BAREKENG: Journal of Mathematics and Its Applications
December 2026 Volume 20 Issue 4 Page 3321-3338
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss4pp3321-3338

MODIFYING MIXED MODEL NEURAL NETWORKS FOR UNIT-LEVEL PROPORTION SMALL AREA ESTIMATION

Ade Riyawan [1*], Muhammad Nur Aidi [2], Budi Susetyo [3], Anang Kurnia [4]

[1,2,3,4] *School of Data Science, Mathematics, and Informatics, IPB University*
Jln. Dramaga, Bogor, 16680, Indonesia

Corresponding author's e-mail: * *ade0717riyawan@apps.ipb.ac.id*

Article Info

Article History:
Received: 17th October 2025
Revised: 03rd May 2026
Accepted: 02nd June 2026
Available online: 01st July 2026

Keywords:
Complexity;
Deep learning with random effects;
Nonlinear covariate relationships;
Statistical learning;
Unit level estimation.

ABSTRACT

This paper develops a Small Area Estimation Mixed Model Neural Network (SAE-MNN) for unit-level estimation of area-level proportions under clustered binary responses. The method is motivated by a limitation of conventional generalized linear mixed model (GLMM)-based small area estimation, namely the linear specification of the fixed-effects component on the latent scale, which may be inadequate when auxiliary variables exhibit nonlinear effects and higher-order interactions. The proposed model replaces the linear predictor with a neural-network-based nonlinear function while retaining area-level random effects within a likelihood-based mixed-model framework. Thus, SAE-MNN combines nonlinear function approximation with the borrowing-strength mechanism required for coherent small area inference. The method is evaluated through a controlled simulation study and a real-data application. The simulation study considers four scenarios generated by combining two predictor structures (linear and interaction-based nonlinear) with two levels of between-area heterogeneity (small and large), and compares SAE-MNN with a conventional GLMM and a standard deep neural network (DNN). SAE-MNN performs robustly across all scenarios and is especially effective when nonlinear interaction and substantial area-level heterogeneity coexist. In the most complex scenario, SAE-MNN attains the lowest median RRMSE (8.389%), compared with 11.474% for GLMM and 18.826% for DNN, while maintaining empirical coverage between 0.89 and 0.94. In the real-data application, using KSA segment-level harvest proportions to estimate subdistrict-level monthly harvest proportions, SAE-MNN shows the closest overall agreement with the direct estimator, reproduces the main temporal pattern more faithfully than GLMM, and yields moderate shrinkage without excessive smoothing. Overall, the results indicate that SAE-MNN provides a principled compromise between the rigidity of GLMM and the purely predictive character of DNN, thereby extending hybrid small area estimation methodology to binary and proportional outcomes in data-rich settings with nonlinear auxiliary information and area-level heterogeneity.

How to cite this article:

A. Riyawan, M. N. Aidi, B. Susetyo and A. Kurnia., "MODIFYING MIXED MODEL NEURAL NETWORKS FOR UNIT-LEVEL PROPORTION SMALL AREA ESTIMATION", *BAREKENG: J. Math. & App.*, vol. 20, no. 4, pp. 3321-3338, Dec, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Estimating reliable statistics for small geographic or administrative domains is a central task in official statistics and many applied fields, because policy and planning decisions increasingly require information at disaggregated levels. Small Area Estimation (SAE) has been developed to improve the precision of domain-specific estimates when direct survey estimators are unreliable due to limited sample sizes [1], [2], [3], [4]. In practice, SAE combines survey data with auxiliary information obtained from census, administrative, or geospatial sources in order to borrow strength across areas and increase estimation efficiency. Classical SAE methodology has been particularly successful for continuous outcomes through area-level and unit-level mixed models, such as the Fay–Herriot and Battese–Harter–Fuller frameworks [1], [4], [5].

When the response of interest is binary or proportional, however, the modeling problem becomes more challenging. A common approach is to use generalized linear mixed models (GLMMs), especially logistic mixed models, which allow domain-specific random effects and provide a coherent framework for empirical best prediction and uncertainty estimation [6], [7], [8], [9]. These models are attractive because they preserve the core inferential principles of SAE, namely borrowing strength across areas and accounting for hierarchical variability [10], [11], [12]. Nevertheless, GLMM-based SAE relies on a restrictive specification of the fixed-effects component, which is typically linear on the latent scale [13], [14], [15]. In many contemporary applications, especially those involving remote sensing or high-dimensional auxiliary data, such an assumption may be too rigid. Complex predictor structures, nonlinear effects, and higher-order interactions may not be adequately captured by a conventional linear predictor, which can lead to model misspecification, degraded predictive performance, and unreliable inference [16], [17].

To overcome this limitation, recent studies have explored machine learning approaches in SAE. Tree-based methods, random forests, and other predictive models offer greater flexibility and reduced sensitivity to parametric misspecification [18], [19], [20], [21]. In particular, mixed-effects random forest approaches such as MERF and its extensions demonstrate that combining nonlinear learners with hierarchical random effects can improve estimation in small area contexts [20], [22], [23], [24]. However, most existing machine-learning-based SAE approaches have focused either on continuous responses or on tree-based methods, while standard deep neural networks, when used directly, typically ignore the clustering structure induced by areas and therefore do not naturally preserve the inferential borrowing-strength mechanism that is fundamental in SAE [19], [24], [25], [26]. As a result, purely predictive deep learning models may capture nonlinear covariate patterns well, but they may also yield unstable domain-level predictions and underestimated uncertainty when hierarchical dependence is present.

A related line of work has emerged in the broader literature on mixed-effects neural networks. Several studies have shown that neural networks can be integrated with random-effects structures to model clustered or correlated data more effectively [27], [28]. These developments indicate that neural networks need not be restricted to independent observations, and that mixed-model principles can be embedded within likelihood-based learning frameworks. Among these, the Linear Mixed Model Neural Network (LMMNN) proposed by Simchoni and Rosset provides an important conceptual foundation by showing how random effects can be incorporated into neural-network-based models through a likelihood-based objective [29]. Likewise, MeNets and Bayesian deep GLMM-type approaches illustrate the broader feasibility of combining nonlinear representation learning with hierarchical modeling [28], [30]. Despite this progress, the literature still lacks a framework specifically tailored to unit-level small area estimation of proportions that simultaneously addresses nonlinear covariate–response relationships, area-level random effects, and SAE-oriented prediction and uncertainty evaluation.

This gap motivates the present study. The problem addressed in this paper is not simply one of prediction accuracy, but of developing a model that remains compatible with the inferential logic of SAE while being flexible enough to handle nonlinear auxiliary information. In the context of harvested-area estimation using satellite-derived predictors, for example, the relationship between vegetation indicators and the probability of harvest is unlikely to be adequately represented by a purely linear predictor. At the same time, observations within the same small area are not independent, since they may share unobserved local characteristics such as soil conditions, irrigation patterns, or management practices. Therefore, a suitable model for this setting must combine nonlinear learning capability with hierarchical variance modeling.

The contribution of this paper is threefold. First, we propose a modified Small Area Estimation Mixed Model Neural Network (SAE-MNN) for unit-level proportion estimation, in which the fixed-effects component of a logistic mixed model is replaced by a neural-network function while the area-specific random effects are retained within a likelihood-based framework. Second, we adapt this formulation to the SAE setting by presenting the corresponding prediction strategy for area-level proportions, including empirical best prediction and synthetic prediction under the proposed model. Third, we evaluate the method through a controlled simulation study designed to assess its performance under varying levels of between-area heterogeneity and nonlinear covariate interaction, and compare it with benchmark alternatives represented by a conventional GLMM-based SAE approach and a standard deep neural network without random effects [9], [20], [24], [29].

It is important to note that the present study is intended primarily as a methodological investigation rather than as an empirical case study. The main objective is to evaluate whether the proposed SAE-MNN framework can appropriately capture and exploit nonlinear covariate–response relationships together with random-effects-driven hierarchical dependence, which constitute the main motivation for its development. In this context, a controlled simulation design provides a more suitable basis for methodological assessment than a real-data application, because the underlying data-generating mechanism in empirical datasets is unknown and cannot be verified with certainty. By contrast, simulation enables the systematic control of nonlinearity, between-area heterogeneity, and sample-size imbalance, which are central features in small area estimation [20], [24]. For this reason, the present paper focuses on a carefully structured simulation study as a first step in establishing the methodological validity of the proposed approach.

The remainder of this paper is organized as follows. Section 2 reviews the conventional unit-level logistic mixed model and develops the proposed SAE-MNN framework, including its likelihood-based estimation strategy, prediction procedures, and uncertainty assessment. Section 3 presents the simulation design and discusses the comparative performance of the proposed method against existing alternatives. The final section concludes with the main findings, limitations, and directions for future research.

