Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2026 (44) 17** : 1–13.
ISSN-0037-8712
doi:10.5269/bspm.82344

A Bayesian Conditional Diffusion Regression Framework with Hierarchical Global-Local Priors for Robust Agricultural Image-to-Yield Prediction

Ch. Shashi Kumar*, G. Upender Reddy and K. Varalakshmi

ABSTRACT: Accurate estimation of crop yield from image-derived data remains difficult due to the combined effects of high predictor dimensionality, strong inter-feature dependence, non-standard noise behavior, and the presence of anomalous observations. Many commonly used regression approaches, including Bayesian models with Gaussian error assumptions, struggle to remain stable under these conditions and often provide unreliable uncertainty estimates. This work presents Bayesian Conditional Diffusion Regression (BCDR), a probabilistic modeling framework that approaches regression through conditional response generation. The method integrates a diffusion-based mechanism for modeling the response variable with hierarchical global–local shrinkage to control model complexity, while a heavy tailed likelihood improves robustness to outliers. Posterior inference is carried out using a hybrid Gibbs and Hamiltonian Monte Carlo strategy to enable tractable estimation in high-dimensional settings. Experiments conducted on synthetic datasets and a real agricultural case study involving pomegranate images and fruit weight measurements demonstrate improved predictive accuracy, better uncertainty calibration, and stronger resistance to data contamination when compared with established statistical and machine-learning baselines. These results suggest that BCDR is a robust Bayesian framework for image-based yield prediction.

Keywords: Bayesian regression, conditional diffusion models, hierarchical global-local priors, Student-t likelihood, robust statistics, Hamiltonian Monte Carlo, agricultural informatics, high-dimensional inference.

Contents

* Corresponding author.
2020 *Mathematics Subject Classification:* 62F15, 62J05, 62F25.
Submitted March 03, 2026. Published June 19, 2026.

1. Introduction

Regression analysis is also a key step in agricultural informatics to relate image-based descriptors with continuous traits (yield, biomass and fruit weight) [1,2]. With the advancements in sensing systems and high-throughput phenotyping, large numbers of visual features that describe crop appearance, texture, and structure have been obtained [3]. Although informative, these features pose significant modeling challenges as the ensuing data are usually high-dimensional, heavily dependent and noise comprised all enabledby environmental variations and imaging noise [4].

In a finite sample setting, these models offer unstable parameter estimates as multicollinearity and small n lead to inflated estimator variances [5]. Regularization methods (such as ridge regression and the lasso) are designed to address this problem by shrinking coefficient estimates or encouraging sparsity [6,7]. However, these methods often yield point predictions and do not conduct principled uncertainty estimation [14], which is important for confidence interval prediction in agriculture [7]. Furthermore, they are vulnerable to outliers (often present in field data) due to the use of squared-error loss.

Probabilistic methods like Bayesian regression models provide a more natural methodology as they can make use of prior information and generating full posterior distributions [9]. Their effectiveness, however, depends strongly on the choice of likelihood and prior structure. Gaussian likelihoods inadequately represent heavy tailed errors [10], while fixed shrinkage priors apply uniform regularization across predictors, often obscuring the distinction between informative and irrelevant variables in high- dimensional feature spaces [11].

Recent progress in diffusion-based generative modeling demonstrates how complex conditional distributions can be learned through iterative noise perturbation and removal [12,13]. Although these models are typically applied to data generation tasks, their conditional formulation motivates a different view of regression, where the response variable is generated conditionally on observed features. Building on this idea, the present work proposes Bayesian Conditional Diffusion Regression, which combines conditional generative modeling with adaptive Bayesian regularization and robust error handling in a unified framework.

1.1. Contributions

The main contributions of this work are summarized as follows

1. Alternative View of Regression:

A Bayesian regression framework is introduced that formulates continuous prediction as a conditional generation process, extending diffusion-based modeling concepts to scalar response Prediction problems [12,13].

2. Robust and Adaptive Structure:

Such attributes give the hierarchical global–local shrinkage in conjunction with heavy-tailed error model a stabilizing effect against outliers and at the same time offer adaptive model complexity control for high-dimensional features [10,14].

3. Practical Inference Strategy:

Efficient posterior estimation for hierarchical and non-linear models is facilitated by a hybrid inference procedure that combines Gibbs sampling and Hamiltonian Monte Carlo [9,30].

4. Application to Agricultural Data:

The effectiveness of the approach is also experimented on both synthetic datasets and an image-based agriculture data of pomegranate fruit weighing towards betterment over some well-known statistical, machine learning counterparts [19,33,34,35,36].

The rest of this paper is structured as follows: Section 2 reviews related work. The mathematical description of the BCDR model is presented in Section 3. The hybrid inference algorithm is presented in Section 4. Section 5 outlines the experimental design. Section 6 presents and discusses results. Section 7 concludes and suggests future research directions.

