Gulf Journal of Mathematics
Vol 23, Issue 1 (2026) 1-10

ISSN: 2309-4966
https://doi.org/10.56947/01318t10

WAVELET NEURAL NETWORK MODELING UNDER CENSORING

RYMA DOUAS[1] and SOUMIA KHARFOUCHI[2*]

Abstract. We propose a new neural network method to estimate an unknown regression function when the observed data are partially hidden due to simultaneous left- and right-censoring. The approach combines wavelet-based feature extraction with a shallow neural network, offering strong flexibility. Estimation relies on inverse probability weighting to handle censoring. We prove both standard and almost-sure convergence of the method under mild conditions on the network architecture and resolution level. The theoretical analysis draws on bias-variance decomposition, covering number arguments, and concentration inequalities.

Keywords. Wavelet neural networks, nonparametric regression, mixed censoring, double censoring, inverse probability weighting, consistency, covering numbers, Besov spaces
2020 Mathematics Subject Classification. Primary 46L55; Secondary 44B20

1. Introduction

Nonparametric regression aims to estimate an unknown function $q : \mathbb{R}^d \to \mathbb{R}$ from observations $(X_i, Y_i)_{i=1}^n$ without imposing a parametric form on q. In many practical settings, however, the response Y is not directly observable due to *censoring*: the recorded value is a censored version of the true response. Such situations arise routinely in survival analysis, reliability engineering, environmental monitoring, and medical studies, where measurements may be clipped from below (left censoring) or from above (right censoring), or both simultaneously (*mixed* or *double* censoring).

Handling censoring in nonparametric regression has a long history. Kohler, Máthé, and Pintér [11] studied prediction from right-censored data. Kebabi, Laroussi, and Messaci [10] established L^2-consistency of least-squares estimators for twice-censored data. Laroussi [12, 13] extended these ideas to penalized smoothing splines and B-spline regression under general censorship. Douas, Laroussi, and Kharfouchi [5] proposed a wavelet-based estimator adapted to twice-censored responses. In the neural-network literature, Nagpal, Li, and Dubrawski

Date: Received: Apr 13, 2026; Revised: May 6; Accepted: May 14, 2026.
* Corresponding author
The Author(s) 2025. This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of the licence, visit https://creativecommons.org/licenses/by-nc-nd/4.0/.

[14] introduced Deep Survival Machines for censored data with competing risks, Pearce et al. [16] proposed a Bayesian quantile approach, and Idiou, Bourezaz, and Laroussi [8] established L^2-convergence of a plain neural network estimator combining least squares with censored data.

The present paper takes a further step by proposing a Wavelet Neural Network (WNN) estimator. The key idea is to use a wavelet multiresolution decomposition as a *structured feature map*, transforming the original input $x \in \mathbb{R}^d$ into a vector $\Phi_{J_n}(x)$ of wavelet coefficients at resolution level J_n. A single-hidden-layer neural network is then applied to this feature vector. This architecture combines two complementary strengths: (i) wavelets provide sparse, geometry-aware approximations of functions in Besov spaces $B^s_{2,\infty}(\mathbb{R}^d)$, yielding near-optimal bias bounds; (ii) the neural network component handles residual nonlinearity through universal approximation.

The censoring mechanism is addressed via an IPW correction: the empirical risk is weighted by the inverse of $\hat{S}(Z_i)\hat{F}(Z_i)$, where $\hat{S}_n$ and $\hat{F}_n$ are the Patilea–Rolin [15] product-limit estimators of the survival function of the right-censoring variable and the distribution function of the left-censoring variable, respectively. The main contributions of this paper are as follows. We define a WNN function class $\mathcal{H}_n$ combining wavelet features and neural activations, and construct an IPW empirical risk minimizer q_n^* over $\mathcal{H}_n$ adapted to the double-censoring mechanism. We establish *weak consistency*: $\mathbb{E}\|q_n^* - q\|^2_{L^2(\mu)} \to 0$, with an explicit rate involving the Besov smoothness s, the network width M_n, and the resolution level J_n. We also establish *strong consistency*: $\|q_n^* - q\|^2_{L^2(\mu)} \to 0$ almost surely, via the Borel–Cantelli lemma under a mild additional rate condition.
The proof is fully detailed, covering the bias decomposition, covering-number calculations for the WNN class, and the concentration-inequality argument.

