theoretically optimal spectral representation.

Analysis of the complexity of the algorithms

Table 1. Comparison of computational complexity and memory of the numerical methods studied

Instructions	Array construction	Solution	Inference	Memory
B-Spline	$O(4^M)$	$O(8^m)$	$O(2^m)$	$O(4^m)$
Hair	$O(m^2)$	$O(m^3)$	O(m)	$O(m^2)$
IINN	$O(N_r N_q N_h)$	—	$O(L N_h^2)$	$O(L N_h^2)$
DeepONet	$O(N_r p)$	—	(O(p))	$O(p N_h)$
FNO	*O(NlogN)*	—	O(NlogN)	(O(k d))

Final Analysis

The dominant cost during training at IINN is:

$$O(N_r N_q N_h),$$

Because numerical squared is integrated into the forward pass, while inference remains low-cost:

$$O(LN_h^2),$$

This makes it suitable for multi-query applications.

7. Numerical Results and Statistical Analysis

Preparing Experiments

a. Justice Protocol

To eliminate the systemic bias caused by unequal transaction control, we adopt the following protocol:

- *Classical Methods*: Setting Transactions to Optimal Values in Original References
- *Learned models*: Selecting supercoefficients using five-fold cross-realization.
- *All methods*: Assessment on identical test networks.
- *Repetition*: Perform each experiment 10 times using different random seeds.

Training without precise solutions. In all of the IINN,DeepONet, and FNO experiments, the remainder of the equation is minimized without the use of analytical solutions. Where some models require reference data, high-fidelity numerical solutions

not analytical solutions are used as training objectives. This reflects realistic deployment conditions.

Smooth Core Issues

Table 2. Comparison of the performance of numerical methods on the problem of the exponential nucleus

Instructions	MSE (Deviation±)	∞|E|	Convergence Rate	Time (s)
B-Spline (m=4, M=3)	1.43×10^{-6}	3.21×10^{-4}	O()h^4	0.12
B-Spline (m=4, M=4)	9.12×10^{-8}	2.04×10^{-5}	(O()h^4	0.31
Hair	5.21×10^{-5}	8.73×10^{-3}	O()h	0.09
Heterogeneous repetition	3.64×10^{-5}	6.41×10^{-3}	—	0.21
Taylor	2.81×10^{-7}	4.32×10^{-5}	(O()h^6	0.18
DeepONet	$4.21 \pm 0.83 \times 10^{-7}$	7.81×10^{-5}	—	23.4
FNO	$2.87 \pm\times 0.41 \times 10^{-7}$	5.23×10^{-5}	—	18.7
IINN	$3.14 \pm\times 0.21 \times 10^{-8}$	8.92×10^{-6}	—	16.3

The results in Table 2 showed that the IINN model clearly outperformed in terms of numerical accuracy, as it achieved the lowest error square mean value compared to all other classical methods and neural models. It is noted that B-Spline methods maintain the theoretical convergence rate O(h4), but this comes at the expense of increasing the computational cost when the accuracy level is raised. In contrast, the Har waves suffer from a decrease in the degree of convergence due to discontinuity, which negatively reflects on the accuracy. Modern models such as DeepONet FNO performs well, but does not reach the accuracy of IINN, despite its higher training cost. This superiority can be explained by the fact that IINN integrates the integrative effector within the network architecture, rather than treating it as an external constraint, resulting in a more accurate representation of the solution and improving numerical stability, which is consistent with the results of the theoretical analysis in Section VI.

Table 3. StudentNet t-test results to compare IINN with the best classical methods

Scale	T-value	P value	Significance
MSE	12.47	0.001>	***
$\|E\|_\infty$	9.83	0.001>	***

The results of Table 3 show that the differences between the IINN model and the best classical methods are statistically significant, with values for both scales $p <$ 0.001 reaching MSE, which allows the null hypothesis to be rejected with great confidence. The high values of $\|E\|$ the t-statistic indicate that there is a substantial difference in the mean performance and not just a random fluctuation resulting from the different samples. Thus, it can be concluded that the improvement provided by the IINN It is a real and statistically stable improvement, and reflects a systematic superiority in the representation of solutions compared to traditional methods.

