



Assessing the influence of sample size and spatial sampling strategies on GeoAI-based digital elevation model error correction: A case study of the Sabaragamuwa University area, Sri Lanka

T.A.T.K.G. Jayathilaka^{1*} and G.N.H. Wanigaratne²

¹IFS R&D International (Pvt) Ltd, Orion Towers 1, Level 18, 736, Dr. Danister De Silva Mawatha, Colombo 09, Sri Lanka

²Department of Surveying and Geodesy, Faculty of Geomatics, Sabaragamuwa University of Sri Lanka

Abstract

Digital Elevation Models (DEMs) are essential for terrain analysis, hydrological modeling, disaster assessment, and environmental monitoring. However, freely available DEMs often contain significant vertical errors due to the terrain complexity, vegetation cover, and sensor-related limitations. Although Machine Learning (ML) techniques have strong potential for DEM corrections, the effect of sample sizes and sampling strategies under limited ground data conditions remains insufficiently studied. Therefore, this study investigated the influence of sample sizes and sampling strategies on Geo-Artificial Intelligence (GeoAI)-based DEM error correction within the Sabaragamuwa University area, Sri Lanka. A Random Forest (RF) regression model was developed by using the elevation, slope, aspect, and roughness derived from Shuttle Radar Topography Mission (SRTM) DEM as predictor variables while DEM vertical error was set as the target variable. Ground survey elevations collected through Total Station and Automatic Level instruments were used as reference data. To simulate the practical limitations in field elevation acquisition, multiple sample sizes, ranging from 50 to 500 points were evaluated under random, clustered, and stratified sampling strategies. Model performance was evaluated using Root Mean Square Error (RMSE), Coefficient of Determination (R^2), percentage improvement, and variogram based geostatistical analysis. The results demonstrated a significant improvement in DEM accuracy after correction. The original SRTM DEM RMSE of 5.010m was reduced to values between 1.541m and 2.575m. The stratified sampling strategy consistently produced the best overall performance (RMSE = 1.541m, R^2 = 0.952). Variogram analysis further confirmed a reduction in spatially autocorrelated errors in the corrected DEM while independent validation using reserved ground survey points further verified the reliability of the proposed framework. Overall, the findings demonstrated the effectiveness of the proposed GeoAI-based approach for cost-efficient DEM correction with limited ground data.

Keywords: DEM, GeoAI, Random Forest, Spatial Sampling, SRTM

Article info

Article history:

Received 20th June 2026

Received in revised form 15th July 2026

Accepted 25th August 2026

Available online 20th September 2026

ISSN (E-Copy): ISSN 3051-5262

ISSN (Hard copy): ISSN 3051-5602

Doi: <https://doi.org/10.4038/jmtr.v11i4.346>

ORCID iD: <https://orcid.org/0009-0004-8047-4704>

*Corresponding author:

E-mail address: thilaj1999@gmail.com (T.A.T.K.G. Jayathilaka)

CC BY-NC-SA © 2026, The Author(s)

Introduction

An accurate representation of the Earth's surface is crucial in a variety of applications, ranging from hydrologic modeling and landslide susceptibility analysis to infrastructure planning and environmental monitoring. The Digital Elevation Model (DEM) represents continuous elevation data and provides one of the primary sources necessary for these analyses. Recent studies have highlighted the importance of DEMs in terrain analysis, disaster warning, and environmental monitoring and have observed that publicly available DEMs, such as Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) and Shuttle Radar Topography Mission (SRTM) have significantly contributed to the quantification of landscape characteristics. (Preety et al., 2022; Zhu et al., 2025).

DEM elevation errors mainly occur due to sensor limitations, data acquisition techniques, and different surface conditions. Radar-based DEMs may exhibit elevation errors in areas covered by dense vegetation or located in complex topography conditions (Guenther et al., 2024). Moreover, Chen et al. (2024) have observed that satellite-based DEMs display systematic positive biases when representing the elevations of urban terrains because of building artifacts. Recent studies have used Machine Learning (ML) and optimization techniques for enhancing DEM accuracy across various terrain conditions (Kandil et al., 2025; Okolie et al., 2024; Tian & Shan, 2025).