2. RESEARCH METHODS

2.1 Conventional Unit-Level Logistic Mixed Model

Classical unit-level small area estimation for binary or proportional outcomes is commonly based on a logistic mixed model [8], [9]. Let y_{dj} denote the binary response for unit j in area d, with $d = 1, \dots, D$ and $j = 1, \dots, n_d$. Conditional on an area-specific random effect, the response is modeled as a Bernoulli variable,

$$y_{dj} \mid v_d \sim \text{Bernoulli}(p_{dj}), \qquad d = 1, \dots, D, \; j = 1, \dots, n_d. \tag{1}$$

The conventional logistic mixed model specifies

$$logit(p_{dj}) = \log\left(\frac{p_{dj}}{1-p_{dj}}\right) = x_{dj}^{\top}\beta + \phi v_d, \tag{2}$$

where x_{dj} is the vector of auxiliary variables, β is the vector of fixed-effect parameters, $\phi > 0$ is a scale parameter, and v_d is the area-specific random effect, commonly assumed to follow a standard normal or Gaussian distribution after appropriate scaling [8], [9]. The conditional Bernoulli probability can therefore be written as

$$p_{dj} = \frac{\exp\{x_{dj}^{\top}\beta + \phi v_d\}}{1 + \exp\{x_{dj}^{\top}\beta + \phi v_d\}}. \tag{3}$$

This formulation is attractive because it preserves the hierarchical borrowing-strength mechanism central to SAE. However, its fixed-effects structure remains linear on the latent logit scale. In applications involving complex auxiliary information, such as spectral indices derived from satellite imagery, this linearity assumption may be too restrictive and may fail to capture nonlinear effects and interactions among predictors [16], [17], [24].

2.2 Proposed Method: Small Area Estimation Mixed Model Neural Network (SAE-MNN)

To address this limitation, we propose a modified Small Area Estimation Mixed Model Neural Network (SAE-MNN) for unit-level estimation of binary outcomes. The key idea is to replace the linear fixed-effects component in the logistic mixed model by a neural-network function while retaining the area-level random effects required for valid small area estimation.

Let y_{di} denote the binary response for unit i in area d. The proposed model is

$$y_{di} \sim Bernoulli(\pi_{di}),\ i = 1, \dots, n_d, \tag{4}$$

with

$$\text{logit}(\pi_{di}) = f_\theta(x_{di}) + u_d, \tag{5}$$

where $f_\theta(\cdot)$ is a feed-forward neural network parameterized by θ, x_{di} is the covariate vector, and $u_d \sim N(0, \sigma_u^2)$ is the area-specific random intercept. Under this formulation, the neural network flexibly captures nonlinear covariate-response relationships, while u_d accounts for unobserved between-area heterogeneity.

Thus, the unit-level success probability becomes

$$\pi_{di}(u_d) = \frac{\exp\{f_{\boldsymbol{\theta}}(x_{di}) + u_d\}}{1 + \exp\{f_{\boldsymbol{\theta}}(x_{di}) + u_d\}}. \tag{6}$$

Conditional on $\boldsymbol{u_d}$, the observations within area d are independent Bernoulli trials, so the conditional likelihood for area d is

$$L_d(\,y_d \mid \boldsymbol{u_d}\,) = \prod_{i=1}^{n_d} \pi_{di}(\boldsymbol{u_d})^{y_{di}}[1 - \pi_{di}(\boldsymbol{u_d})]^{1-y_{di}}. \tag{7}$$

Because u_d is latent, the marginal likelihood for area d is obtained by integrating over the random-effects distribution,

$$L_d(y_d; \theta, \sigma_u^2) = \int L_d(\,\boldsymbol{y_d} \mid u_d\,)\phi(u_d; 0, \sigma_u^2)du_d, \tag{8}$$

where $\phi(u_d; 0, \sigma_u^2)$ denotes the Gaussian density with mean 0 and variance σ_u^2. The full marginal log-likelihood is then

$$l(\theta, \sigma_u^2) = \sum_{d=1}^{D} \log L_d\,(y_d; \theta, \sigma_u^2) \tag{9}$$

The corresponding objective function is the negative log-likelihood,

$$\mathcal{L}(\theta, \sigma_u^2) = -\sum_{d=1}^{D} \log L_d\,(y_d; \theta, \sigma_u^2) \tag{10}$$

which is minimized with respect to the neural-network parameters θ and the variance component σ_u^2.

Since the integral in Eq. (8) does not admit a closed-form solution, we approximate it using Gauss-Hermite quadrature [8], [9]. Let $\{z_k, w_k\}_{k=1}^{K}$ denote the quadrature nodes and weights. Then the area-specific marginal likelihood is approximated by

$$L_d(y_d; \theta, \sigma_u^2) \approx \sum_{k=1}^{K} w_k \prod_{i=1}^{n_d} \pi_{di}\left(\sqrt{2}\sigma_u z_k\right)^{y_{di}}\left[1 - \pi_{di}\left(\sqrt{2}\sigma_u z_k\right)\right]^{1-y_{di}} \tag{11}$$

Equivalently, in log-likelihood form,

$$\mathrm{l}_d(\theta, \sigma_u^2) \approx \log\left[\sum_{k=1}^{K} w_k \exp\left\{\sum_{i=1}^{n_d}[y_{di}\eta_{dik} - \log(1 + \exp(\eta_{dik}))]\right\}\right], \tag{12}$$

where

$$\eta_{dik} = f_\theta(x_{di}) + \sqrt{2}\sigma_u z_k \tag{13}$$

In practice, the network is trained by minimizing Eq. (10) using a gradient-based optimizer such as Adam. In our implementation, $f_\theta(\cdot)$ is a feed-forward neural network with one or two hidden layers and ReLU activation, followed by a linear output layer producing the latent logit component. Regularization may be introduced through dropout and/or weight decay when necessary.

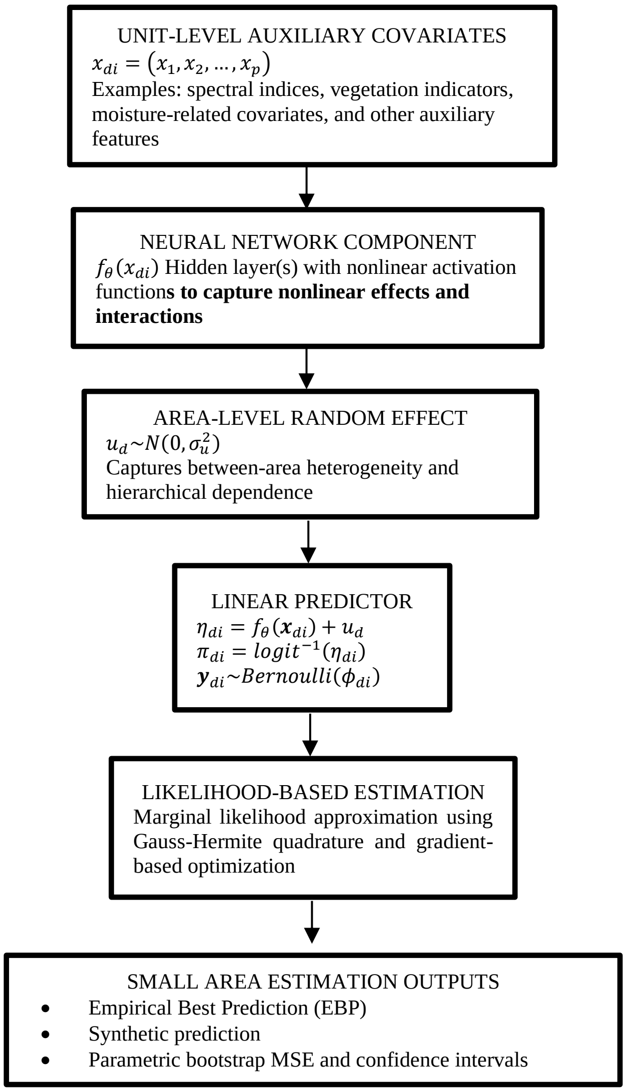


Figure 1. Schematic Diagram of the Proposed SAE-MNN Framework for Unit-Level Proportion Small Area Estimation. The Framework Combines a Neural-Network-Based Nonlinear Fixed-Effects Component with an Area-Level Random Effect to Model Clustered Binary Responses

Fig. 1 illustrates the overall structure of the proposed SAE-MNN framework. Unit-level auxiliary covariates are first processed through a neural-network component, $f_{\boldsymbol{\theta}}(\mathbf{x}_{di})$, which flexibly captures nonlinear effects and interactions among predictors. This nonlinear fixed-effects component is then combined with an area-level random effect, $u_d \sim N(0, \sigma_u^2)$, to form the linear predictor $\eta_{di} = f_{\boldsymbol{\theta}}(\boldsymbol{x}_{di}) + u_d$. The resulting predictor is transformed through the logistic link to obtain the unit-level probability π_{di} , from which the binary response is modeled. Estimation is carried out through a likelihood-based approach using Gauss-Hermite quadrature, and the fitted model is subsequently used for empirical best prediction, synthetic prediction, and bootstrap-based uncertainty estimation.