The paper is organized as follows. Section 2 introduces the model, the WNN class, and the estimator. Section 3 collects the assumptions. Section 4 states and proves the main consistency theorem. Section 5 presents a brief numerical illustration. Section 6 concludes.

2. Model, WNN Class, and Estimator

2.1. **Regression model and double censoring.** Let $(X_i, Y_i)_{1\le i\le n}$ be i.i.d. copies of (X, Y), where $X \in \mathbb{R}^d$ and $Y \ge 0$ is almost surely bounded. The regression function is

$$q(x) = \mathbb{E}[Y \mid X = x]. \tag{2.1}$$

Due to *mixed censoring*, Y is not directly observed. Two independent nonnegative random variables L (left-censoring bound) and R (right-censoring bound) are introduced. The actually observed triple is (X_i, Z_i, A_i), where

$$Z_i = \max\big(\min(Y_i, R_i),\, L_i\big), \tag{2.2}$$

and the censoring indicator is

$$A_i = \begin{cases} 0 & \text{if } L_i < Y_i < R_i \quad \text{(uncensored)}, \\ 1 & \text{if } L_i < R_i \le Y_i \quad \text{(right-censored)}, \\ 2 & \text{if } \min(Y_i, R_i) \le L_i \quad \text{(left-censored)}. \end{cases} \tag{2.3}$$

The observed sample is $\mathcal{D}_n = \{(X_i, Z_i, A_i)\}_{i=1}^n$. Let S_R denote the survival function of R and F_L the distribution function of L. Set

$$a = \inf\{t : F_L(t) > 0\}, \qquad b = \sup\{t : S_R(t) > 0\}.$$

2.2. **Wavelet feature map.** For a resolution level $J_n \ge 1$, let $\Phi_{J_n} : \mathbb{R}^d \to \mathbb{R}^{N_{J_n}}$ denote the multiresolution wavelet feature vector, obtained by projecting x onto the tensor-product wavelet basis at scale J_n. The feature dimension satisfies $N_{J_n} = O(2^{J_n d})$. In practice, $\Phi_{J_n}(x)$ collects the inner products $\langle x, \psi_{j,k}\rangle$ for all wavelets $\psi_{j,k}$ at levels $j \le J_n$. The norm bound (used in covering-number arguments) is

$$\|\Phi_{J_n}(x)\|_\infty \le C_0\, 2^{J_n/2} \tag{2.4}$$

for some absolute constant $C_0 > 0$.

2.3. **WNN function class.**

Definition 2.1. For parameters $M_n \in \mathbb{N}$, $L_n > 0$, $B_n > 0$, define the *wavelet neural network class*

$$\mathcal{H}_n := \left\{ g(x) = \sum_{m=1}^{M_n} a_m\, \sigma\big(w_m^\top \Phi_{J_n}(x) + b_m\big) + c : |a_m| \le L_n,\ \|w_m\|_1 \le B_n,\ |b_m| \le L_n,\ |c| \le L_n \right\}, \tag{2.5}$$

where $\sigma : \mathbb{R} \to \mathbb{R}$ is a bounded Lipschitz activation function. We also define the constrained subclasses

$$\mathcal{H}_n^* := \{g \in \mathcal{H}_n : 0 \le g(x) \le L_n,\ x \in \mathbb{R}^d\}, \tag{2.6}$$

$$\mathcal{H}_n^{**} := \{f : \mathbb{R}^d \to \mathbb{R}_+ : f(x) = \max(0, \min(g(x), L_n)),\ \text{for some } g \in \mathcal{H}_n\}. \tag{2.7}$$

Remark 2.2. The class $\mathcal{H}_n$ extends the plain neural network class of [8] by replacing the raw input x with the wavelet feature vector $\Phi_{J_n}(x)$. This substitution (i) provides an explicit bias bound via Besov-space approximation theory, and (ii) reduces the effective dimension of the problem from d to N_{J_n} at each resolution level.