Table 4. Numerical convergence rates of the studied methods at different cutting levels

Instructions	(h=1/8)	(h=1/16)	(h=1/32)	Rank
B-Spline (m=2)	4.21×10^{-3}	1.05×10^{-3}	2.63×10^{-4}	2.00
B-Spline (m=4)	8.73×10^{-5}	3.41×10^{-6}	3.21×10^{-7}	4.00
Hair	2.14×10^{-3}	1.07×10^{-3}	5.35×10^{-4}	1.00
Taylor	1.32×10^{-4}	8.27×10^{-6}	5.17×10^{-7}	3.99

Instructions	(h=1/8)	(h=1/16)	(h=1/32)	Rank
IINN	1.87×10^{-7}	4.31×10^{-8}	2.31×10^{-8}	—

Table 4 shows that classical methods, such as B-Spline and Taylor expansion, accurately achieve the expected theoretical convergence rates, where the degree of convergence is proportional to the degree of the functions used. In contrast, Har waves exhibit only a first-order convergence order due to their non-continuous nature. The IINN model does not rely on the h-cutting coefficient, yet achieves an accuracy that surpasses all methods at the highest level of refining, demonstrating its ability to bypass the traditional constraints of lattice approximation. This behavior is explained by the fact that The model learns a direct spectral representation of the solution rather than relying on local approximations, which gives it a clear edge in accuracy at high levels.

Weak Nuclei

Table 5. Comparison of the performance of numerical methods in the case of weakly unique cores

Instructions	MSE	Note
B-Spline	3.42×10^{-3}	Deterioration
Hair	1.87×10^{-2}	Instability
Heterogeneous repetition	8.93×10^{-3}	Slow
DeepONet	2.14×10^{-5}	Good
FNO	1.83×10^{-5}	Good
IINN	4.21×10^{-6}	Excellent

Table 5 showing that the performance of classical methods deteriorates significantly in the case of poorly differentiated nuclei, where error values are higher due to the difficulty of representing irregular behavior near the point of singularity. Har waves exhibit apparent instability due to their non-continuous nature, while B-Spline suffers from limited singularity capture despite its high rank. In contrast, neural models perform better, but IINN It remains the most precise, as it benefits from the integration of numerical squared within the mechanism of integrative attention, allowing it to adapt to the nature of the nucleus. This behavior confirms that the proposed model is able to deal with irregular cases efficiently, which is a critical advantage over traditional methods.

Nonlinear Equations

Table 6. Comparison of the performance of numerical methods on nonlinear Hammerstein equations

Instructions	MSE	Rapprochement
Simpson	2.41×10^{-4}	yes
Galerkin	8.73×10^{-5}	yes
Adomian	9.81×10^{-5}	yes
DeepONet	3.21×10^{-6}	yes
FNO	2.87×10^{-6}	yes
IINN	4.82×10^{-7}	yes

The results of Table 6 show that all methods are able to achieve convergence in solving a nonlinear equation, but the differences are clearly evident in the level of accuracy. The IINN model achieves the lowest error square mean value compared to all other methods, including modern neural models such as DeepONet and FNO. This indicates that IINN is not limited to handling linear states, but efficiently extends to complex nonlinear systems. This superiority is due to the integration of the integrative effect within the architecture. This allows for a more accurate representation of the nonlinear interaction within the equation and improves the stability of the affinity.

Engineering Applications

Table 7. Results of the application of numerical methods to realistic engineering problems

Instructions	Error (%)	Time
B-Spline	2.34	0.8s
SN	1.87	2.1w
DeepONet	0.43	High Training
IINN	0.12	Acceptable

The results of Table 7 show that the IINN model achieves the highest accuracy in realistic engineering applications, with the lowest error rate compared to all other methods. Although the initial cost of training the model is relatively higher, the inference phase is efficient and fast, making it suitable for applications that require a large number of post-training evaluations. In contrast, classical methods provide less computational time in direct execution, but suffer from low accuracy. Thus, IINN showsAn effective balance between accuracy and practicality, making it a solid choice in multi-query applications.