2. Literature Review

2.1. Robust Bayesian Regression

Handling outliers and departures from Gaussian noise has long been a focus of regression research. Classical robust methods include M-estimators and least absolute deviation approaches, which limit the influence of extreme observations [8,15]. Bayesian formulations commonly address robustness by adopting heavy-tailed error models, most notably the Student t distribution, often implemented through scale-mixture representations to preserve computational tractability [10,16,17]. These approaches have been shown to improve stability in the presence of contaminated data, particularly in linear and moderately high-dimensional settings [18,19]. However, their effectiveness can diminish when predictor dimensionality is large and strong correlations are present.

2.2. Shrinkage Priors and Sparse Bayesian Learning

High-dimensional regression requires mechanisms to control model complexity and suppress irrelevant predictors. Early Bayesian solutions, such as the Bayesian lasso, impose Laplace priors to mimic L1 regularization [20]. While effective for inducing sparsity, such priors may excessively shrink large signals. This limitation motivated the development of global-local shrinkage priors, including the Horseshoe, Dirichlet–Laplace, and R2-D2 priors, which combine an overall shrinkage parameter with predictor-specific local scales [14,21,22]. These priors offer improved separation between signal and noise and have strong theoretical support for sparse recovery in high-dimensional problems.

2.3. Diffusion Probabilistic Models

Inspired by non-equilibrium thermodynamics, diffusion models [12,23] have emerged as a leading class of deep generative models. They operate by progressively adding noise to data (forward process) and then learning to reverse this process (reverse process) using a neural network. Their iterative, noise-prediction nature often yields higher sample quality and diversity compared to Generative Adversarial Networks (GANs) [24]. Conditional diffusion models [25,26] learn p(data|condition)) , enabling controlled generation. They have seen success in image-to-image translation, inpainting, and super-resolution. The application of diffusion models to regression, as proposed here, is a significant departure from their standard use for high-dimensional output generation, treating the scalar response as a 1D data point to be generated conditioned on a potentially high-dimensional input.

2.4. Bayesian Deep Learning

Bayesian deep learning aims to combine the representational capacity of neural networks with principled uncertainty quantification [27]. Common inference strategies include variational methods and Markov chain Monte Carlo techniques [28,29]. Among these, Hamiltonian Monte Carlo and its adaptive variants are particularly effective for sampling from complex posterior distributions due to their use of gradient information [30]. Hybrid schemes that exploit conditional conjugacy alongside gradient-based sampling have been shown to improve efficiency in hierarchical and high-dimensional models.

2.5. Agricultural Informatics and Yield Prediction

Image-based modeling has become central to agricultural informatics, supporting applications such as disease detection, fruit counting, and yield estimation [31,32,33]. Yield prediction models usally use one or several hand-crafted or deep-learned image features, which could be noisy and highly correlated [34,35]. New research into pomegranate imaging addresses the affect of occlusion feature quality, and thus complicating downstream regression [36]. The literature review presents the challenges of learning from such noisy data that highlights two key properties of regression models: their ability to deal with anomalies in the training data, and our trust on their uncertainty estimates.

3. Model Formulation

3.1. Problem Statement and Notation

We consider a supervised regression setting with (n) independent and identically distributed observations.

Let $x = [x_1, \ldots, x_N]^t \in \mathbb{R}^{n\times p}$ denote the design matrix of p -dimensional predictors (image features), and $y = (y_1, \ldots, y_N)^t \in \mathbb{R}^n$ be the corresponding vector of continuous response variables (e.g., fruit weights). We assume a general relationship $y = f(x) + \epsilon$, where $f : \mathbb{R}^p \to \mathbb{R}$ is an unknown function and ϵ is a noise vector. The objective is to learn the conditional distribution $p(y|x)$ in a manner that is robust to outliers, provides valid uncertainty estimates, and performs effectively when (p) is large relative to n.

3.2. Conditional Diffusion Process for Regression

We reframe the prediction of y given as a conditional generative modeling task. Inspired by denoising diffusion models, we define a forward noising process that gradually corrupts a clean" response (y_0) (conditional on x) into pure Gaussian noise over T discrete timesteps.

The forward process is a fixed Markov chain defined by the variance schedule

$$\{\beta_t \in (0,1)\}_{t=1}^t : q(y_t|y_{t-1}) = N\left(y_t;\ \sqrt{1-\beta_t}y_{t-1}, \beta_t\right), t = 1, \ldots, t. \tag{3.1}$$

A notable property is that we can sample y_t directly from y_0

$$q(y_t|y_0) = N\left(y_t;\ \sqrt{\hat{a}_t}\ y_0, (1-\hat{a}_t)\right), \tag{3.2}$$

Where $a_t = 1 - \beta_t$ and $\hat{a}_t = \prod_{s=1}^t a_s$.