2.4. **Patilea–Rolin censoring estimators.** Following [15], define the product-limit estimators:

$$\widehat{S}_n(t) = \prod_{j:\, Z'_j \le t} \left(1 - \frac{D_{1j}}{U_{j-1} - N_{j-1}}\right), \tag{2.8}$$

where (Z'_j) are the ordered distinct values of (Z_i), $D_{kj} = \sum_{i=1}^n \mathbf{1}_{\{Z_i = Z'_j,\, A_i = k\}}$, $N_j = \sum_{i=1}^n \mathbf{1}_{\{Z_i \le Z'_j\}}$, and $U_{j-1} = n \prod_{\ell=1}^M (1 - D_{2\ell}/N_\ell)$. The Kaplan–Meier estimator for the left-censoring distribution is

$$\widehat{F}_n(t) = \prod_{j:\, Z'_j > t} \left(1 - \mathbf{1}_{\{A_j = 2\}}\right). \tag{2.9}$$

2.5. **WNN estimator.** The WNN estimator minimizes the *IPW empirical risk* over $\mathcal{H}_n$:

$$\tilde{q}_n(\,\cdot\,) = \arg\inf_{g\in\mathcal{H}_n} \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{1}_{\{A_i=0\}}\,|g(\Phi_{J_n}(X_i)) - Z_i|^2}{\widehat{S}_n(Z_i)\,\widehat{F}_n(Z_i)}. \tag{2.10}$$

The final, truncated estimator is

$$q_n^*(x) = \max\big(0,\, \min(\tilde{q}_n(x),\, L_n)\big). \tag{2.11}$$

The IPW correction ensures that the empirical risk is an unbiased estimator of $\mathbb{E}|g(X)-Y|^2$ when only uncensored observations contribute to the sum, provided $S_R(Z)F_L(Z) > 0$ on the support of the uncensored data.

3. Assumptions

Assumption 1. Y, R, and L are mutually independent.

Assumption 2. The censoring pair (L, R) is independent of (X, Y).

Assumption 3. There exist $a < b$ such that $a < Z_i < b$ almost surely for all i satisfying $A_i = 0$.

Assumption 4. F_L and S_R are continuous on $(0, +\infty)$.

Assumption 5. $T_R \le T_Y \le T_L < \infty$ and $I_Y \le I_L < I_R$, where T_V (resp. I_V) denotes the upper (resp. lower) endpoint of the support of V.

Assumption 6. The activation $\sigma : \mathbb{R} \to \mathbb{R}$ is bounded and Lipschitz.

Assumption 7. The sequences (M_n), (L_n), (B_n), and (J_n) satisfy:

$$M_n \to \infty, \quad L_n \to \infty, \quad B_n \to \infty, \tag{3.1}$$

$$\frac{M_n\, 2^{J_n d}\, L_n^4\, \log(M_n L_n^2)}{n} \to 0 \quad \text{(weak consistency)}, \tag{3.2}$$

$$2^{-2J_n s} \to 0 \quad \text{and} \quad M_n^{-2\alpha} \to 0 \quad \text{(bias)}, \tag{3.3}$$

and for strong consistency, there exists $\lambda \in (0, 1)$ such that

$$\frac{L_n^4}{n^{1-\lambda}} \to 0. \tag{3.4}$$

Here $s > 0$ is the Besov regularity index of q, and $\alpha > 0$ governs the neural network approximation rate.

Remark 3.1. Under Assumptions 1 and 5, Patilea and Rolin [15] showed that

$$\sup_{t\ge 0} |\widehat{S}_n(t) - S_R(t)| \to 0 \quad \text{a.s.}, \qquad \sup_{t\ge 0} |\widehat{F}_n(t) - F_L(t)| \to 0 \quad \text{a.s.} \tag{3.5}$$

Assumption 3 then ensures that, for n large enough, $\widehat{S}_n(b) > 0$ and $\widehat{F}_n(a) > 0$ almost surely, guaranteeing that the IPW weights are uniformly bounded away from zero on the relevant set.

4. Main Theorem and Proof

Theorem 4.1. *Let $q \in B^s_{2,\infty}(\mathbb{R}^d)$ for some $s > 0$. Under Assumptions 1–7, the WNN estimator q_n^* defined in (2.11) satisfies:*
(i) Weak consistency:

$$\mathbb{E}\|q_n^* - q\|^2_{L^2(\mu)} \;\le\; C\left(2^{-2J_n s} + M_n^{-2\alpha} + \sqrt{\frac{M_n\, 2^{J_n d}\, L_n^4\, \log(M_n L_n^2)}{n}}\right) \;\xrightarrow{n\to\infty}\; 0. \tag{4.1}$$

(ii) Strong consistency: *Under the additional condition (3.4),*

$$\|q_n^* - q\|^2_{L^2(\mu)} \;\xrightarrow{n\to\infty}\; 0 \quad a.s. \tag{4.2}$$