Stability under noise

Table 8. Effect of Different Noise Levels on the Accuracy of Numerical Methods

The	B-Spline	Hair	IINN	DeepONet
0%	9.12×10^{-8}	5.21×10^{-5}	3.14×10^{-8}	4.21×10^{-7}
1%	1.34×10^{-5}	7.82×10^{-5}	4.73×10^{-7}	8.14×10^{-6}
5%	3.21×10^{-4}	2.41×10^{-4}	3.12×10^{-6}	4.87×10^{-5}
10%	1.87×10^{-3}	8.73×10^{-4}	1.43×10^{-5}	2.31×10^{-4}

The results of Table 8 show that the IINN model has a high degree of stability in the presence of noise, maintaining acceptable accuracy even at high noise levels of up to 10%. In contrast, the accuracy of classical methods deteriorates rapidly with increased noise, reflecting their high sensitivity to data disturbances. The performance of DeepONet is also significantly affected compared to IINN. This behavior can be explained by the IINN integrative attention mechanism acting as an implicit filter that distributes the noise effect over the entire field, reducing its local impact. This conclusion is consistent with stability theory (6.2), where the error degradation rate is determined by a factor providing a direct theoretical explanation of the numerical results$(1-|\lambda|\|K\|)^{-1}$.Integrative Attention Acts as an Internal Noise Filter

8. Critical Discussion and Analysis

Interpretation of the Results

The numerical results in Section VII show a consistent pattern that supports the research's basic hypothesis, namely that the integration of the Friedholm effect within the neural network architecture achieves a systematic superiority over classical methods and modern effects learning models. This superiority can be explained across three interrelated levels.

At the level of numerical accuracy, IINN achieves an average squares error value of 3.14×10^{-8} compared to 9.12×10^{-8} for the fourth-order B-Spline method, an improvement of a full rank, and consistent with the proven error limit in theory, where the total error is broken down into independent, controllable components.

At the level of convergence, classical methods adhere to their known rates such as $O(h^4)$ for B-Spline and O(h) for Har wavelengths, while IINN bypasses these limitations by learning a direct spectroscopic representation of the opposite effect $(I-\lambda K)^{-1}$.

In terms of stability, the stability theory supports the experimental results, with B-Spline performance deteriorating sharply under noise, while IINN maintains higher stability, which directly reflects the effect of the factor $(1-|\lambda|\|K\|)^{-1}$.

a. Critical Comparison of Methodological Groups

- **Classic Routes vs. IINN**

Classical methods rely on representing the solution within a finite space and then converting the problem into an algebraic system. This approach is effective when the solution corresponds to the chosen space, but it imposes obvious limitations, including high computational cost with increased accuracy, poor performance in the case of irregular cores, and high noise sensitivity due to the air conditioning number.In contrast, IINN does not restrict the solution to a pre-space, but rather learns an adaptive spectral representation that reflects the structure of the effector itself, giving it greater flexibility and accuracy.

- **DeepONet and FNO vs . IINN**

Although DeepONet and FNO represent an important development in effects learning, they rely primarily on data without investing in the known mathematical structure of the influencer. This leads to a wider hypothesis space and a greater need for data.

IINN incorporates explicit structural constraints such as integrative squaring and spectral modulation, which reduces the research space and significantly improves the convergence and stability characteristics.

b. Analysis of the results of weakly unique nuclei

The results of weakly unique cores reveal fundamental differences between approaches. B-Spline suffers from low accuracy due to the smoothness of its core functions, and Har wavelengths show instability due to amplification of values near singularity points.In contrast, deep learning models perform better thanks to their representative flexibility, while IINN achieves the best results due to its ability to implement implicit adaptive integration that focuses computational around regions of uniqueness without the need for explicit modification of the structure.

c. Statistical Significance of Results

Statistical tests confirm that the differences between methods are not random. The t-test shows very high statistical significance in favor of the IINN, with p values less than 0.001, demonstrating that the improvement is systematic and stable across multiple iterations.

The confidence ranges also show no overlap between the IINN results and other methods, enhancing the reliability of the conclusions.

d. Computational efficiency: a balanced reading

Although the cost of training IINN is higher than classical methods, this comparison should be taken in the context of use. Traditional methods require recalculation for each new coercive function, while IINN provides immediate generalization after training. Therefore, in applications that require repeat queries such as engineering design or sensitivity analysis, the total cost is in favor of IINN, which is confirmed by applied results that show its superiority in accuracy and efficiency in the long run.

9. Conclusion

a. The IINN model showed a clear numerical superiority over all the comparative methods, achieving a lower value of one to two ranks over traditional methods. $MSE = 3.14 \times 10^{-8}$،
b. The B-Spline outperformed the fourth rank in the classical methods in terms of accuracy, but its performance was significantly reduced when dealing with weakly unique cores compared to the proposed model.
c. Har waves showed the lowest level of accuracy among all the methods studied, due to their limited O(h) numerical convergence rate
d. The IINN model achieved a more stable and efficient convergence than traditional methods, as a result of its reliance on direct inverse spectral representation rather than localized approximations.

e. The statistical results proved the superiority of the proposed model in practical applications, as it recorded a relative error of 0.12% compared to 2.34% for B-Spline, with a strong statistical significance ($p < 0.001$ according to the Student's t-test).