The reverse denoising process is a parameterized Markov chain that starts from standard normal noise $y_t \sim N(0,1)$ and iteratively denoises it, conditioned on the predictor vector x :

$$p_\theta(y_{t-1}|y_t, x) = N\left(y_{t-1}; \mu_\theta(y_t, t, x), \tilde{\beta}_t\right), \tag{3.3}$$

where $\tilde{\beta}_t = \frac{1-\hat{a}_{t-1}}{1-\hat{a}_t}\beta_t$. The mean function μ_θ is parameterized by a neural network with parameters θ. A common parameterization, following Ho et al. [12], predicts the noise component (ϵ) added at step t

$$\mu_\theta(y_t, t, x) = \frac{1}{\sqrt{a_t}}\left(y_t - \frac{\beta_t}{\sqrt{1-\hat{a}_t}}\epsilon_\theta(y_t, t, x)\right). \tag{3.4}$$

Thus, the network $\epsilon_\theta : \mathbb{R} \times \mathbb{Z}^+ \times \mathbb{R}^p \to \mathbb{R}$ takes the noisy response y_t the time step index t, and the feature vector x as input, and outputs an estimate of the noise. For scalar regression, this network can be a simple multi-layer perceptron (MLP).

The joint distribution implied by the reverse process is:

$$p_\theta(y_{0:T}|x) = p(y_T) \prod_{t=1}^{T} p_\theta(y_{t-1}|y_t, x), \tag{3.5}$$

with $p(y_t) = N(y_t; 0, 1)$. The marginal $p_\theta(y_0|x)$ is the model's prediction for the response given (x).

3.3. Robust Likelihood via Student-t Scale Mixture

Instead of using a simple Gaussian distribution for the final observation y_0, we employ a Student-t distribution to robustify the model against outliers. We leverage its representation as a scale mixture of normals, which facilitates conjugate Gibbs sampling. For each observation i, we introduce a latent precision variable $\lambda_i > 0$

$$y_i|x_i, \theta, \lambda_i, \sigma^2 \sim N\left(\mu_{\theta,0}(x_i), \sigma^2/\lambda_i\right), \tag{3.6}$$

$$\lambda_i \sim \text{ Gamma } \left(\frac{\nu}{2}, \frac{\nu}{2}\right), \tag{3.7}$$

Where $\mu_{\theta,0}(x_i)$ denotes the denoised mean at $t = 0$ obtained by running the reverse diffusion process from $(y_t \sim N(0, 1))$ conditioned on(x_i). The marginal distribution of y_i integrating out λ_i is Student-$t(\nu, \mu_{\theta,0}(x_i), \sigma^2)$. The degrees of freedom parameter $\nu > 0$ controls the heaviness of the tails; smaller values imply greater robustness. Here, (σ^2) is a global scale parameter.

3.4. Hierarchical Global-Local Prior on Diffusion Network Parameters

To prevent overfitting and enable adaptive feature selection from the high-dimensional input x, we place a hierarchical Horseshoe prior [14] on the weights (θ) of the noise-prediction network (ϵ_θ).

Let $\theta \in \mathbb{R}^K$ represent the flattened vector of all network weights and biases. The prior is structured as:

$$\begin{aligned} \theta_j|\tau,\ \phi_j\ \& \sim N\left(0, t^2\phi_j^2\right), \quad j = 1, \ldots, K, \\ \phi_j \sim \text{Half} - \text{Cauchy}(0, 1). \\ \tau \sim \text{Half} - \text{Cauchy}(0, 1). \end{aligned} \tag{3.8}$$

The parameter τ is the global shrinkage parameter; a small value shrinks all weights strongly toward zero. The parameters ϕ_j are local shrinkage parameters; they allow individual weights to escape the global shrinkage if the data provide sufficient evidence. This prior is known to have excellent theoretical properties for sparse signal recovery, placing substantial mass near zero while maintaining heavy tails to avoid over-shrinking true signals.

3.5. Hyperpriors and Complete Hierarchical Model

We complete the model with weakly informative priors on the remaining hyperparameters:

$$\begin{aligned} \sigma \sim \text{Half} - \text{Cauchy}(0, 1), \\ \nu \sim \text{Exponential}(0.1) + 2, \ \ (\text{ensuring } \nu > 2). \end{aligned} \tag{3.9}$$

The shift(+2) on the Exponential prior for (ν) ensures the variance of the Student-t distribution is defined. The complete generative process for a single observation (x_i, y_i) is:

1. Draw global shrinkage $\tau \sim \text{Half} - \text{Cauchy } (0, 1)$.
2. For each network weight $j = 1, \ldots, K$, draw local shrinkage $\phi_j \sim \text{Half} - \text{Cauchy } (0, 1)$.
3. For each network weight j, draw $\theta_j \sim N(0, \tau^2\phi_j^2)$.