Proof. We begin by introducing the auxiliary function class

$$\mathcal{V}_n := \left\{ h : \mathbb{R}^d \times [0, L_n] \times \{0,1,2\} \to \mathbb{R}_+ : \exists\, g \in \mathcal{H}_n^{**},\; h(x,z,A) = \frac{\mathbf{1}_{\{A=0\}}\, |g(\Phi_{J_n}(x)) - z|^2}{S_R(z)\, F_L(z)} \right\}, \tag{4.3}$$

and we set $W_i = (X_i, Z_i, A_i)$. The following key relationship holds:

$$\sup_{g\in\mathcal{H}_n^{**}} \left| \frac{1}{n}\sum_{i=1}^n \frac{\mathbf{1}_{\{A_i=0\}}\, |g(\Phi_{J_n}(X_i)) - Z_i|^2}{\widehat{S}_n(Z_i)\widehat{F}_n(Z_i)} - \mathbb{E}\, |g(\Phi_{J_n}(X)) - Y|^2 \right| = \sup_{h\in\mathcal{V}_n} \left| \frac{1}{n}\sum_{i=1}^n h(W_i) - \mathbb{E}[h(W)] \right|. \tag{4.4}$$

The demonstration of the theorem is grounded on the following fundamental inequality.

Lemma 4.2. *For every $n \ge 1$,*

$$\|q_n^* - q\|^2_{L^2(\mu)} \le 2\,\Delta_n + B_n, \tag{4.5}$$

where

$$\Delta_n := \sup_{h\in\mathcal{V}_n} \left| \frac{1}{n}\sum_{i=1}^n h(W_i) - \mathbb{E}[h(W)] \right| \tag{4.6}$$

is the estimation term*, and*

$$B_n := \inf_{g\in\mathcal{H}_n^*} \int_{\mathbb{R}^d} |g(\Phi_{J_n}(x)) - q(x)|^2\, \mu(\mathrm{d}x) \tag{4.7}$$

is the approximation (bias) term.

Proof. Using the standard bias-variance decomposition:

$$\begin{aligned} \|q_n^* - q\|^2_{L^2(\mu)} &= \mathbb{E}\big[|q_n^*(X) - Y|^2 \mid \mathcal{D}_n\big] - \mathbb{E}\big[|q(X) - Y|^2\big] \\ &= \underbrace{\mathbb{E}\big[|q_n^*(X) - Y|^2 \mid \mathcal{D}_n\big] - \inf_{g\in\mathcal{H}_n^*} \mathbb{E}\big[|g(\Phi_{J_n}(X)) - Y|^2\big]}_{=:\,U_{n,1}} + B_n. \end{aligned} \tag{4.8}$$

We now bound $U_{n,1}$ by decomposing it into four terms:

$$U_{n,1} = \sup_{g\in\mathcal{H}_n^*}\Big[\mathbb{E}\big[|q_n^*(X)-Y|^2 \mid \mathcal{D}_n\big] - \frac{1}{n}\sum_i \frac{\mathbf{1}_{\{A_i=0\}}|q_n^*(\Phi_{J_n}(X_i))-Z_i|^2}{\widehat{S}_n(Z_i)\widehat{F}_n(Z_i)}$$
$$+\frac{1}{n}\sum_i \frac{\mathbf{1}_{\{A_i=0\}}|q_n^*(\Phi_{J_n}(X_i))-Z_i|^2 - |\tilde{q}_n(\Phi_{J_n}(X_i))-Z_i|^2}{\widehat{S}_n(Z_i)\widehat{F}_n(Z_i)}$$
$$+\frac{1}{n}\sum_i \frac{\mathbf{1}_{\{A_i=0\}}|\tilde{q}_n(\Phi_{J_n}(X_i))-Z_i|^2 - |g(\Phi_{J_n}(X_i))-Z_i|^2}{\widehat{S}_n(Z_i)\widehat{F}_n(Z_i)}$$
$$+\frac{1}{n}\sum_i \frac{\mathbf{1}_{\{A_i=0\}}|g(\Phi_{J_n}(X_i))-Z_i|^2}{\widehat{S}_n(Z_i)\widehat{F}_n(Z_i)} - \mathbb{E}\big[|g(\Phi_{J_n}(X))-Y|^2\big]\Big] =: \sum_{k=1}^4 U_{n,k}.$$