2.2.1 Prediction of Area-Level Proportions

Let N_d denote the population size of area d. The target small area proportion is

$$P_d = \frac{1}{N_d}\sum_{i=1}^{N_d} \pi_{di} \tag{14}$$

where π_{di} is the probability of success for unit i in area d. Since u_d is unobserved, the empirical best predictor of P_d is defined as the conditional expectation given the observed data,

$$\hat{P}_d^{EBP} = E(P_d \mid \mathbf{y_d}, \mathbf{X_d}) \tag{15}$$

Substituting the model formulation, this becomes

$$\hat{P}_d^{EBP} = \frac{1}{N_d}\sum_{i=1}^{N_d} \int \pi_{di}(u_d)\, p\big(u_d \mid \mathbf{y_d}, \mathbf{X_d}; \widehat{\boldsymbol{\theta}}, \hat{\sigma}_u^2\big) du_d \tag{16}$$

where $p\big(u_d \mid \mathbf{y_d}, \mathbf{X_d}; \widehat{\boldsymbol{\theta}}, \hat{\sigma}_u^2\big)$ is the posterior density of the random effect. Using Gauss-Hermite quadrature, Eq. (16) is approximated by

$$\hat{P}_d^{EBP} \approx \frac{1}{N_d}\sum_{\boldsymbol{i}=1}^{N_d} \sum_{\boldsymbol{k}=1}^{\boldsymbol{K}} \widetilde{w}_{dk}\pi_{di}(\hat{u}_{dk}) \tag{17}$$

where $\widetilde{w}_{dk}$ are normalized posterior quadrature weights and $\hat{u}_{dk}$ denotes the corresponding quadrature-based support points.

2.2.2 Uncertainty Estimation

A crucial component of SAE is the estimation of uncertainty associated with small area predictors. For the empirical best predictor, the mean squared error (MSE) is

$$MSE\big(\hat{P}_d^{EBP}\big) = E\left(\big(\hat{P}_d^{EBP} - P_d\big)^2\right)$$

Because a closed-form MSE expression is generally unavailable under the proposed model, we adopt a parametric bootstrap approach. For bootstrap replication $b = 1, \dots, B$, the following steps are performed:

1. Generate bootstrap random effects
$$u_d^{*(b)} \sim N(0, \hat{\sigma}_u^2). \tag{18}$$
2. Simulate bootstrap outcomes
$$y_{di}^{*(b)} \sim \text{Bernoulli}\left(\pi_{di}\left(u_d^{*(b)}\right)\right), \quad b = 1, \dots, B \tag{19}$$
3. Refit the model on each bootstrap sample to obtain $\hat{P}_d^{EBP*(b)}$.
4. Estimate the MSE as
$$\widehat{\text{MSE}}\big(\hat{P}_d^{EBP}\big) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{P}_d^{EBP*(b)} - P_d^{*(b)}\right)^2 \tag{20}$$
where $P_d^{*(b)}$ is the true proportion computed from the b-th bootstrap population.

This bootstrap estimator captures both sampling variation and model-estimation uncertainty, and is consistent with the model-based uncertainty assessment commonly used in SAE [8], [9], [20], [24].

2.3 Simulation Design

The simulation study was designed to evaluate the proposed SAE-MNN under controlled settings that reflect key challenges in small area estimation: between-area heterogeneity, nonlinear predictor-response relationships, and limited area-level information. Following the general simulation logic adopted in recent model-based SAE studies [20], [24], we considered four scenarios formed by the combination of two dimensions:

1. Predictor structure:
Normal: predictors enter without nonlinear interaction terms;
Interaction: predictors enter through nonlinear interaction structures.
2. Between-area heterogeneity:
Small: low random-effect variance ($\sigma_u^2 = 0.1$);
Large: high random-effect variance ($\sigma_u^2 = 1.0$).
3. For each area $d = 1, \dots, D$, an area random effect was generated as
$$u_d \sim N(0, \sigma_u^2) \tag{21}$$
Next, covariates x_1 and x_2 were generated independently from a standard normal distribution,
$$x_1, x_2 \sim N(0,1) \tag{22}$$
The latent predictor η and the success probability μ were then defined under four scenarios:
i. Normal-Small
$$\eta = 0.5 - 0.8x_1 - 0.6x_2 + u_d, \qquad \mu = \text{sigmoid}(\eta) \tag{23}$$
ii. Interaction-Small

$$\eta = 1 - x_1x_2 - 0.6x_1^2 + u_d, \qquad \mu = \text{sigmoid}(\eta) \tag{24}$$

iii. Normal-Large

$$\eta = 0.5 - 0.1x_1 - 0.2x_2 + u_d, \qquad \mu = \text{sigmoid}(\eta) \tag{25}$$

iv. Interaction-Large

$$\eta = 1 - x_1x_2 - 0.6x_1^2 + u_d, \qquad \mu = \text{sigmoid}(\eta) \tag{26}$$

Finally, the binary response was generated from

$$y_{di} \sim \text{Bernoulli}(\mu_{di}) \tag{27}$$

This design allows direct assessment of how different methods behave under linear versus nonlinear fixed-effect structures and under weak versus strong area-level heterogeneity. Such controlled simulation settings are particularly appropriate for a methodological study, since they make the nonlinear signal and hierarchical dependence explicitly verifiable.

Three competing estimators were compared:

1. GLMM-based SAE, using a conventional logistic mixed model with linear fixed effects,
2. Standard DNN, using a feed-forward neural network without random effects,
3. Proposed SAE-MNN, using a neural network with area-level random effects.

For fairness, the DNN and SAE-MNN models used the same basic feed-forward architecture, differing only in the inclusion of the random-effects structure and the likelihood-based training criterion.

2.4 Performance Measures

The performance of the competing methods was evaluated by comparing estimated area proportions with the true simulated area proportions. Let $\hat{\mu}_d^{(m)}$ denote the estimated proportion for area d in replication m, and let $\hat{\mu}_d^{(m)}$ denote the corresponding true proportion. Over M simulation replications, the following criteria were used.

The relative bias (RB) is defined as

$$RB_d = \frac{1}{M}\sum_{m=1}^{M}\left(\frac{\hat{\mu}_d^{(m)} - \mu_d^{(m)}}{\mu_d^{(m)}}\right), \tag{28}$$

The relative root mean squared error (RRMSE) is

$$RRMSE_d = \frac{\sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\hat{\mu}_d^{(m)} - \mu_d^{(m)}\right)^2}}{\frac{1}{M}\sum_{m=1}^{M}\mu_d^{(m)}} \tag{29}$$

To assess the quality of MSE estimation, we also considered the relative bias of the RMSE estimator,

$$RB\text{-}RMSE_d = \frac{\sqrt{\frac{1}{M}\sum_{m=1}^{M} MSE_{est_d}^{(m)}} - RMSE_{emp_d}}{RMSE_{emp_d}} \tag{30}$$

and the relative RMSE of the RMSE estimator,

$$RRMSE\text{-}RMSE_d = \frac{\sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\sqrt{MSE_{est_d}^{(m)}} - RMSE_{emp_d}\right)^2}}{RMSE_{emp_d}}. \tag{31}$$

In addition, uncertainty calibration was evaluated using empirical coverage probability and average confidence interval width, both computed over areas and replications. These performance measures are basically from [31], [32] and modified based on [8], [9], [20], [24] for simulation-based comparison of SAE estimators and uncertainty estimators.

2.5 Computational Complexity

The computational complexity of the proposed SAE-MNN is driven by three main components:

1. forward and backward propagation through the neural network,
2. Gauss-Hermite quadrature over the area-level random effects, and
3. bootstrap-based uncertainty estimation.

Let $N = \sum_{d=1}^{D} n_d$ denote the total number of observed units, K the number of quadrature points, and P the effective number of neural-network parameters. The cost of a single likelihood evaluation is approximately of order

$$O(NKP), \tag{32}$$

since each quadrature node requires evaluation of the network and the Bernoulli likelihood contributions over the observed units. The total training cost depends additionally on the number of optimization epochs. For uncertainty estimation, if B bootstrap replications are used, the total complexity increases approximately to

$$O(BNKP) \tag{33}$$

excluding constant factors associated with repeated optimization. Therefore, SAE-MNN is computationally more demanding than a conventional GLMM and substantially more demanding than a standard DNN without quadrature or bootstrap. However, this additional cost is the consequence of explicitly retaining the hierarchical likelihood structure required for valid small area inference. In the present study, the method is intended as a methodological proof-of-concept, and future work may consider more efficient approximation and optimization strategies for larger-scale applications.