4. Draw scale parameter sigma $\sim$ Half $-$ Cauchy $(0, 1)$.

5. Draw degrees of freedom $\nu \sim$ Exponential $(0, 1) + 2$,

6. Draw latent precision $\lambda_i \sim$ Gamma $(\nu/2, \nu/2)$.

7. Conditioned on x_i and θ, run the reverse diffusion process (Equation (3.5)) to obtain the denoised mean $\mu_{\theta,0}(x_i)$

8. Draw the response $y_i \sim N(\mu_{\theta,0}(x_i), \sigma^2/\lambda_i)$.

The joint posterior distribution of all unknown parameters $\Theta = \{\theta, \tau, \{\phi_j\}, \sigma^2, \mu, \{\lambda_i\}\}$ given the data $D = \{x, y\}$ is:

$$\begin{aligned} p(\Theta|D) \propto & \left[\prod_{i=1}^{N} N\left(y_i|\mu_{\theta,0}(x_i), \sigma^2/\lambda_i\right) \text{ Gamma } (\lambda_i|\nu/2, \nu/2)\right] \\ & \times \left[\prod_{j=1}^{K} N(\theta_j|0, \tau^2\phi_j^2) \text{ Half} - \text{Cauchy } (\phi_j|0, 1)\right] \\ & \times \text{Half} - \text{Cauchy } (\tau|0, 1) \times \text{Half} - \text{Cauchy } (\sigma|0, 1) \times p(\nu). \end{aligned} \tag{3.10}$$

4. Posterior Inference via Hybrid Gibbs-HMC

Direct sampling from the posterior in Equation (3.10) is intractable due to the complex dependence of $\mu_{\theta,0}(x_i)$ on θ through the diffusion process. We design a hybrid Markov Chain Monte Carlo (MCMC) algorithm that exploits conditional conjugacy where possible and uses Hamiltonian Monte Carlo (HMC) for the remaining non-conjugate blocks.

4.1. Gibbs Sampling Steps

Conditional on the network parameters θ and the hyperparameters, the full conditional distributions for $\{\lambda_i\}, \sigma^2$, and parts of the prior hierarchy are standard.

1. Sampling the latent precisions λ_i:

$$\lambda_i|\cdots \sim \text{Gamma } \left(\frac{\nu+1}{2}, \frac{\nu + (y_i - \mu_{\theta,0}(x_i))^2/\sigma^2}{2}\right), \quad i = 1, \ldots, n. \tag{4.1}$$

2. Sampling the scale parameter (σ^2) : Define the weighted residual sum of squares $RSS_\lambda = \sum_{i=1}^{n} \lambda_i (y_i - \mu_{\theta,0}(x_i))^2$. The conditional posterior is:

$$\sigma^2|\cdots \sim \text{Inverse} - \text{Gamma } \left(\frac{n}{2}, \frac{RSS_\lambda}{2}\right). \tag{4.2}$$

3. Sampling the degrees of freedom (ν): The conditional for ν is non-standard due to its role in the Gamma prior for λ_i. We use a Metropolis-Hastings step with a symmetric log-normal random walk proposal.

4.2. Hamiltonian Monte Carlo for (θ, τ, ϕ)

The parameters θ, τ, and $\phi = \{\phi_j\}$ have non-conjugate full conditionals due to the complex dependence in the diffusion mean $\mu_{\theta,0}(x_i)$. We employ Hamiltonian Monte Carlo to sample this high-dimensional block jointly.

The negative log-posterior (potential energy) $U(\theta, \tau, \phi)$ for this block, up to a constant, is:

$$\begin{aligned} U(\theta,\tau,\phi) =& \frac{1}{2\sigma^2}\sum_{i=1}^{N}\lambda_i(y_i-\mu_{\theta,0}(x_i))^2 \\ &+\frac{1}{2}\sum_{j=1}^{K}\left[\frac{\theta_j^2}{\tau^2\phi_j^2}+\log(\tau^2\phi_j^2)\right] \\ &+\sum_{j=1}^{K}\log(1+\phi_j^2)+\log(1+\tau^2). \end{aligned} \tag{4.3}$$

The key challenge is computing the gradient $\nabla_\theta U$, which requires backpropagation through the entire reverse diffusion process (or its simplified training objective). In practice, we use the denoising score matching objective [12]. For a given mini-batch, we sample a random timestep $t \sim \text{Uniform } \{1, \ldots, T\})$, compute the noisy target $y_t^{(i)}$ from $y_0^{(i)} = y_i$ using Equation (3.2), and define a stochastic estimate of the gradient:

$$\nabla_\theta \widehat{U} \approx \frac{|\mathcal{B}|}{N}\sum_{i\in\mathcal{B}}\lambda_i\left(\epsilon_\theta(y_t^{(i)},t,x_i)-\epsilon^{(i)}\right)\nabla_\theta\epsilon_\theta+\nabla_\theta\left(\frac{1}{2}\sum_j\frac{\theta_j^2}{\tau^2\phi_j^2}\right), \tag{4.4}$$

where $(\mathcal{B})$ is a mini-batch of indices, and $\epsilon^{(i)}$ is the true noise used to generate $y_t^{(i)}$ from y_i. Gradients for τ and ϕ_j are computed similarly, involving terms from their log-priors.