Moreover, recent studies have extensively explored different ways to improve DEM precision by using sparse reference data. Tian and Shan (2025) emphasized that creating a continuous surface from sparse data remains challenging because of the high cost and the limited availability of data. Various researchers have confirmed the feasibility of using small samples of high-quality reference data to correct the DEM. Guenther et al. (2024) demonstrated that it is possible to substantially reduce DEM errors based on sparse ICESat-2 elevation data.

Finally, the use of ML algorithms has allowed for the development of effective and inexpensive approaches for enhancing the DEM. For example, Okolie et al. (2024) have shown that optimized models provide high-predictive accuracy under different conditions. However, most previous works have mainly focused on improving the quality of models, while limited attention to the influence of the sampling methods and sample sizes on DEM correction process.

In practice, the estimation of the minimum number of ground elevation points needed to ensure accurate DEM corrections is still uncertain, especially in areas with insufficient data availability. Although the larger sample sizes generally improve predictive capability of the models used, it increases surveying expenses and logistics involved. As such, the necessity of establishing a balance between data and modeling is important when it comes to applying this approach to real-world problems.

Accordingly, this study proposes a Geo-Artificial Intelligence (GeoAI)-based framework for SRTM DEM error correction using the ground survey points acquired within the study area. The dataset was divided systematically into multiple sample sizes under different sampling strategies to simulate real-world data scarcity scenarios. This study evaluates the influence of sample size on GeoAI-based DEM error correction, compares the performance of random, clustered, and

stratified sampling strategies and identifies a practically useful sampling configuration for DEM correction using limited ground survey data.

Methodology

Study area

Sabaragamuwa University of Sri Lanka is located in Belihuloya, Sabaragamuwa province of Sri Lanka, approximately at 6.714°N latitude and 80.787°E longitude (Figure 1). This area contains moderately complex topographic features such as gentle slopes and steeper terrain which makes it suitable for evaluating DEM accuracy and terrain dependent elevation error. This topographic diversity plays a significant role in investigating the influence of terrain characteristics on DEM error behavior and the performance of the models under different spatial conditions.