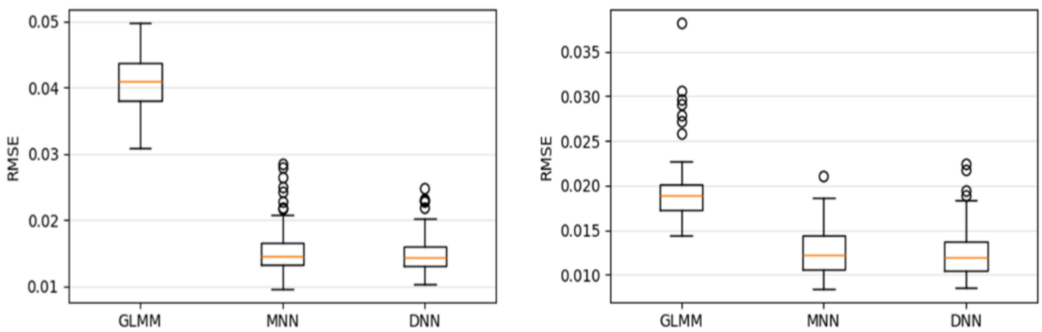
3. RESULTS AND DISCUSSION

3.1 Simulation Study Results

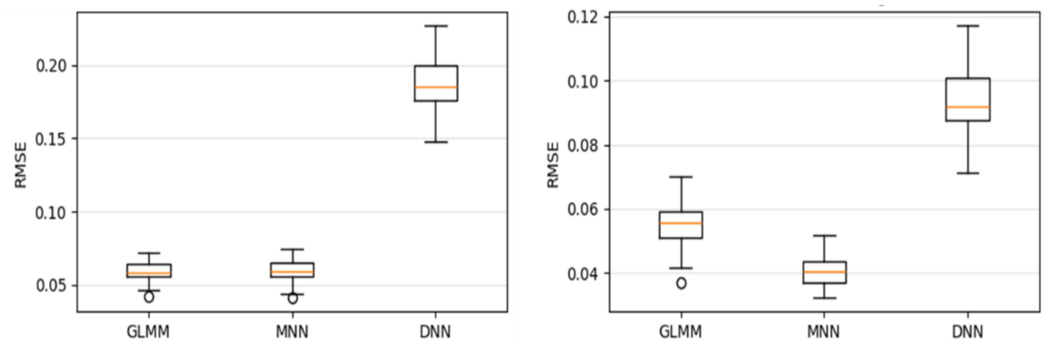
3.1.1 Point Estimation Performance

Fig. 2 compares the empirical RMSE of the competing methods for area-level proportion estimates under the four simulation scenarios. A clear pattern emerges across settings. Under Normal-Small, where the predictor structure is relatively simple and between-area heterogeneity is limited, both DNN and SAE-MNN outperform the conventional GLMM, with DNN showing a slight numerical advantage. A similar result is observed under Interaction-Small, where nonlinear interaction terms are present but the between-area variance remains low. In this case, SAE-MNN achieves the lowest RMSE, while DNN remains competitive and GLMM performs worst. These findings indicate that when the domain-level random effect is weak, modeling flexibility becomes the dominant factor, favoring neural-network-based methods over a strictly linear mixed model [20], [24], [29].

The pattern changes substantially when between-area heterogeneity increases. Under Normal-Large and Interaction-Large, where $\sigma_u^2 = 1.0$, SAE-MNN consistently provides the best point estimation performance, followed by GLMM, whereas DNN exhibits markedly higher RMSE. This result highlights the importance of explicitly incorporating area-level random effects when domain heterogeneity is substantial. In such settings, the ability to borrow strength across areas becomes essential, and a model that ignores hierarchical dependence loses stability. SAE-MNN performs well because it combines the representational flexibility of neural networks with the hierarchical variance structure required in SAE [8], [9], [24].



(a) (b)



(c) (d)

Figure 2. Empirical RMSE Comparison of Point Estimates for Area-Level Proportions under Four Simulation Scenarios (a) Normal–Small, (b) Interaction–Small, (c) Normal–Large, (d) Interaction–Large

Table 1 supports this interpretation through the area-level summaries of relative bias (RB) and relative root mean squared error (RRMSE). Under Normal-Small, GLMM shows substantial negative bias and the largest RRMSE, whereas DNN and SAE-MNN both achieve much smaller errors. Under Interaction-Small, GLMM again performs worst, reflecting its difficulty in accommodating nonlinear interactions within the linear predictor. In contrast, SAE-MNN and DNN remain substantially more accurate, with SAE-MNN slightly outperforming DNN in median and mean RRMSE. This suggests that the neural-network component successfully captures the nonlinear structure while the random-effects component does not impose excessive shrinkage when heterogeneity is weak.

Table 1. Mean and Median of RB and RRMSE over Areas for Point Estimates

Scenario	Model	RB [%]		RRMSE [%]	
		Median	**Mean**	**Median**	**Mean**
Normal–Small	GLMM	-4.360	-4.252	7.247	7.266
	MNN	0.217	0.263	2.778	2.797
	DNN	-0.406	-0.407	2.674	2.675
Interaction–Small	GLMM	1.516	1.571	3.959	3.996
	MNN	0.153	0.187	2.606	2.628
	DNN	-0.289	-0.295	2.576	2.593
Normal–Large	GLMM	-0.013	-0.079	9.955	10.022
	MNN	1.665	1.703	9.965	10.088
	DNN	13.893	15.236	31.098	31.369
Interaction–Large	GLMM	2.502	2.441	11.474	11.411
	MNN	0.579	0.485	8.389	8.388
	DNN	3.810	4.063	18.826	19.292

Under Normal-Large, both GLMM and SAE-MNN remain stable, while DNN deteriorates sharply. Although GLMM and SAE-MNN have similar RRMSE in this setting, SAE-MNN remains competitive while allowing a more flexible functional form. Under Interaction-Large, SAE-MNN clearly outperforms both GLMM and DNN, achieving the smallest bias and lowest RRMSE. This is the most relevant scenario for the proposed method, since it combines strong nonlinear effects with pronounced area-level heterogeneity. The results show that neither GLMM nor DNN alone is sufficient in this setting: GLMM preserves borrowing strength but lacks nonlinear flexibility, while DNN captures complex patterns but fails to account for the hierarchical structure. SAE-MNN is effective precisely because it addresses both requirements simultaneously.

Taken together, the point-estimation results reveal a central trade-off. When heterogeneity is low, neural-network-based methods dominate because flexibility is more important than shrinkage. When heterogeneity is high, however, the inclusion of random effects becomes critical. SAE-MNN performs robustly across all scenarios and is especially advantageous when nonlinear covariate effects and hierarchical variation coexist. This is consistent with the broader literature arguing that hybrid models are needed when neither classical mixed models nor purely predictive machine learning methods can adequately represent the full data-generating structure [20], [24], [28], [29].

3.1.2 Performance of Bootstrap MSE Estimators

Point estimation alone is not sufficient in small area estimation, where uncertainty quantification is equally important. Table 2 summarizes the performance of the bootstrap MSE estimators in terms of RB-RMSE and RRMSE-RMSE across areas. These measures evaluate how well the estimated MSE reflects the empirical RMSE observed across simulation replications. Table 2 shows bootstrap MSE estimator performance varies across scenarios. Normal-Small scenario, in this low-variability and well-specified setting, both GLMM and MNN perform very well. Their RB–RMSE values are small (around −1 % for GLMM, −6 % for MNN), indicating nearly unbiased MSE estimation. The corresponding RRMSE–RMSE values (≈ 10–11 % for GLMM and ≈ 38 % for MNN) show that GLMM produces slightly more stable estimates but MNN remains acceptable. In contrast, DNN exhibits very large negative RB–RMSE (−43 %) and inflated RRMSE–RMSE (≈ 61 %), confirming severe underestimation of the true uncertainty. This pattern highlights that when random effects are weak, conventional mixed models outperform purely predictive neural networks in variance estimation.

Under Normal-Small, both GLMM and SAE-MNN perform reasonably well, with small RB-RMSE values and acceptable dispersion. GLMM yields slightly more stable MSE estimation, while SAE-MNN remains close enough to indicate satisfactory uncertainty quantification. By contrast, DNN shows very large negative RB-RMSE and inflated RRMSE-RMSE, indicating severe underestimation of uncertainty. Even though DNN performs well in point estimation under low heterogeneity, it does not provide reliable model-based uncertainty measures. This difference is important: strong predictive accuracy does not automatically translate into valid inference.

Table 2. Performance of Bootstrap MSE Estimators In Model-Based Simulation: Mean and Median of RB-RMSE and RRMSE-RMSE Over Areas

Scenario	Model	RB-RMSE [%]		RRMSE-RMSE [%]	
		Median	Mean	Median	Mean
Normal–Small	GLMM	-0.73	-1.15	10.13	10.91
	MNN	-5.97	-6.25	38.09	38.76
	DNN	-43.29	-44.23	61.09	61.30
Interaction–Small	GLMM	-34.83	-34.60	79.47	79.84
	MNN	-4.28	-4.57	37.82	38.31
	DNN	-25.83	-24.36	63.17	62.35
Normal–Large	GLMM	-1.07	-1.19	8.15	9.11
	MNN	-3.27	-2.64	8.70	9.38
	DNN	-98.36	-98.35	98.37	98.36
Interaction–Large	GLMM	0.52	0.76	9.74	10.53
	MNN	-3.63	-3.06	9.86	10.65
	DNN	-90.11	-90.06	90.41	90.37

Under Interaction-Small, the effect of model misspecification becomes more visible. GLMM displays strong negative bias in MSE estimation and much larger dispersion, reflecting the inability of a linear latent predictor to account for nonlinear interactions. SAE-MNN performs substantially better, with markedly smaller bias and lower dispersion, indicating that the neural-network component improves not only prediction but also the accuracy of the uncertainty estimates. DNN again remains unstable, showing that flexible prediction without structured variance modeling is insufficient for valid interval estimation.