We use the No-U-Turn Sampler (NUTS) [30], an adaptive variant of HMC, to automatically tune the step size and number of leapfrog steps. For computational efficiency, we update the (θ, τ, ϕ) block every (M) iterations (e.g., M=5), while performing Gibbs updates for (λ, σ^2, ν) more frequently.

4.3. Predictive Distribution

After obtaining S posterior samples $(\{\Theta^{(s)}\}_{s=1}^{S})$, the posterior predictive distribution for a new feature vector x^* is approximated via Monte Carlo:

$$p(y^*|x^*,D) \approx \frac{1}{S}\sum_{s=1}^{S}p(y^*|x^*,T^{(s)}). \tag{4.5}$$

For each sample $\Theta^{(s)}$, drawing from $p(y^*|x^*, T^{(s)})$ involves: (1) drawing $\lambda^* \sim \text{Gamma}(\nu^{(s)}/2, \nu^{(s)}/2)$ (2) running the reverse diffusion process conditioned on x^* and $\theta^{(s)}$ to get $\mu_{\theta,0}^{(s)}(x^*)$, and (3) drawing $y^* \sim N(\mu_{\theta,0}^{(s)}(x^*), (\sigma^2)^{(s)}/\lambda^*)$.

Predictive intervals are constructed from the empirical quantiles of these samples.

4.4. Algorithm Summary

Algorithm 1: Hybrid Gibbs-HMC for BCDR

1. Initialize parameters $(\Theta^{(0)})$.
2. For iteration $(m = 1)$ to (M_{total})
   A. Gibbs Step: Sample $(\{\lambda_i^{(m)}\})$ from Equation (4.1).
   B. Gibbs Step: Sample $(\sigma^2)^{(m)}$ from Equation (4.2).
   C. MH Step: Sample $\nu^{(m)}$ using a Metropolis step.
   D. If $(m \bmod M = 0)$: Use NUTS (HMC) to jointly sample$(\theta, \tau, \phi)^{(m)})$ targeting Equation (4.3), using gradient estimate Equation (4.4).
   E. Else: Set $(\theta, \tau, \phi)^{(m)} = (\theta, \tau, \phi)^{(m-1)}$.
3. Discard burn-in samples and thin the chain.
4. Output: Posterior samples $\{\Theta^{(s)}\}$.

5. Experimental Design

5.1. Datasets

We evaluate BCDR on one synthetic and one real-world agricultural dataset. This study mainly uses the Pomegranate Fruit Dataset: https://www.kaggle.com/datasets/kumararun37/pomegranate-fruit-dataset,

5.1.1. Synthetic Data. We generate data to mimic the challenges of high-dimensional, collinear, and noisy agricultural features.

Feature Generation: $x \in \mathbb{R}^{n \times p}$ with $n = 500$, $p = 100$. Rows $x_i \sim N_p(0,\Sigma)$,

where $\Sigma_{jk} = \rho^{|j-k|}$(Toeplitz/AR(1)structure) withρ= 0.7 to induce multicollinearity.

TrueCoefficient:

Only the first 10 coefficients are non-zero: β_j^*= 1.5 for $j = 1, \ldots, 5; \beta_j^* = -1.0$ for $j = 6, \ldots, 10$; β_j^*= 0 for $j > 10$.

Response Generation: $y_i = x_i^t \beta^* + \epsilon_i$.

Scenario A (Gaussian): $\epsilon_i \sim N(0, 1)$.

Scenario B (Student-t): ϵ_i $Student - t(\nu = 3, 0, 1)$.

Scenario C (Contaminated): 80% of $\epsilon_i \sim N(0, 1), 20\%$(randomly chosen)$(\epsilon_i \sim N(0,5^2)$.

5.1.2. Pomegranate Image-Yield Dataset. We construct a real dataset by linking image features to physical measurements.

Image Source: We use the pomegranate image dataset from Zhang et al. [36], which includes original and completed images of pomegranates under occlusion.

Feature Extraction: For each fruit image, we extract a 512-dimensional feature vector using a pre-trained VGG-16 network [37], specifically the activations from the final fully connected layer (fc1). Images are resized to (224×224)as input.

Response Variable: We synthetically pair these features with realistic fruit weights (in grams). Weights are generated using a ground-truth linear model on the first 20 principal components of the features, plus heavy-tailed noise, ensuring a known but complex relationship. The final dataset contains n=1500 samples, p=512.

Split: 70% training, 15% validation, 15% testing.