Recommendation

a. Adopting the IINN model in engineering and scientific applications that require high accuracy and remarkable stability under the influence of noise.
b. Use B-Spline methods in cases where computational efficiency and speed of execution are a priority, especially when dealing with regular and seamless solutions.
c. Avoid using HAR waves in matters involving poorly characterized cores or requiring high levels of precision.
d. Extend the IINN model to include fractional integral equations and multidimensional problems to enhance its applicability.
e. Develop the theoretical framework of the model by combining it with Bayesian optimization and uncertainty analysis, as well as studying convergence cases outside of standard conditions.

REFERENCES

[1] A. D. Polyanin and A. V Manzhirov, *Handbook of Integral Equations*, 2nd ed. CRC Press, 2008.

[2] K. Maleknejad and M. N. Sahlan, "The method of moments for solution of second kind Fredholm integral equations based on B-spline wavelets," *Int. J. Comput. Math.*, vol. 87, no. 7, pp. 1602–1616, 2010.

[3] M. Javidi and A. Golbabai, "Modified homotopy perturbation method for solving non-linear Fredholm integral equations," *Chaos, Solitons and Fractals*, vol. 40, no. 3, pp. 1408–1412, 2009.

[4] L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, "Learning nonlinear operators via {DeepONet} based on the universal approximation theorem of operators," *Nat. Mach. Intell.*, vol. 3, no. 3, pp. 218–229, 2021.

[5] S. Lanthaler, S. Mishra, and G. E. Karniadakis, "Error estimates for {DeepONets}: A deep learning framework in infinite dimensions," *Trans. Math. Its Appl.*, vol. 6, no. 1, p. tnac001, 2022.

[6] A. Vaswani *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.

[7] A. M. Hafiz, I. Faiq, and M. Hassaballah, "Solving partial differential equations using large-data models: A literature review," *Artif. Intell. Rev.*, vol. 57, p. 152, 2024.

[8] C. Allouch, D. Sbibih, and M. Tahrichi, "Spectral approximation methods for Fredholm integral equations with non-smooth kernels," *Math. Model. Anal.*, vol. 27, no. 4, pp. 652–667, 2022.

[9] N. Kovachki, S. Lanthaler, and S. Mishra, "On universal approximation and error bounds for Fourier neural operators," *J. Mach. Learn. Res.*, vol. 22, no. 1, pp. 13237–13312, 2021.

[10] M. Lakestani, M. Razzaghi, and M. Dehghan, "Solution of nonlinear Fredholm--Hammerstein integral equations by using semiorthogonal spline wavelets," *Math. Probl. Eng.*, vol. 2005, no. 1, pp. 113–121, 2005.

[11] X.-Y. Lin, J.-S. Leng, and Y.-J. Lu, "A Haar wavelet solution to Fredholm equations," in *Proceedings of the International Conference on Computational Intelligence and Software Engineering (CiSE 2009)*, 2009, pp. 1–4.

[12] J.-H. He, "Variational iteration method --- Some recent results and new interpretations," *J. Comput. Appl. Math.*, vol. 207, no. 1, pp. 3–17, 2007.

[13] K. Maleknejad, N. Aghazadeh, and M. Rabbani, "Numerical solution of second kind Fredholm integral equations system by using a Taylor-series expansion method," *Appl. Math. Comput.*, vol. 175, no. 2, pp. 1229–1234, 2006.

[14] Z. Li *et al.*, "Fourier neural operator for parametric partial differential equations," in *Proceedings of the 9th International Conference on Learning Representations (ICLR)*, 2021.

[15] N. Kovachki *et al.*, "Neural operator: Learning maps between function spaces with applications to {PDEs}," *J. Mach. Learn. Res.*, vol. 24, no. 89, pp. 1–97, 2023.

[16] T. Chen and H. Chen, "Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems," *IEEE Trans. Neural Networks*, vol. 6, no. 4, pp. 911–917, 1995.

[17] S. Wang, S. Sankaran, and P. Perdikaris, "Respecting causality is all you need for training physics-informed neural networks," *arXiv Prepr. arXiv2203.07404*, 2022.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[19] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2010, pp. 249–256.

[20] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019.

[21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, 2015.