Under Normal-Large, both GLMM and SAE-MNN achieve near-unbiased MSE estimation with very low relative dispersion, whereas DNN collapses almost completely. Under Interaction-Large, GLMM and SAE-MNN again show strong performance, but SAE-MNN maintains slightly better balance between flexibility and inferential stability. DNN continues to produce severely biased and unstable uncertainty estimates. This reinforces the conclusion that uncertainty estimation in SAE requires explicit hierarchical modeling and cannot be recovered adequately from a purely predictive neural model without random effects [6], [8], [24].

Across all scenarios, SAE-MNN achieves the most favorable compromise. It remains close to GLMM under settings where the mixed-model structure is correctly specified, while substantially outperforming GLMM when nonlinear interactions are present. This is a key result of the study: the proposed method does not merely improve point prediction, but also preserves inferential stability to a much greater extent than standard DNN. From a methodological SAE perspective, this is crucial, because the usefulness of a model-based estimator depends not only on bias and RMSE but also on whether its uncertainty measures are trustworthy [8], [9], [16].

3.1.3 Confidence Interval Coverage and Width

Fig. 3 and Table 3 summarize the inferential behavior of the competing methods through empirical coverage probability and average confidence interval width. A narrow interval is desirable only if it still achieves acceptable coverage; otherwise, it reflects overconfidence rather than inferential efficiency. Under Normal-Small, GLMM and SAE-MNN both achieve coverage close to the nominal level, with GLMM producing wider intervals and SAE-MNN producing narrower but still reasonably valid intervals. DNN, on the other hand, produces the narrowest intervals but with clear under-coverage. This shows that DNN's apparently precise intervals are misleading, since they do not adequately reflect the true uncertainty. The result also illustrates the difference between predictive sharpness and inferential validity.

Under Interaction-Small, GLMM coverage drops sharply, indicating misspecification under nonlinear interactions. SAE-MNN, however, maintains high coverage with relatively narrow intervals. This is an important finding: the proposed model is able to accommodate nonlinear structure without sacrificing interval validity. DNN still underestimates uncertainty, with intervals that remain too narrow. Thus, in this scenario SAE-MNN clearly dominates both alternatives by achieving a better balance between flexibility and valid uncertainty estimation.

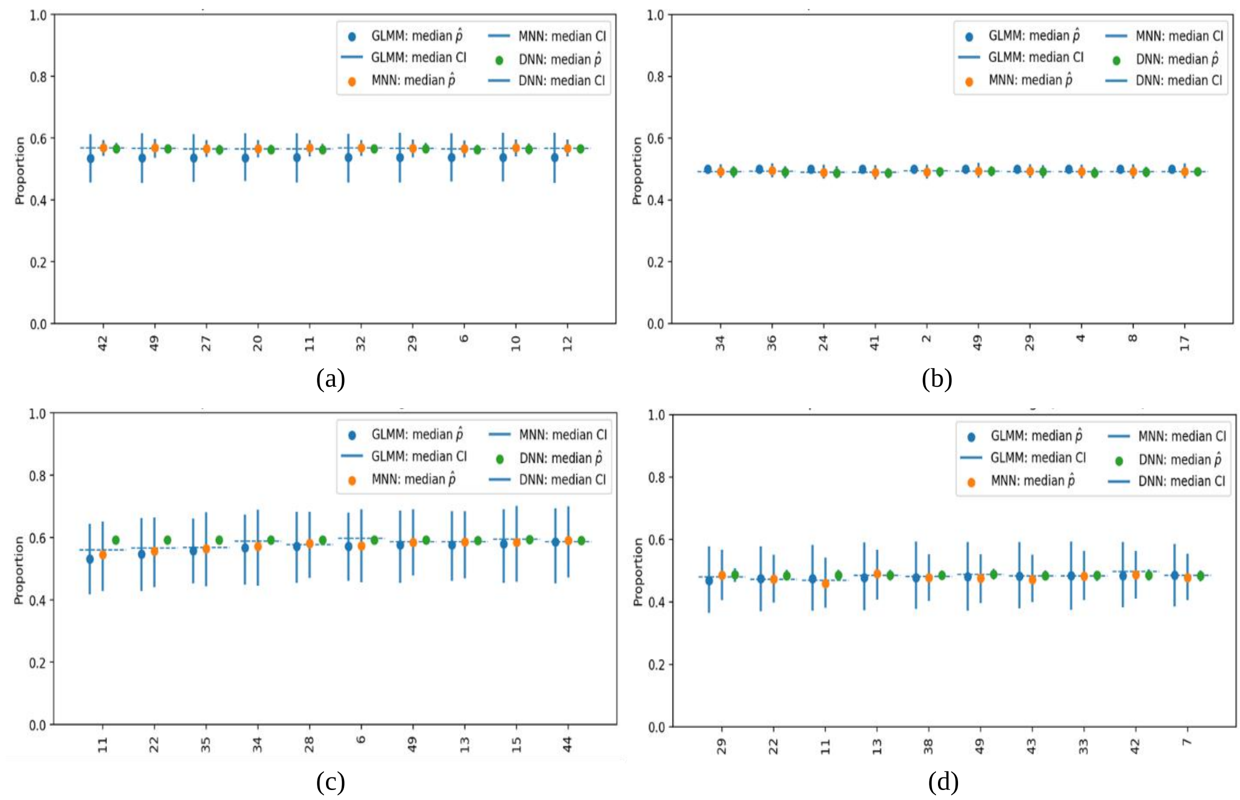


Figure 3. Confidence interval (CI) plots across 10 domains of four simulation scenarios
(a) Normal–Small, (b) Interaction–Small, (c) Normal–Large, (d) Interaction–Large

Under Normal-Large, both GLMM and SAE-MNN achieve nearly identical and near-nominal coverage, with similar interval widths. In this setting, the two methods are inferentially comparable, but SAE-MNN retains the advantage of a more flexible fixed-effects structure. DNN again performs poorly, with extremely low coverage and unrealistically narrow intervals. Under Interaction-Large, both GLMM and SAE-MNN achieve excellent coverage, but SAE-MNN does so with narrower intervals. This indicates a genuine efficiency gain: unlike DNN, which produces narrow but invalid intervals, SAE-MNN achieves narrower intervals without compromising coverage.

Table 3. Performance of Bootstrap MSE Estimators in Model-Based Simulation: Mean of Coverage and Width of Confidence Interval Over Areas

	Normal-Small			Interaction-Small		
	DNN	**GLMM**	**MNN**	**DNN**	**GLMM**	**MNN**
Coverage	0.60	0.94	0.89	0.68	0.49	0.89
CI's width	0.03	0.16	0.06	0.04	0.05	0.05
	Normal-Large			**Interaction-Large**		
	DNN	**GLMM**	**MNN**	**DNN**	**GLMM**	**MNN**
Coverage	0.02	0.93	0.93	0.14	0.94	0.94
CI's width	0.01	0.23	0.23	0.04	0.22	0.15

The contrast among methods illustrates the fundamental trade-off between predictive flexibility and inferential validity. GLMM remains the most conservative method, often producing wider intervals that ensure reliable coverage. DNN is the opposite: it may fit nonlinear patterns well, but it lacks a principled variance structure and therefore produces unreliable confidence intervals. SAE-MNN occupies the middle ground in the most favorable sense: it preserves the hierarchical mixed-model mechanism required for valid uncertainty quantification, while also introducing nonlinear flexibility through the neural-network component. This is precisely the type of balance needed in modern SAE applications involving complex auxiliary information [6], [24], [28], [29].

3.1.4 Overall Discussion of Simulation Results

The results across all simulation scenarios consistently support the methodological rationale of the proposed SAE-MNN. First, when nonlinear relationships are present but area-level heterogeneity is weak, neural-network-based approaches outperform GLMM in point estimation because they are able to adapt to the true predictor-response structure more effectively. Second, when between-area heterogeneity is strong, methods that explicitly incorporate random effects are clearly superior, since they stabilize small area estimates by borrowing strength across domains. Third, only SAE-MNN is able to perform well across both dimensions simultaneously, because it integrates nonlinear representation learning with hierarchical variance modeling.

These findings are aligned with recent developments in hybrid machine-learning-based SAE, especially those emphasizing that flexibility alone is not sufficient without an explicit mechanism for handling clustered dependence [20], [24]. At the same time, the results also show that inferential performance cannot be inferred from point-estimation accuracy alone. Standard DNN performs competitively in some low-heterogeneity settings, but its uncertainty estimates and confidence intervals are consistently unreliable. Conversely, GLMM remains inferentially stable but may lose efficiency or become biased when nonlinear interactions are ignored. SAE-MNN overcomes this dichotomy by retaining the small area estimation logic of mixed models while replacing the restrictive linear fixed-effects structure with a learned nonlinear function.

From a methodological standpoint, the simulation study confirms that the proposed approach is especially relevant for unit-level proportion estimation problems in which auxiliary variables have complex nonlinear effects and domain-level clustering cannot be ignored. In such settings, SAE-MNN provides a principled alternative to both classical GLMM-based SAE and purely predictive deep learning methods. The overall evidence therefore supports the view that mixed-model neural-network frameworks are a promising direction for extending SAE methodology to data-rich environments characterized by nonlinear structure and hierarchical dependence [8], [9], [24], [29].