5.2. Compared Methods

We compare BCDR against a range of established and state-of-the-art regression methods:

1. OLS: Ordinary Least Squares.
2. Ridge: L2-penalized regression with penalty chosen via 10-fold CV.
3. Lasso: L1-penalized regression with penalty chosen via 10-fold CV.
4. Bayesian Lasso (BL): The standard Bayesian Lasso with Laplace prior [20], implemented with Gibbs sampling.
5. Robust Bayesian Regression (RBR): A Bayesian linear model with a Student-t likelihood and independent Gaussian priors on coefficients [19], using Gibbs sampling.
6. Feedforward Neural Network (NN): A 3-layer MLP with ReLU activations and dropout, tuned on validation loss.
7. Gaussian Process Regression (GP): With a radial basis function (RBF) kernel, optimized via marginal likelihood maximization.
8. CDR (Ours): The proposed model with T=100 diffusion steps, a 3-layer MLP for(ϵ_θ), and the Horseshoe prior.

5.3. Evaluation Metrics

Performance is assessed using the following metrics on the held-out test set:

1. Prediction Accuracy:
   - Root Mean Squared Error (RMSE)
   - Mean Absolute Error (MAE)
   - Coefficient of Determination ($\mathbb{R}^2$)
2. Uncertainty Quantification:

   Coverage: Empirical coverage of 90% and 95% posterior predictive intervals.

   Interval Width: Average width of the 95% predictive interval.

   Calibration Plot: Observed vs. predicted cumulative probabilities.
3. Model Robustness: Percentage increase in RMSE when moving from the clean (Scenario A) to the contaminated (Scenario C) synthetic dataset.
4. Sparsity Recovery (Synthetic Data):

   False Discovery Rate (FDR): Proportion of selected (95% CI excludes zero) variables that are truly zero.

   True Positive Rate (TPR): Proportion of truly non-zero variables that are selected.
5. Computational Efficiency: Wall-clock time per 1000 MCMC iterations (for Bayesian methods) and total training time.

5.4. Implementation Details

BCDR: Implemented in PyTorch. The (ϵ_θ) network is an MLP with layers [p, 128, 128, 1]. The HMC step uses the 'pyro' library's NUTS implementation. T=100, linear noise schedule. Chain length: 10,000 iterations (5,000 burn-in). Mini-batch size for stochastic gradients: 128.
Baselines: Implemented using 'scikit-learn', 'statsmodels', and 'PyMC3' libraries.
Hardware: All experiments run on a system with an Intel i9-13900K CPU and an NVIDIA RTX 4090 GPU.

6. Results and Discussion

6.1. Synthetic Data Results

Table 1 shows BCDR achieves the best or competitive RMSE/MAE across all scenarios. Its superiority is most pronounced under heavy-tailed (Scenario B) and contaminated (Scenario C) noise, highlighting its robustness. RBR also performs well under heavy tails, but BCDR's generative flexibility gives it an edge under complex contamination.

As shown in Table 2, BCDR leverages the Horseshoe prior to recover all relevant predictors while producing the smallest proportion of false selections, successfully isolating the ten true signals from a set of one hundred candidate variables.

6.2. Pomegranate Yield Prediction Results

Results reported in Table 3 indicate that BCDR delivers the strongest overall performance on the real-world dataset. The method records the smallest prediction error (RMSE of 21.7 g), explains the largest proportion of outcome variability ($\mathbb{R}^2 = 0.87$), and produces prediction intervals that are both well calibrated and relatively narrow, achieving 94% coverage with an average width of 79.8 g.

Table 1: Prediction Performance (RMSE, MAE) on Synthetic Test Data (n=150).

Method	Scenario A(Gaussian)		Scenario B (Student-t)		Scenario C (Contaminated)	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
OLS	1.02	0.81	2.51	1.85	4.33	3.12
Ridge	0.98	0.78	2.15	1.62	3.87	2.85
Lasso	0.95	0.75	2.02	1.54	3.65	2.71
BL	0.96	0.76	1.97	1.50	3.41	2.52
RBR	0.94	0.74	1.61	1.25	2.15	1.68
NN	0.91	0.72	1.89	1.48	2.58	2.01
GP	0.93	0.73	195	1.52	3.10	2.40
BCDR	0.92	0.73	1.58	1.23	1.72	1.35

Table 2: Sparsity Recovery Metrics on Synthetic Data (Scenario A).

Method	**TPR**	**FDR**	**Num. Selected**
Lasso	1.00	0.53	21.4
BL	0.95	0.45	18.2
RBR	0.98	0.60	25.0
BCDR	1.00	0.32	14.7

Table 3: Performance on Pomegranate Image-Yield Test Set.