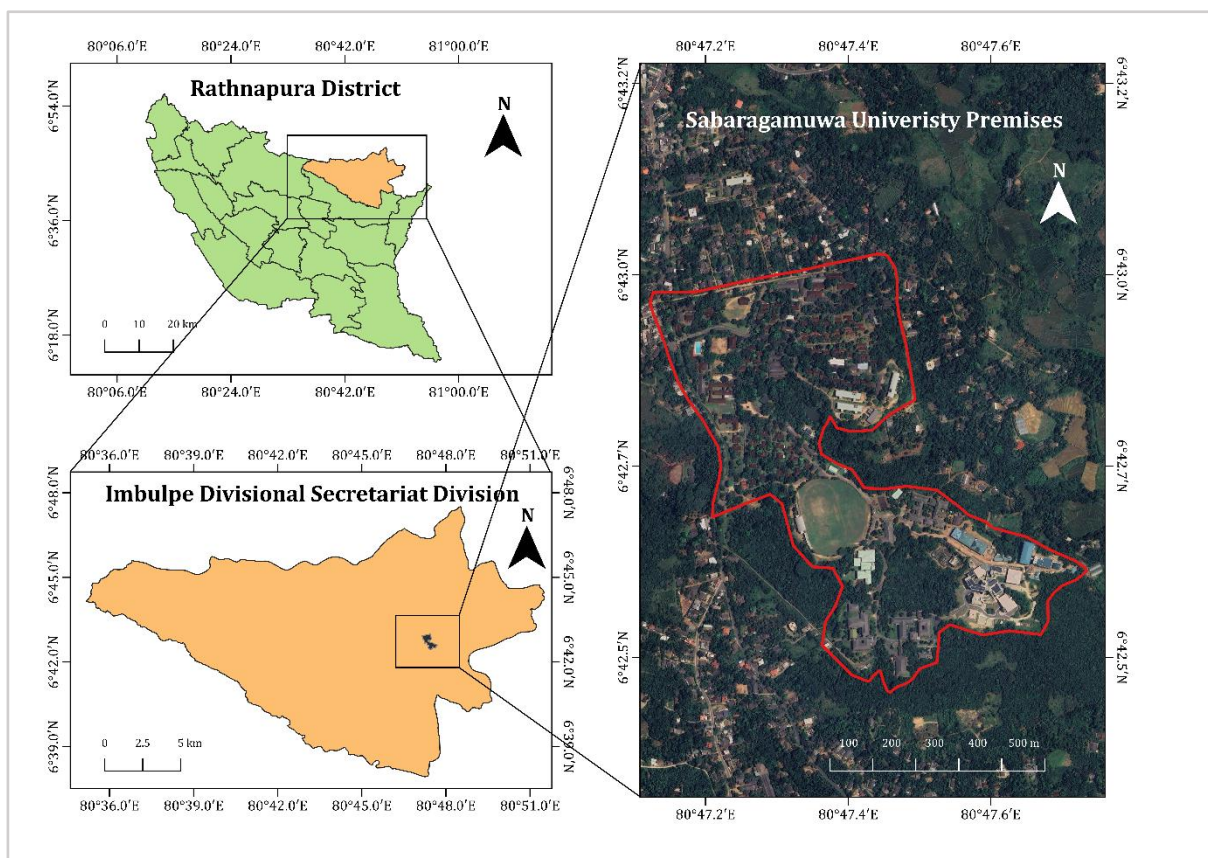


Figure 1. Study area (Location of Sabaragamuwa University of Sri Lanka)

Data sources and DEM error calculation

Figure 2 illustrates the full workflow that was followed during this study. A high-density elevation dataset, collected using Total Station and Automatic Level machines, was obtained from the Department of Surveying and Geodesy, Faculty of Geomatics, Sabaragamuwa University of Sri Lanka as the primary reference dataset. These survey points were used to quantify the vertical errors of the selected DEM and train the GeoAI model. The 30m SRTM DEM downloaded from the