3.2 Real-Data Application: Estimation of Subdistrict-Level Harvest Proportions

3.2.1 Overall Comparison with the Direct Estimator

Table 4 summarizes the overall agreement of the competing EBP-based estimators with the direct estimator across all subdistrict-month combinations. Because the true area-level harvest proportions are not observed in the real data, the direct estimator is used here as an empirical benchmark rather than as a ground truth. Therefore, the reported correlation, RMSE, MAE, and mean signed difference should be interpreted as measures of empirical agreement with the direct estimator, rather than as accuracy relative to a known truth.

Table 4. Summary of Comparison EBP-Based Estimators with The Direct Estimator

Method	Corr	RMSE	MAE	Mean Signed Diff	Mean Estimate	SD Estimate	Mean Direct	SD Direct
SAE-MNN	0.776	0.167	0.128	-0.007	0.234	0.166	0.241	0.259
GLMM	0.551	0.239	0.206	0.094	0.335	0.103	0.241	0.259
DNN	0.738	0.177	0.135	-0.007	0.233	0.164	0.241	0.259

The results show that SAE-MNN provides the closest overall agreement with the direct estimator. Among the three competing methods, SAE-MNN achieves the highest correlation with the direct estimator (0.7756), together with the lowest RMSE (0.1672) and MAE (0.1283). The standard DNN ranks second, with a correlation of 0.7380, RMSE of 0.1769, and MAE of 0.1352. By contrast, GLMM shows the weakest agreement with the direct benchmark, with a substantially lower correlation (0.5510) and the largest RMSE (0.2390) and MAE (0.2065). These results indicate that SAE-MNN is better able to preserve the empirical area-level pattern observed in the direct estimator while still operating within a model-based estimation framework.

A similar pattern is observed in the signed differences and in the overall dispersion of the estimates. Both SAE-MNN and DNN yield mean estimates that are only slightly below the mean direct estimator, with mean signed differences of approximately -0.007, indicating relatively mild downward shrinkage. In contrast, GLMM produces a notably higher mean estimate than the direct benchmark, with a mean signed difference of about +0.094, suggesting a much stronger systematic deviation. In terms of variability, the direct estimator exhibits the largest dispersion across subdistrict-month combinations (SD = 0.2590), whereas GLMM yields the smallest dispersion (SD = 0.1034), reflecting the strongest smoothing effect. SAE-MNN and DNN occupy an intermediate position (SD = 0.1659 and 0.1641, respectively), indicating that both methods regularize the direct estimator while still preserving substantially more empirical variation than GLMM.

Overall, these findings suggest that SAE-MNN achieves the most favorable balance between empirical fidelity and model-based stabilization in the real-data application. Compared with GLMM, SAE-MNN follows the direct estimator much more closely while avoiding the excessive smoothing implied by the linear mixed-model structure. Compared with DNN, SAE-MNN also shows slightly better agreement with the direct estimator across all overall measures. Thus, at the level of overall subdistrict-month comparison, the real-data results support SAE-MNN as a compromise estimator that remains close to the empirical benchmark while still preserving the borrowing-strength mechanism required in small area estimation.

3.2.2 Monthly Pattern of Area-Level Estimates

Fig. 4 presents the monthly mean of subdistrict-level harvest proportions obtained from the direct estimator, SAE-MNN, GLMM, and DNN over the twelve months of 2024. The figure reveals that all three model-based estimators broadly capture the temporal pattern of harvest dynamics, but they differ substantially in their level and smoothness relative to the direct estimator.

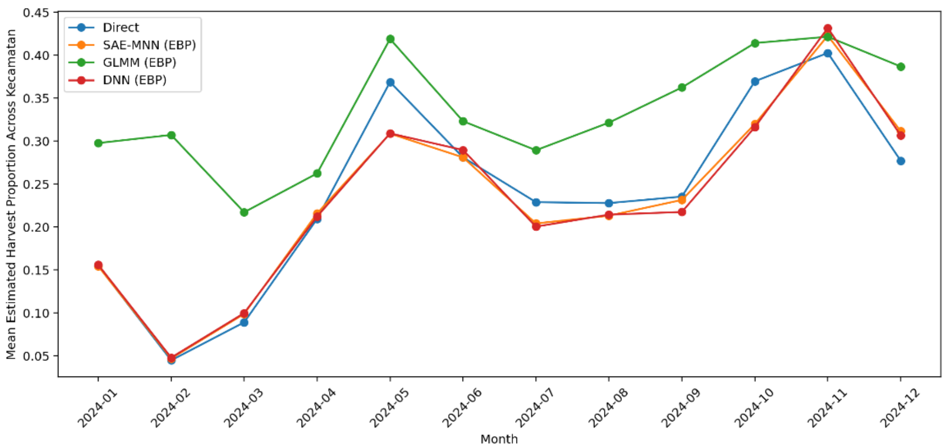


Figure 4. Monthly Mean of Subdistrict-Level Harvest Proportions

Overall, SAE-MNN and DNN track the direct estimator much more closely than GLMM. Across most months, the trajectories of SAE-MNN and DNN lie near the direct series and reproduce the main turning points of the observed harvest pattern, including the sharp increase from March to May, the decline from June to July, and the subsequent rise toward October–November. In contrast, GLMM remains consistently above the direct estimator in nearly all months and produces a smoother, more elevated trajectory. This pattern suggests that the linear mixed-model structure induces stronger regularization and a more pronounced upward shift in the real-data application, whereas the nonlinear methods preserve the empirical monthly movement more closely.

A closer inspection shows that SAE-MNN provides a favorable compromise between fidelity to the direct estimator and temporal stabilization. In January to April, SAE-MNN almost overlaps with DNN and remains very close to the direct estimator, indicating that both nonlinear approaches are able to reproduce the low-to-moderate harvest proportions observed early in the year. In May, when the direct estimator reaches a local peak, SAE-MNN and DNN both remain below the direct value, but the difference is notably smaller than that observed for GLMM, which overshoots the direct benchmark. During July to September, when the direct estimator stays at a moderate level, SAE-MNN continues to follow the direct trajectory closely, whereas GLMM remains systematically higher. In October and November, all methods rise sharply, but SAE-MNN and DNN show a steeper adjustment toward the direct pattern, while GLMM again appears more conservative and smoothed. In December, SAE-MNN remains closer to the direct estimator than GLMM, although all three model-based estimators lie above the direct benchmark.

The comparison between SAE-MNN and DNN is also informative. Their monthly mean trajectories are very similar throughout the year, which indicates that the nonlinear predictive component is the dominant driver of the observed temporal pattern. However, SAE-MNN generally appears slightly more regular while still remaining close to the direct benchmark. This behavior is consistent with the role of the area-level random effect in SAE-MNN, which introduces hierarchical borrowing strength without forcing the estimates toward the stronger smoothing pattern observed in GLMM.

Taken together, the monthly mean comparison supports the same conclusion obtained from the overall agreement measures: SAE-MNN preserves the main empirical temporal pattern of the direct estimator more closely than GLMM, while offering a more structured small-area model than DNN. Thus, at the aggregate monthly level, SAE-MNN achieves a useful balance between nonlinear adaptability and model-based stabilization in estimating subdistrict-level harvest proportions.

3.2.3 Shrinkage Behavior Relative to the Direct Estimator

Table 5 summarizes the shrinkage behavior of the competing EBP-based estimators relative to the direct estimator. In this context, shrinkage is assessed through the mean and median absolute deviation from the direct estimator, as well as through the relative dispersion of the model-based estimates compared with the dispersion of the direct benchmark. These measures are useful for evaluating how strongly each method regularizes the direct estimator in the real-data application.

Table 5. Summarizes of The Shrinkage behavior of EBP-Based Estimators Relative to The Direct Estimator

Method	Mean Absolute	Median Absolute	Standard Dev Estimate	Standard Dev Direct	SD Ratio Estimate to direct
SAE-MNN	0.128	0.094	0.166	0.259	0.640
GLMM	0.206	0.207	0.103	0.259	0.399
DNN	0.135	0.099	0.164	0.259	0.635

The results show that GLMM induces the strongest shrinkage among the three competing methods. It has the largest mean absolute deviation from the direct estimator (0.2065) and the largest median absolute deviation (0.2074), indicating that its area-level estimates depart more substantially from the direct benchmark than those of the nonlinear approaches. This stronger shrinkage is also reflected in its much smaller standard deviation (0.1032) compared with the direct estimator (0.2590), corresponding to a variance ratio of only 0.3985. Thus, GLMM produces the most compressed distribution of area-level estimates, which is consistent with the smoother and systematically elevated trajectory observed in the monthly mean comparison.

By contrast, SAE-MNN and DNN exhibit substantially weaker shrinkage and remain much closer to the direct estimator. SAE-MNN yields the smallest mean absolute deviation (0.1283) and the smallest median absolute deviation (0.0943), while DNN shows slightly larger values (0.1352 and 0.0985, respectively). In terms of dispersion, both methods preserve considerably more of the empirical variability of the direct estimator than GLMM, with standard deviations of 0.1657 for SAE-MNN and 0.1643 for DNN. Their corresponding variance ratios relative to the direct estimator are 0.6400 and 0.6345, respectively, indicating that both nonlinear approaches retain a substantially larger share of area-level variation.