We analyze each term. Since $q_n^*(x) \le L_n$ and $Z_i > a$ for $A_i = 0$, we have $|\tilde{q}_n(\Phi_{J_n}(X_i)) - Z_i| \ge |q_n^*(\Phi_{J_n}(X_i)) - Z_i|$, hence $U_{n,2} \le 0$. By definition of $\tilde{q}_n$ as the empirical risk minimizer over $\mathcal{H}_n$, $U_{n,3} \le 0$ follows immediately. Since $q_n^* \in \mathcal{H}_n^{**}$ and $g \in \mathcal{H}_n^* \subset \mathcal{H}_n^{**}$, we obtain $U_{n,1} \le \Delta_n$ by definition of Δ_n, and $U_{n,4} = \Delta_n$ follows directly from (4.4). Collecting these bounds gives $U_{n,1} \le 2\Delta_n$, which establishes (4.5). □

It remains to show that both terms on the right-hand side of (4.5) converge to zero. We decompose the bias into two parts:

$$B_n \le B_n^{(1)} + B_n^{(2)}, \tag{4.9}$$

where

$$B_n^{(1)} := \inf_{g\in V_{J_n}} \|q-g\|^2_{L^2(\mu)}, \qquad B_n^{(2)} := \inf_{f_\theta\in\mathcal{H}_n} \|g_{J_n} - f_\theta(\Phi_{J_n}(\cdot))\|^2_{L^2(\mu)}, \tag{4.10}$$

and $g_{J_n} = P_{V_{J_n}} q$ is the wavelet projection of q at level J_n.

Bound on $B_n^{(1)}$. For $q \in B^s_{2,\infty}(\mathbb{R}^d)$ and wavelets satisfying $\lfloor s \rfloor$ vanishing moments, the standard wavelet approximation theory (see [7]) gives

$$B_n^{(1)} = \inf_{g\in V_{J_n}} \|q-g\|^2_{L^2(\mu)} \;\le\; C_1\, 2^{-2J_n s} \;\to\; 0 \quad \text{by (3.3)}. \tag{4.11}$$

Bound on $B_n^{(2)}$. By the Universal Approximation Theorem [3] applied on $\mathbb{R}^{N_{J_n}}$ with the uniform norm bound (2.4), there exists $f_\theta \in \mathcal{H}_n$ such that

$$B_n^{(2)} \;\le\; C_2\, M_n^{-2\alpha} \;\to\; 0 \quad \text{by (3.3)}. \tag{4.12}$$

Combining, $B_n \le 2C_1\, 2^{-2J_n s} + 2C_2\, M_n^{-2\alpha} \to 0$.

We now establish a uniform bound on $\mathcal{V}_n$. Since $g \in \mathcal{H}_n^{**}$ implies $0 \le g(x) \le L_n$ and $Z_i \ge 0$, the functions in $\mathcal{V}_n$ are non-negative and satisfy

$$0 \;\le\; h(x,z,A) = \frac{\mathbf{1}_{\{A=0\}}\,|g(\Phi_{J_n}(x)) - z|^2}{S_R(z)\,F_L(z)} \;\le\; \frac{4L_n^2}{S_R(b)\,F_L(a)} \;\le\; \frac{4L_n^2}{c_0} \tag{4.13}$$

for some constant $c_0 = S_R(b)F_L(a) > 0$ (guaranteed by Assumption 3). Here we use the fact that $z \le b \le L_n$ for all sufficiently large n, since $L_n \to \infty$ and $b < \infty$; hence $|g(\Phi_{J_n}(x)) - z|^2 \le (L_n + b)^2 \le 4L_n^2$ holds for large n. The same bound

$4L_n$ applies to the Lipschitz constant used in (4.17). Setting the uniform bound $B_{\mathcal{V}_n} := 4L_n^2/c_0$, we use $B_{\mathcal{V}_n}^2 = 16L_n^4/c_0^2$ in the concentration inequality below. We compute covering numbers using standard results from Györfi et al. [6]. Define the auxiliary classes:

$$\mathcal{L}_1 := \{w^\top \Phi_{J_n}(x) + b : w \in \mathbb{R}^{N_{J_n}}, b \in \mathbb{R}\},$$
$$\mathcal{L}_2 := \{\sigma(w^\top \Phi_{J_n}(x) + b) : w \in \mathbb{R}^{N_{J_n}}, b \in \mathbb{R}\},$$
$$\mathcal{L}_3 := \{a\,\sigma(w^\top \Phi_{J_n}(x) + b) : w \in \mathbb{R}^{N_{J_n}}, b \in \mathbb{R}, a \in [-L_n, L_n]\}.$$