Method	RMSE (g)	MAE (g)	$\mathbb{R}^2$	95% Cov	Avg. Width (g)
Ridge	28.4	22.1	0.78	-	-
Lasso	26.8	20.8	0.8	-	-
BL	259	20.1	0.82	0.88	98.5
RBR	24.1	189	0.84	0.91	85.2
NN	23.5	18.5	0.85	-	-
GP	26.2	20.5	0.81	0.93	102.3
BCDR	21.7	17.1	087	0.94	79.8

6.3. Computational Performance

The computational cost of Bayesian Conditional Diffusion Regression was evaluated relative to competing regression methods. As expected, BCDR requires greater computational effort than classical linear and penalized models due to the iterative diffusion process and the use of Markov chain Monte Carlo sampling for posterior inference [9,30]. In particular, Hamiltonian Monte Carlo updates of large parameter blocks cause high overhead on training time.

Table 4: Computational Cost (Training Time).

Method	**Training Time (mins)**	**Inference Time per sample (ms)**
OLS/Ridge/Lasso	< 0.1	< 0 .1
BL	2.5	1.2
RBR	3.1	1.5
NN	5.1	0.5
GP	8.2	15.0
BCDR	45.0	8.5

Rate, urgency of training, and cost trade–offs Nevertheless with this overhead the observed computational cost is acceptable for medium-sized agricultural datasets where data-gathering is expensive and model reliability is more critical than fast training. Prediction time per test instance is also reasonable for

off-line analysis and decision-support. It contrasted to Gaussian process regression, that between matching the likelihood and computational efficiency-char-acteristics of DBD; instead, BCDR tends to provide a better match between those two aspects (in high-dimensional cases). 27, 33]; however in high-dimension environments.

6.4. Discussion of Results

The experimental results demonstrate several advantages of the proposed framework. First, a heavy-tailed likelihood in conjunction with conditional response generation ensures strong robustness to non-Gaussian noise and contaminated observations. This is reflected in the maintainability little amount of predictive power under a severe noisy condition compared with their classical regression and neural network counterparts [8,10].

Second, the hierarchical global–local shrinkage prior is a successful and appropriate approach to high-dimensional feature spaces by putting irrelevant predictors into suppression mode without incurring signal loss. This adaptive behaviour enable the model to gather useful information from noisy image-feature without depending on manual feature selection or over-eager dimensionality reduction [14,21].

Third, the Bayesian formulation provides well calibrated and informative predictive intervals, enabling accurate uncertainty quantification in agricultural yield prediction task [9]. Finally, since regression is framed as a conditional generation process, the proposed model captures non-linear image feature-to-yield relationships via its enhanced performance.over linear approaches and competitive results with respect to more flexible machine-learningapproaches.

7. Conclusion

7.1. Limitations

The only limitation of the proposed framework is its computational complexity. The Eulerian model-based (to mimic events themselves) strategy and HMC inference approach are computationally intensive, limiting its use to very large data with very high dimension under current computational hardware constraints. Furthermore, the model contains several meta-parameters, e.g., the number of diffusion steps and inference configuration), which are needed to be carefully fine-tuned to achieve good performance.

7.2. Future Research Directions

The main drawback of the proposed model is the computational complexity. The diffusion-based modeling approach together with Hamiltonian Monte Carlo inference have sizable training overheads that could limit its applicability to very large data sets of much higher dimensionality than are currently hardware-limited. Furthermore, the model contains a number of hyperparameters (e.g., the number of diffusion steps, settings for inference) which need to be carefully tuned for the best possible performance.

References

1. Li, J., Tang, L. and Wang, Y., *High-throughput plant phenotyping using computer vision*, Plant Methods 16, 115, (2020).
2. Singh, A. et al., *Machine learning for high-throughput stress phenotyping in plants*, Trends Plant Sci. 21, 110–124, (2016).
3. Sankaran, S., Khot, L. R. and Ehsani, R., *Visible–near infrared spectroscopy for nondestructive evaluation of apple quality*, Trans. ASABE 54, 1927–1934, (2011).
4. van Wijk, M. T., *Challenges and prospects for agricultural carbon sink research*, Front. Agric. Sci. Eng. 8, 1–10, (2021).
5. Seber, G. A. F. and Lee, A. J., *Linear Regression Analysis*, Wiley, Hoboken, NJ, (2003).
6. Hoerl, A. E. and Kennard, R. W., *Ridge regression: biased estimation for nonorthogonal problems*, Technometrics 12, 55–67, (1970).
7. Tibshirani, R., *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B 58, 267–288, (1996).
8. Huber, P. J. and Ronchetti, E. M., *Robust Statistics*, Wiley, Hoboken, NJ, (2009).
9. Gelman, A. et al., *Bayesian Data Analysis*, Chapman & Hall/CRC, Boca Raton, FL, (2013).
10. Lange, K. L., Little, R. J. A. and Taylor, J. M. G., *Robust statistical modeling using the t distribution*, J. Amer. Statist. Assoc. 84, 881–896, (1989).