These results suggest that SAE-MNN provides the most balanced shrinkage behavior in the real-data application. Compared with GLMM, it avoids excessive compression of the area-level estimates and remains much closer to the empirical direct pattern. Compared with DNN, SAE-MNN yields slightly smaller deviation from the direct estimator while maintaining a very similar level of dispersion. Therefore, the shrinkage analysis supports the interpretation that SAE-MNN achieves a useful compromise between borrowing strength and local adaptability: it regularizes the direct estimator enough to stabilize area-level estimates, but not so strongly that meaningful empirical variation is suppressed.

Taken together with the overall agreement measures and the monthly mean comparison, the shrinkage results reinforce the conclusion that SAE-MNN is the most favorable compromise estimator among the three competing methods. GLMM appears overly conservative in the real-data setting, whereas DNN preserves local variation but lacks the explicit hierarchical structure of a mixed model. SAE-MNN, in contrast, combines nonlinear flexibility with moderate and interpretable model-based shrinkage, making it well suited for estimating subdistrict-level harvest proportions from segment-level KSA data.

3.2.4 Overall Interpretation of the Real-Data Application

Taken together, the real-data results show a consistent pattern across the three comparative perspectives—overall agreement with the direct estimator, monthly temporal behavior, and shrinkage relative to the direct benchmark. In all three aspects, SAE-MNN emerges as the most balanced estimator. It achieves the closest overall agreement with the direct estimator, reproduces the main monthly harvest pattern more faithfully than GLMM, and exhibits moderate shrinkage that stabilizes the area-level estimates without excessively compressing their variability. By contrast, GLMM produces the strongest smoothing effect and departs most substantially from the direct benchmark, indicating that the linear mixed-model structure may be overly restrictive in the present real-data setting. DNN follows the direct estimator more closely than GLMM and preserves local variation well, but its behavior remains purely predictive and does not explicitly incorporate the hierarchical borrowing-strength mechanism required in small area estimation. Therefore, the real-data application supports the same general conclusion obtained from the simulation study: SAE-MNN provides a useful compromise between nonlinear adaptability and model-based regularization, making it a promising approach for estimating subdistrict-level harvest proportions from KSA segment-level data with complex auxiliary covariates.

3.3 General Discussion

The simulation study and the real-data application lead to a consistent overall conclusion. In the simulation setting, the controlled data-generating mechanisms show that the main challenge in unit-level proportion small area estimation is not only to capture nonlinear predictor-response relationships, but also to preserve the hierarchical borrowing-strength mechanism required for stable area-level estimation. Neural-network-based methods perform well when between-area heterogeneity is weak, because their flexibility allows them to adapt to nonlinear effects more effectively than GLMM. However, when area-level heterogeneity becomes substantial, methods that ignore the hierarchical dependence structure deteriorate noticeably. This confirms that flexibility alone is not sufficient in modern SAE problems.

The real-data application reinforces this conclusion from an empirical perspective. Although the true area-level proportions are unknown and the direct estimator serves only as an empirical benchmark, the same pattern remains visible. SAE-MNN shows the closest overall agreement with the direct estimator, reproduces the monthly harvest pattern more faithfully than GLMM, and exhibits more moderate shrinkage than the linear mixed-model benchmark. DNN also follows the direct pattern relatively closely, but it does so without explicitly preserving the hierarchical borrowing-strength mechanism central to small area estimation. Thus, the real-data results suggest that predictive closeness alone is not enough; the estimator must also retain an interpretable model-based structure for stabilizing area-level inference.

Taken together, the results clarify the roles of the three competing approaches. GLMM provides a stable inferential benchmark, but its linear fixed-effects formulation may be too restrictive when auxiliary variables exhibit strong nonlinear interactions. DNN offers greater predictive flexibility, but without explicit random effects it remains essentially a predictive tool rather than a full SAE estimator. SAE-MNN occupies the methodologically important middle ground by combining nonlinear adaptability with hierarchical regularization. In this sense, the proposed method should not be viewed simply as a more flexible prediction model, but as an extension of the SAE framework itself to settings where classical parametric assumptions are no longer adequate.

Overall, the study highlights a fundamental trade-off between predictive flexibility and inferential validity. DNN emphasizes flexibility but lacks an explicit mechanism for hierarchical variance modeling, while GLMM preserves inferential structure but may lose efficiency under nonlinear misspecification. SAE-MNN addresses this trade-off by integrating both components in a single framework. Accordingly, the proposed method is especially relevant for unit-level small area estimation problems characterized by complex auxiliary information, nonlinear covariate-response relationships, and non-negligible area-level heterogeneity.

4. CONCLUSION

This study proposes a modified Small Area Estimation Mixed Model Neural Network (SAE-MNN) for unit-level proportion estimation under clustered binary data. By replacing the linear fixed-effects component of the conventional logistic mixed model with a neural-network-based nonlinear function while retaining area-level random effects, SAE-MNN addresses a key limitation of classical GLMM-based small area estimation in representing complex nonlinear covariate-response relationships. The simulation results show that SAE-MNN achieves a favorable balance between predictive flexibility and inferential validity. It performs robustly across all scenarios and is especially advantageous when nonlinear interactions and substantial between-area heterogeneity coexist. Unlike a standard DNN, SAE-MNN improves not only point estimation but also the reliability of bootstrap-based uncertainty estimates and confidence interval coverage. The real-data application leads to a consistent conclusion: using KSA segment-level harvest proportions to estimate subdistrict-level monthly harvest proportions, SAE-MNN shows the closest overall agreement with the direct estimator, reproduces the temporal pattern more faithfully than GLMM, and provides moderate shrinkage without excessive smoothing.

Overall, the findings confirm that neither a purely linear mixed model nor a purely predictive neural model is sufficient for unit-level small area estimation of proportions when the underlying process combines nonlinear fixed effects with area-level dependence. SAE-MNN addresses this gap by integrating nonlinear representation learning with area-level random effects in a coherent mixed-model framework, thereby extending hybrid SAE methodology to binary and proportional outcomes. Nevertheless, the study remains limited by its primary reliance on controlled simulation for methodological validation, the relatively high computational burden of the proposed method, and the use of a simple random-intercept area structure. Future research may therefore consider broader real-data applications, more efficient estimation strategies, richer hierarchical structures, alternative uncertainty approximation methods, and extensions to other non-Gaussian response types.

Author Contributions

Ade Riyawan: Conceptualization, Methodology, Formal Analysis, Software, Resources, Writing-Original Draft. Muhammad Nur Aidi: Supervision, Validation, Methodology, Writing - Review and Editing. Budi Susetyo: Supervision, Validation, Methodology, Visualization, Writing - Review and Editing. Anang Kurnia: Supervision, Validation, Methodology, Data Curation, Funding Acquisition, Resources, Writing - Review and Editing. All authors discussed the results and contributed to the final manuscript.

Funding Statement

This research was funded by Statistics Indonesia (BPS) through a doctoral scholarship grant from the national budget (APBN) of Indonesia.

Acknowledgment

The author gratefully acknowledges the guidance and support provided by academic supervisors and lectures at School of data science, Mathematics, and Informatics (SSMI), IPB University. Appreciation is also extended to Statistics Indonesia (BPS) for awarding the doctoral scholarship funded by the national budget (APBN).

Declarations

The authors declare that he/she has no conflicts of interest to report study.

Declaration of Generative AI and AI-assisted technologies

Generative AI tools were used solely for language refinement (grammar, spelling, and clarity). The scientific content, analysis, interpretation, and conclusions were developed entirely by the authors. The authors reviewed and approved all final text.

REFERENCES

[1] J. D. Merfeld, H. Chen, P. Lahiri, and D. Newhouse, "SMALL AREA ESTIMATION WITH GEOSPATIAL DATA: A PRIMER," 2024. [Online]. Available: https://unstats.un.org/iswghs/documents/

[2] J. Jiang and P. Lahiri, "MIXED MODEL PREDICTION AND SMALL AREA ESTIMATION," 2006. doi: https://doi.org/10.1007/BF02595419

[3] E. Berg, "EMPIRICAL BEST PREDICTION OF SMALL AREA MEANS BASED ON A UNIT-LEVEL GAMMA-POISSON MODEL," *J. Surv. Stat. Methodol.*, vol. 11, no. 4, pp. 873–894, Sep. 2023. doi: https://doi.org/10.1093/jssam/smac026

[4] N. Tzavidis, L.-C. Zhang, A. Luna, T. Schmid, and N. Rojas-Perilla, "FROM START TO FINISH: A FRAMEWORK FOR THE PRODUCTION OF SMALL AREA OFFICIAL STATISTICS," 2018. [Online]. Available: https://academic.oup.com/jrsssa/article/181/4/927/7072057. doi: https://doi.org/10.1111/rssa.12364

[5] D. Morales, M. D. Esteben, A. Perez, and T. Hobza, "A COURSE ON SMALL AREA ESTIMATION AND MIXED MODEL," 2021. doi: https://doi.org/10.1007/978-3-030-63757-6

[6] O. Faltys, T. Hobza, and D. Morales, "SMALL AREA ESTIMATION UNDER AREA-LEVEL GENERALIZED LINEAR MIXED MODELS," *Commun. Stat. Simul. Comput.*, vol. 51, no. 12, pp. 7404–7426, 2022. doi: https://doi.org/10.1080/03610918.2020.1836216

[7] P. J. Farrell, B. Macgibbon, and T. J. Tomberlin, "EMPIRICAL BAYES ESTIMATORS OF SMALL AREA PROPORTIONS IN MULTISTAGE DESIGNS," 1997.