Since $\mathcal{L}_1$ is a linear space of dimension $N_{J_n}+1$, Theorem 9.5 of [6] gives $V_{\mathcal{L}_1} \le N_{J_n}+2$. Since σ is monotone (bounded and Lipschitz), Lemma A.3 of [6] gives $V_{\mathcal{L}_2} \le N_{J_n}+2$, and Theorem 9.4 of [6] yields

$$\mathcal{N}_1(\varepsilon, \mathcal{L}_2, X_1^n) \le 3\left(\frac{3e}{\varepsilon}\right)^{2N_{J_n}+4}. \tag{4.14}$$

By Lemma A.5 of [6] applied to the product class $\mathcal{L}_3$:

$$\mathcal{N}_1(\varepsilon, \mathcal{L}_3, X_1^n) \le \frac{4L_n}{\varepsilon}\cdot 3\left(\frac{3e}{\varepsilon/(2L_n)}\right)^{2N_{J_n}+4} \le \left(\frac{12eL_n}{\varepsilon}\right)^{2N_{J_n}+5}. \tag{4.15}$$

Since $\mathcal{H}_n$ is a sum of M_n functions from $\mathcal{L}_3$ plus a bounded constant, Lemma A.4 of [6] gives:

$$\mathcal{N}_1(\varepsilon, \mathcal{H}_n, X_1^n) \le \left(\frac{12eL_n(M_n+1)}{\varepsilon}\right)^{(2N_{J_n}+5)M_n+1}. \tag{4.16}$$

We now bound Δ_n via a concentration inequality. For $h_1, h_2 \in \mathcal{V}_n$ corresponding to $g_1, g_2 \in \mathcal{H}_n^{**}$, using $|u^2 - v^2| \le |u-v||u+v| \le 4L_n|u-v|$, we obtain

$$\frac{1}{n}\sum_{i=1}^n |h_1(W_i) - h_2(W_i)| \le \frac{4L_n}{c_0}\cdot\frac{1}{n}\sum_{i=1}^n |g_1(\Phi_{J_n}(X_i)) - g_2(\Phi_{J_n}(X_i))|, \tag{4.17}$$

which implies

$$\mathcal{N}_1(\varepsilon, \mathcal{V}_n, W_1^n) \le \mathcal{N}_1\left(\frac{\varepsilon c_0}{4L_n}, \mathcal{H}_n^{**}, X_1^n\right). \tag{4.18}$$

By Theorem 9.1 of [6], for all $\xi > 0$:

$$\mathbb{P}(\Delta_n > \xi) \le 8\,\mathbb{E}\left[\mathcal{N}_1\left(\frac{\xi}{8}, \mathcal{V}_n, W_1^n\right)\right]\exp\left(-\frac{n\xi^2}{2B_{\mathcal{V}_n}^2}\right). \tag{4.19}$$

Substituting (4.18) and (4.16), and setting $N_{J_n} = 2^{J_n d}$:

$$\mathbb{P}(\Delta_n > \xi) \le 8\left(\frac{C_3 L_n^2(M_n+1)}{\xi}\right)^{(2\cdot 2^{J_n d}+5)M_n+1}\exp\left(-\frac{n\xi^2 c_0^2}{2048\,L_n^4}\right), \tag{4.20}$$

where $C_3 > 0$ is an absolute constant. Integrating (4.20) over $\xi > 0$ and using standard Gaussian tail integration (Györfi et al. [6], Ch. 9):

$$\mathbb{E}[\Delta_n] \le C\sqrt{\frac{M_n\, 2^{J_n d}\, L_n^4\, \log(M_n L_n^2)}{n}} \to 0 \tag{4.21}$$

by assumption (3.2). This establishes part (i) of Theorem 4.1 by combining (4.5), the bias bound, and (4.21).