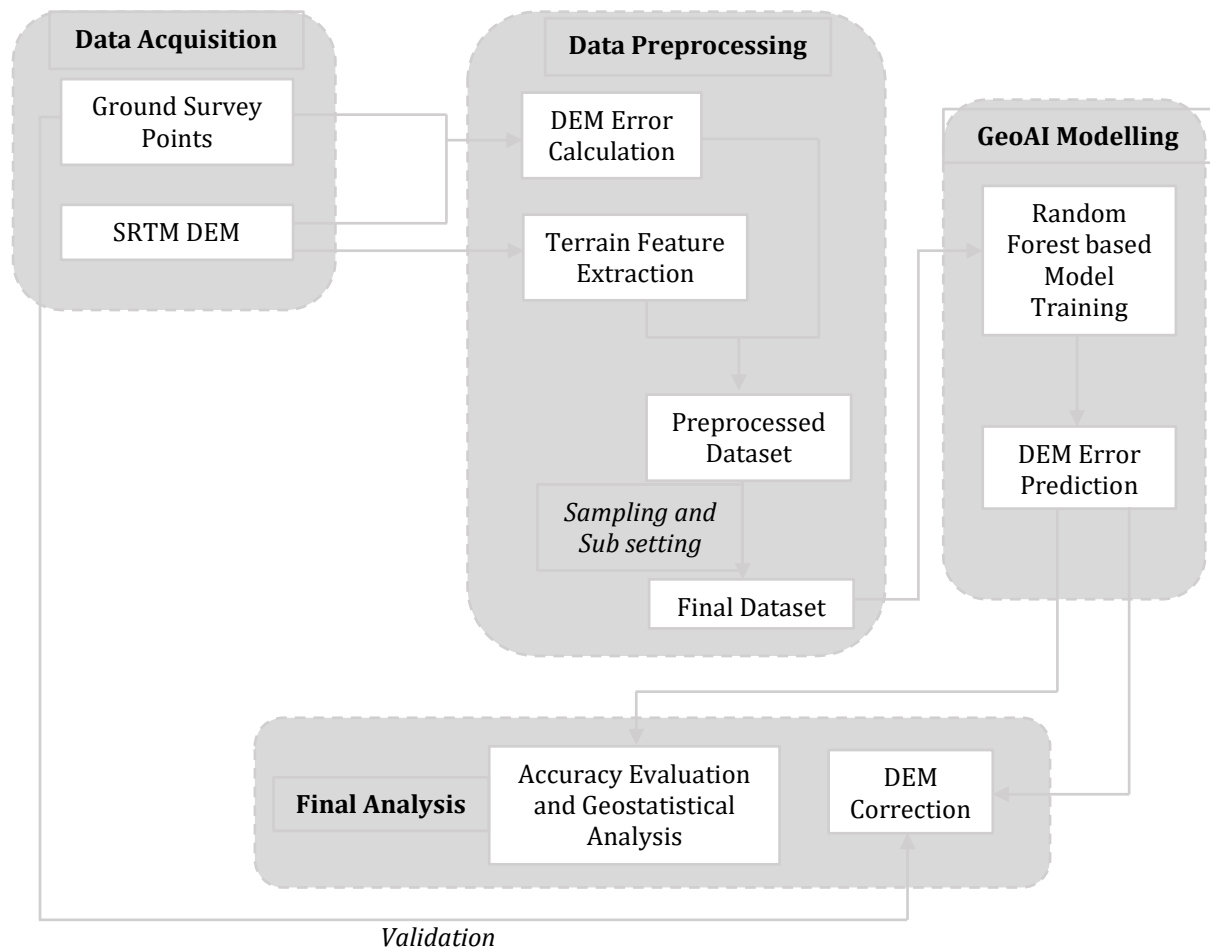


Figure 2. Flowchart of the methodology

United States Geological Survey - Earth Explorer (<https://earthexplorer.usgs.gov/>) was used as the key DEM elevation dataset for the study due to its high applicability in most of the terrain analysis and hydrological modelling.

Initially, the DEM error at each point was calculated as the difference between the DEM-derived elevation and the corresponding reference elevation obtained from the ground survey using Equation (1) where E_i is the vertical DEM error, Z_{DEM_i} represents the elevation extracted from SRTM DEM and the Z_{Survey_i} represents the reference elevation obtained from ground observations. The calculated errors were used as the target variable for the entire machine learning based modelling process.

$$E_i = Z_{DEM_i} - Z_{Survey_i} \quad (1)$$

Terrain derivatives and independent validation dataset

Terrain derivatives including slope, aspect and roughness were generated from the original SRTM DEM. Previous studies have identified the strong relationship between the DEM error patterns with them and elevations errors are strongly influenced by surface morphology (Chen et al., 2024;

Li & Heap, 2014). The values of elevation, slope, aspect and roughness corresponding to each ground surveyed points were extracted and integrated with the calculated DEM errors to form the final dataset.

Prior to the model development, a well spatially distributed dataset throughout the entire study area was separated from the original dataset for the final independent data validation of the corrected DEM. This data was reserved separately without using for any training, testing, sampling and model selection work processes in order to ensure the full unbiased final accuracy assessment of the developed model. These points were selected by dividing the study area into a 10×10 grid of spatial bins, defined by deciles of the X- and Y-coordinates of all survey points, and randomly drawing one point from each occupied bin. This procedure ensured the validation set was distributed across the full extent of the study area rather than concentrated in any sub-region.

Sampling strategies and experimental design

In order to train the model in different data scarcity conditions, the full data set was divided into multiple sample sizes ranging from 50 to 500 data points. For each sample size three main sample strategies were applied. For random sampling, the required number of observations was selected randomly without replacement from the available training dataset. This strategy represents an approximately unbiased selection of field observations without explicitly considering terrain characteristics. The Clustered sampling strategy was designed to represent a situation in which field observations are concentrated in relatively low-slope and potentially more accessible terrain. Specifically, observations were first restricted to locations with slope values below the 30th percentile of the training dataset's slope distribution. The required sample size was then randomly selected from this low-slope subset. For stratified sampling, the training data set was divided into four slope-based strata using the 25th, 50th, and 75th percentiles of the slope distribution. The four strata represented approximately equal-frequency groups ranging from lower to higher slope conditions. Approximately equal numbers of observations were selected from each stratum to obtain the required total sample size (Preety et al., 2022; Tian & Shan, 2025; Wang et al., 2012).