[8] J. Jiang and P. Lahiri, "EMPIRICAL BEST PREDICTION FOR SMALL AREA INFERENCE WITH BINARY DATA," *Ann. Inst. Stat. Math.*, vol. 53, pp. 217–243, 2001. doi: https://doi.org/10.1023/A:1012410420337

[9] T. Hobza and D. Morales, "EMPIRICAL BEST PREDICTION UNDER UNIT-LEVEL LOGIT MIXED MODELS," *J. Off. Stat.*, vol. 32, no. 3, pp. 661–692, Sep. 2016. doi: https://doi.org/10.1515/jos-2016-0034

[10] S. Sugasawa and T. Kubokawa, "TRANSFORMING RESPONSE VALUES IN SMALL AREA PREDICTION," *Comput. Stat. Data Anal.*, vol. 114, pp. 47–60, Oct. 2017. doi: https://doi.org/10.1016/j.csda.2017.03.017

[11] P. Corral, I. Molina, A. Cojocaru, and S. Segovia, "GUIDELINES TO SMALL AREA ESTIMATION FOR POVERTY MAPPING," 2022. doi: https://doi.org/10.1596/37728

[12] A. Kurnia, D. Surida, S. Muchlisoh, I. Wulandari, and D. Handayani, "SMALL AREA ESTIMATION UNDER A PSEUDO-UNIT LEVEL MODEL WITH FIRST-ORDER AUTOREGRESSIVE TIME EFFECTS AND MEASUREMENT ERROR," *IAENG International Journal of Applied Mathematics*, vol. 54, no. 3, pp. 365–372, 2024.

[13] N. Tzavidis, N. Salvati, M. Pratesi, and R. Chambers, "M-QUANTILE MODELS WITH APPLICATION TO POVERTY MAPPING," *Stat. Methods Appt.*, vol. 17, no. 3, pp. 393–411, 2008. doi: https://doi.org/10.1007/s10260-007-0070-8

[14] H. Li, Y. Liu, and R. Zhang, "SMALL AREA ESTIMATION UNDER TRANSFORMED NESTED-ERROR REGRESSION MODELS," *Statistical Papers*, vol. 60, no. 4, pp. 1397–1418, Aug. 2019. doi: https://doi.org/10.1007/s00362-017-0879-7

[15] Y. Kurniawati, H. Wijayanto, A. Kurnia, D. D. Domiri, and B. Susetyo, "SELECTION OF MULTINOMIAL LOGIT MODELS BASED ON ACCURACY RECLASSIFICATION OF THE AREA SAMPLING FRAME LABELS," *Science & Technology Asia* |, vol. 28, no. 2, 2023.

[16] J. Jiang and J. S. Rao, "ANNUAL REVIEW OF STATISTICS AND ITS APPLICATION ROBUST SMALL AREA ESTIMATION: AN OVERVIEW," 2020. doi: https://doi.org/10.1146/annurev-statistics-031219-041212

[17] D. Newhouse, J. D. Merfeld, A. Pudugramam, R. T. Swartz, and P. Lahiri, "SMALL AREA ESTIMATION OF MONETARY POVERTY IN MEXICO USING SATELLITE IMAGERY AND MACHINE LEARNING," 2022. [Online]. Available: http://www.worldbank.org/prwp. doi: https://doi.org/10.1596/1813-9450-10175

[18] P. Bilton, G. Jones, S. Ganesh, and S. Haslett, "CLASSIFICATION TREES FOR POVERTY MAPPING," *Comput. Stat. Data Anal.*, vol. 115, pp. 53–66, Nov. 2017. doi: https://doi.org/10.1016/j.csda.2017.05.009

[19] M. Viljanen, L. Meijerink, L. Zwakhals, and J. van de Kassteele, "A MACHINE LEARNING APPROACH TO SMALL AREA ESTIMATION: PREDICTING THE HEALTH, HOUSING AND WELL-BEING OF THE POPULATION OF NETHERLANDS," *Int. J. Health Geogr.*, vol. 21, no. 1, Dec. 2022. doi: https://doi.org/10.1186/s12942-022-00304-5

[20] P. Krennmair and T. Schmid, "FLEXIBLE DOMAIN PREDICTION USING MIXED EFFECTS RANDOM FORESTS," *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 71, no. 5, pp. 1865–1894, Nov. 2022. doi: https://doi.org/10.1111/rssc.12600

[21] P. Krennmair, W. ¨ Nora, and T. Schmid, "TREE-BASED MACHINE LEARNING IN SMALL AREA ESTIMATION," 2022.

[22] A. Hajjem, F. Bellavance, and D. Larocque, "MIXED-EFFECTS RANDOM FOREST FOR CLUSTERED DATA," *J. Stat. Comput. Simul.*, vol. 84, no. 6, pp. 1313–1328, Jun. 2014. doi: https://doi.org/10.1080/00949655.2012.741599

[23] R. Ananda, K. A. Notodiputro, and M. N. Aidi, "MODIFIED MIXED EFFECTS RANDOM FOREST IN SMALL AREA ESTIMATION USING PCA AND ROTATION FOREST WITH CORRELATED AUXILIARY VARIABLES," *Scientific Journal of Informatics*, vol. 11, no. 3, pp. 705–720, Aug. 2024. doi: https://doi.org/10.15294/sji.v11i3.10633

[24] N. Frink and T. Schmid, "SMALL AREA ESTIMATION WITH GENERALIZED RANDOM FORESTS: ESTIMATING POVERTY RATES IN MEXICO," Jun. 2024, [Online]. Available: http://arxiv.org/abs/2406.03861

[25] D. W. Triscowati, B. Sartono, A. Kurnia, D. Dirgahayu, and A. W. Wijayanto, "CLASSIFICATION OF RICE-PLANT GROWTH PHASE USING SUPERVISED RANDOM FOREST METHOD BASED ON LANDSAT-8 MULTITEMPORAL DATA," *International Journal of Remote Sensing and Earth Sciences (IJReSES)*, vol. 16, no. 2, p. 187, Apr. 2020. doi: https://doi.org/10.30536/j.ijreses.2019.v16.a3217

[26] D. W. Triscowati, B. Sartono, A. Kurnia, D. D. Domiri, and A. W. Wijayanto, "MULTITEMPORAL REMOTE SENSING DATA FOR CLASSIFICATION OF FOOD CROPS PLANT PHASE USING SUPERVISED RANDOM FOREST," in *Sixth Geoinformation Science Symposium*, S. B. Wibowo, A. B. Rimba, A. A. Aziz, S. Phinn, J. T. Sri Sumantyo, H. Widyasamratri, and S. Arjasakusuma, Eds., SPIE, Nov. 2019, p. 10. doi: https://doi.org/10.1117/12.2547216

[27] K. Kita *et al.*, "BIMODAL ARTIFICIAL INTELLIGENCE USING TABNET FOR DIFFERENTIATING SPINAL CORD TUMORS—INTEGRATION OF PATIENT BACKGROUND INFORMATION AND IMAGES," *iScience*, vol. 26, no. 10, Oct. 2023. doi: https://doi.org/10.1016/j.isci.2023.107900

[28] Y. Xiong, H. J. Kim, and V. Singh, "MIXED EFFECTS NEURAL NETWORKS (MENETS) WITH APPLICATIONS TO GAZE ESTIMATION," 2019. doi: https://doi.org/10.1109/CVPR.2019.00793

[29] G. Simchoni and S. Rosset, "INTEGRATING RANDOM EFFECTS IN DEEP NEURAL NETWORKS," Jun. 2023, [Online]. Available: http://arxiv.org/abs/2206.03314

[30] M. N. Tran, N. Nguyen, D. Nott, and R. Kohn, "BAYESIAN DEEP NET GLM AND GLMM," *Journal of Computational and Graphical Statistics*, vol. 29, no. 1, pp. 97–113, Jan. 2019. doi: https://doi.org/10.1080/10618600.2019.1637747

[31] B. Yang *et al.*, "BRAN2SAT: REDUNDANT SATISFIABILITY LOGIC IN LYAPUNOV-BASED DISCRETE HOPFIELD NEURAL NETWORK," *J. Comput. Des. Eng.*, vol. 12, no. 4, pp. 185–204, Apr. 2025. doi: https://doi.org/10.1093/jcde/qwaf039

[32] Y. Chang, M. S. Mohd Kasihmuddin, W. N. Aqlili Ruzai, Y. Guo, and J. Chen, "WEIGHTED C-TYPE RANDOM 2 SATISFIABILITY IN DISCRETE HOPFIELD NEURAL NETWORK," *Eng. Appl. Artif. Intell.*, vol. 160, p. 111760, 2025. doi: https://doi.org/10.1016/j.engappai.2025.111760