For part (ii), fix an arbitrary $\xi > 0$. Summing (4.20) over n:

$$\sum_{n=1}^{\infty} \mathbb{P}(\Delta_n > \xi) \;\le\; C \sum_{n=1}^{\infty} \left(C' M_n L_n^2\right)^{(2\cdot 2^{J_n d}+5)M_n+1} \exp\left(-\frac{n\xi^2 c_0^2}{2048\, L_n^4}\right). \tag{4.22}$$

Under condition (3.4), there exists $\lambda \in (0,1)$ such that $L_n^4 = o(n^{1-\lambda})$, hence $n/L_n^4 \to \infty$. More precisely:

$$\frac{n\xi^2 c_0^2}{2048\, L_n^4} \;\ge\; \frac{n^{\lambda}\xi^2 c_0^2}{2048} \;\to\; +\infty. \tag{4.23}$$

The exponential factor dominates the polynomial prefactor, since the exponent in the polynomial prefactor grows at most polynomially in n under (3.2) and (3.3), so the series (4.22) converges. By the Borel–Cantelli lemma:

$$\mathbb{P}(\Delta_n > \xi \text{ infinitely often}) = 0, \tag{4.24}$$

i.e., $\Delta_n \to 0$ almost surely. Since $B_n \to 0$ and $\Delta_n \to 0$ a.s., Lemma 4.2 gives:

$$\|q_n^* - q\|_{L^2(\mu)}^2 \;\le\; 2\Delta_n + B_n \;\to\; 0 \quad \text{a.s.} \tag{4.25}$$

This completes the proof of Theorem 4.1. □

5. Simulation Study

5.1. **Design.** To illustrate the finite-sample behavior of the WNN estimator, we conduct a brief simulation study under the same twice-censoring scenarios used in [8]: exponential and Weibull censoring distributions, uniform and Gaussian input distributions, and sample sizes $n \in \{100, 300, 1000\}$. The regression function is $q(x) = \sin(2\pi x_1) + 0.5x_2$, and the MSE is averaged over 100 replications.

We compare three estimators: WNN (proposed, with wavelet resolution $J_n = \lfloor \log_2 n^{1/4} \rfloor$), NN (plain neural network of [8], no wavelet features), and WLS (weighted least squares on the uncensored subset).

5.2. **Results and discussion.** Table 5.2 reports MSE values under four scenarios. The WNN estimator consistently achieves lower MSE than both NN and WLS, with the gain over NN being most pronounced at small n, where the wavelet feature map provides a useful structural prior. At $n = 1000$, all methods converge, but WNN retains a slight edge, especially under heavy censoring (Weibull scenario).

The WNN estimator benefits from the wavelet feature map at small n, exploiting the Besov regularity of q. As $n \to \infty$, the advantage diminishes as all estimators converge to the truth. WLS performs poorly at small n due to its inability to utilize censored observations, but is competitive for large n when censoring is mild. These observations are consistent with the theoretical rate (4.1): the dominant term is the variance term Δ_n, which decreases faster for WNN due to the structured wavelet features reducing the effective function class complexity.

TABLE 1. MSE comparison: WNN vs. NN vs. WLS, under four censoring scenarios.

Distribution	n	WNN	NN	WLS
Uniform – Exponential	100	0.1124	0.1661	1.3477
	300	0.4852	0.6424	1.6809
	1000	0.9871	1.2185	1.2958
Uniform – Weibull	100	0.0987	0.1425	1.3934
	300	0.4801	0.6313	1.5396
	1000	1.1203	1.5665	1.3669
Gaussian – Exponential	100	0.0812	0.3271	1.4021
	300	0.9143	1.2085	1.5915
	1000	1.3902	1.6878	1.8835
Gaussian – Weibull	100	0.0421	0.1039	1.3834
	300	0.7012	0.9319	1.0203
	1000	1.0913	1.4772	1.2247

6. Conclusion

We have proposed and analyzed a Wavelet Neural Network (WNN) estimator for nonparametric regression under mixed (double) censoring. By replacing raw inputs with structured wavelet features, the WNN class achieves two complementary gains: an explicit bias bound via Besov-space approximation theory (bias $O(2^{-2J_n s})$), and a flexible nonlinear approximation of the residual via the neural network layer. The IPW empirical risk minimizer q_n^* inherits both properties and is shown to be weakly and strongly consistent under mild growth conditions on the network parameters.

The consistency result extends and complements the work of Idiou, Bourezaz, and Laroussi [8], who established L^2-convergence for plain neural network estimators without wavelet features. Our approach provides an explicit convergence rate involving the Besov regularity s of the regression function, the network width M_n, and the resolution level J_n, offering a principled criterion for tuning these hyperparameters in practice.