11. Mitchell, T. J. and Beauchamp, J. J., *Bayesian variable selection in linear regression*, J. Amer. Statist. Assoc. 83, 1023–1032, (1988).

12. Ho, J., Jain, A. and Abbeel, P., *Denoising diffusion probabilistic models*, Adv. Neural Inf. Process. Syst. 33, 6840–6851, (2020).

13. Dhariwal, P. and Nichol, A., *Diffusion models beat GANs on image synthesis*, Adv. Neural Inf. Process. Syst. 34, 8780–8794, (2021).

14. Carvalho, C. M., Polson, N. G. and Scott, J. G., *The horseshoe estimator for sparse signals*, Biometrika 97, 465–480, (2010).

15. Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A., *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York, NY, (1986).

16. Dunson, D. B. and Park, J. H., *Kernel stick-breaking processes*, Biometrika 95, 307–323, (2008).

17. Fernández, C. and Steel, M. F. J., *Multivariate Student-t regression models: pitfalls and inference*, Biometrika 86, 153–167, (1999).

18. Lange et al., "Robust statistical modeling...," 1989.

19. Kumar, C. S., *A robust Bayesian regression framework for high-dimensional data using adaptive prior structuring and noise-aware likelihood modelling*, (2025).

20. Park, T. and Casella, G., *The Bayesian lasso*, J. Amer. Statist. Assoc. 103, 681–686, (2008).

21. Bhattacharya, A., Pati, D., Pillai, N. S. and Dunson, D. B., *Dirichlet–Laplace priors for optimal shrinkage*, J. Amer. Statist. Assoc. 110, 1479–1490, (2015).

22. Zhang, Y., Dai, B. and Wang, Z., *The R2–D2 prior for Bayesian linear regression with sparse signals*, J. Amer. Statist. Assoc. 117, 274–286, (2022).

23. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N. and Ganguli, S., *Deep unsupervised learning using nonequilibrium thermodynamics*, Proc. Int. Conf. Mach. Learn., 2256–2265, (2015).

24. Goodfellow, I. et al., *Generative adversarial nets*, Adv. Neural Inf. Process. Syst. 27, (2014).

25. Nichol, A. Q. and Dhariwal, P., *Improved denoising diffusion probabilistic models*, Proc. Int. Conf. Mach. Learn., 8162–8171, (2021).

26. Saharia, C. et al., *Palette: image-to-image diffusion models*, ACM SIGGRAPH Conf. Proc., (2022).

27. Gal, Y., *Uncertainty in Deep Learning*, Cambridge Univ. Press, Cambridge, UK, (2016).

28. Kingma, D. P. and Welling, M., *Auto-encoding variational Bayes*, Int. Conf. Learn. Represent., (2014).

29. Neal, R. M., *Bayesian Learning for Neural Networks*, Springer, New York, NY, (1996).

30. Hoffman, M. D. and Gelman, A., *The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo*, J. Mach. Learn. Res. 15, 1593–1623, (2014).

31. Mohanty, S. P., Hughes, D. P. and Salathé, M., *Using deep learning for image-based plant disease detection*, Front. Plant Sci. 7, 1419, (2016).

32. Rocha, D. R., Barbedo, J. G. A. and Koenigkan, L. V., *Image-based fruit counting: a review*, Comput. Electron. Agric. 189, 106383, (2021).

33. Khot, L. R., Sankaran, S., Carter, J. M. and Ehsani, R., *UAV-based remote sensing for tree crop yield prediction*, Trans. ASABE 59, 547–555, (2016).

34. Wang, H. Q., Tang, W. H. and Chen, S. Y., *A deep learning approach for crop yield prediction based on remote sensing data*, IEEE Access 8, 163925–163937, (2020).

35. Kamilaris, A. and Prenafeta-Boldú, F. X., *Deep learning in agriculture: a survey*, Comput. Electron. Agric. 147, 70–90, (2018).

36. Zhang, D. et al., *A conditional segmentation-guided network for pomegranate image completion under occlusion*, Plant Methods 21, 153, (2025).

37. Simonyan, K. and Zisserman, A., *Very deep convolutional networks for large-scale image recognition*, Int. Conf. Learn. Represent., (2015).

Ch. Shashi Kumar,
Department of Mathematics,
VNR Vignana Jyothi Institute of Engineering and Technology,
Hyderabad-500090, Telangana, India.
E-mail address: skch17@gmail.com

and

G. Upender Reddy,
Department of Mathematics,
Univeristy college of Engineering(A), Osmania Univeristy,
Hyderabad-500007, Telangana, India.
E-mail address: yuviganga@gmail.com

and

K. Varalakshmi,
Department of Mathematics,
MJPTBCWR Degree College (W),
Wargal, Siddipet-502279, Telangana, India.
E-mail address: varak2121@gmail.com