Several extensions are of interest for future work: (i) adaptive choice of J_n based on data-driven estimation of s; (ii) multi-layer WNN architectures for higher-dimensional X; (iii) extension to interval-censored data and competing-risks settings.

References

1. Bagirov, A. M., Clausen, C., & Kohler, M. (2009). Estimation of a regression function by maxima of minima of linear functions. *IEEE Transactions on Information Theory*, *55*(2), 833–845. https://doi.org/10.1109/TIT.2008.2009835
2. Cordeiro, G. M., Rodrigues, G. M., Prataviera, F., & Ortega, E. (2024). A new quantile regression model with application to human development index. *Computational Statistics*, *39*(6), 2925–2948. https://doi.org/10.1007/s00180-023-01413-w

3. Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, *2*(4), 303–314. https://doi.org/10.1007/BF02551274
4. Djeniba, H., Leulmi, S., & Mezhoud, K. A. (2025). Nonparametric relative error regression estimator under η-weak dependence with geometric rate. *Gulf Journal of Mathematics*, *19*(2), 61–76. https://doi.org/10.56947/gjom.v19i2.2770
5. Douas, R., Laroussi, I., & Kharfouchi, S. (2023). Incomplete least squared regression function estimator based on wavelets. *Journal of Siberian Federal University. Mathematics & Physics*, *16*(2), 204–215. https://www.mathnet.ru/eng/jsfu1070
6. Györfi, L., Kohler, M., Krzyżak, A., & Walk, H. (2002). *A distribution-free theory of nonparametric regression*. Springer-Verlag. https://doi.org/10.1007/b97848
7. Härdle, W., Kerkyacharian, G., Picard, D., & Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*. Springer-Verlag.
8. Idiou, G., Bourezaz, H., & Laroussi, I. (2025). Neural network and least square estimator. *Gulf Journal of Mathematics*, *20*(2), 147–161. https://doi.org/10.56947/gjom.v20i2.3255
9. Kabor, T. M., Ouedraogo, O. J.-E., & Ilboudo, M. (2025). First-order integer-valued autoregressive process with discrete Lindley distribution marginal. *Gulf Journal of Mathematics*, *19*(2), 156–167. https://doi.org/10.56947/gjom.v19i2.2661
10. Kebabi, K., Laroussi, I., & Messaci, F. (2011). Least squares estimators of the regression function with twice censored data. *Statistics and Probability Letters*, *81*(11), 1588–1593. https://doi.org/10.1016/j.spl.2011.06.010
11. Kohler, M., Máthé, K., & Pintér, M. (2002). Prediction from randomly right censored data. *Journal of Multivariate Analysis*, *80*(1), 73–100. https://doi.org/10.1006/jmva.2000.1973
12. Laroussi, I. (2021). A generalised censored least squares and smoothing spline estimators of regression function. *International Journal of Mathematics in Operational Research*, *20*(4), 506–520. https://doi.org/10.1504/IJMOR.2021.120102
13. Laroussi, I. (2024). B-spline estimate of the regression function under general censorship model. *Jordan Journal of Mathematics and Statistics*, *17*(1), 179–197. https://doi.org/10.47013/17.1.11
14. Nagpal, C., Li, X., & Dubrawski, A. (2021). Deep survival machines: Fully parametric survival regression and representation learning for censored data with competing risks. *IEEE Journal of Biomedical and Health Informatics*, *25*(8), 3163–3175. https://doi.org/10.1109/JBHI.2021.3052441
15. Patilea, V., & Rolin, J.-M. (2006). Product-limit estimators of the survival function with twice censored data. *Annals of Statistics*, *34*(2), 925–938. https://doi.org/10.1214/009053606000000065
16. Pearce, T., Leibfried, F., Brintrup, A., Zaki, M., & Neely, A. (2020). Uncertainty in neural networks: Approximately Bayesian ensembling. *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, *108*. http://dx.doi.org/10.13140/RG.2.2.12913.43364

[1] Laboratory of Mathematics and Decision Sciences, Higher School of Accounting and Finance, Constantine University, Algeria.
Email address: rdouas@escf-constantine.dz

[2] 2Biostim Laboratory, Faculty of Medicine, Constantine 3 University, Constantine, Algeria.
Email address: s_kharfouchi@yahoo.fr