Since random sample selection can lead to differences in model performance, each sampling configuration was repeated 30 times with an independent Monte Carlo sample selection (Hermans et al., 2025). As such, 21 sampling configurations (3 sampling strategies $\times$ 7 sample sizes) were tested, representing 630 model runs. For each configuration, the mean RMSE, standard deviation (SD), and 95% confidence interval (CI) of the 30 repeated cross-validation were calculated. This procedure was followed to assess the variability introduced by the selection of ground-survey observations, rather than a single random sample.

Random forest-based DEM error correction

A Random Forest (RF) Regression model was used as the core DEM error predictor of GeoAI framework which is a machine learning algorithm based on decision trees as the relationship between terrain derivatives and vertical DEM errors is well suited for handling non-linear and relatively small training datasets. As shown in Equation (2), the model was trained using the

selected predictor variables for each dataset (Chen et al., 2024; Okolie et al., 2024). A separate RF regression model was developed for each combination of sampling strategy and sample size. The RF model consisted of 100 decision trees ($n_{\text{estimators}} = 100$), and $\text{random_state} = 42$ was specified to provide reproducibility of the stochastic model-building process. Other RF hyperparameters were retained at the default settings of the scikit-learn 'RandomForestRegressor' implementation.

$$\hat{E} = [Z_{DEM_i}, Slope_i, Aspect_i, Roughness_i] \quad (2)$$

The modelling framework involved the use of distinct data roles: (i) training data used for developing the RF model and sample selection and (ii) independent validation data, which were reserved before model development and, therefore, were not used for training or sample selection. The independent validation data were used to assess the final prediction accuracy of the corrected DEM.

After model training, the RF model was used for predicting the DEM error values for the entire surveyed dataset. Then the corrected DEM was generated by subtracting the predicted error from the original DEM data, as represented in Equation (3), where $Z_{corrected}$ represents the corrected DEM elevation, Z_{DEM} represents the original SRTM elevation, and $\hat{E}$ represents the DEM error predicted by the RF model.

$$Z_{corrected} = Z_{DEM} - \hat{E} \quad (3)$$

Accuracy assessment and statistical analysis

Performance of the models was evaluated through commonly used accuracy measures such as Root Mean Square Error (RMSE), the Coefficient of Determination (R^2), and percent improvement in accuracy compared to that of the original DEM. RMSE provided information on the average magnitude of elevation errors, while R^2 measured the variance in elevation values that was explained by the predictive model. The percentage improvement measure was allowed us to assess the relative enhancement in DEM accuracy through the use of the proposed correction method. Also, spatial patterns of DEM error were analyzed through variogram-based geostatistical framework to identify changes in the degree of spatial correlation before and after correction.

Variogram analysis and DEM refinement

Spatial patterns of DEM error were analyzed through a variogram-based geostatistical framework to identify changes in the degree of spatial correlation before and after correction.

The mean RMSE, RMSE standard deviation and 95% confidence interval were calculated to characterize the variability of the results over the 30 Monte Carlo repetitions. Statistical comparisons among sampling strategies were carried out using Welch's analysis of variance and pairwise comparisons (Midway et al., 2020). Holm adjustment was applied to control for multiple comparisons (Rogerson, 2024). Cohen's d_z was used as an effect-size measure for comparisons

between consecutive sample sizes (Azaad, 2026). The optimal sample size and the sampling strategy was selected based on the above calculated results.

After that original SRTM DEM Raster dataset was converted into point-based dataset in which each raster cell was represented as individual point containing DEM elevations and terrain derivative values. Then selected RF model was applied to the raster derived dataset and corrected DEM values were calculated accordingly. Finally, the corrected continuous DEM was prepared using the corrected DEM values obtained through the model using the Ordinary Kriging interpolation. It was selected because it explicitly accounts for spatial autocorrelation through variogram-based weighting, which was considered appropriate given the spatial structure identified in the DEM errors. Previous DEM interpolation studies have also demonstrated that Kriging can perform effectively where elevation exhibits strong spatial structure, although its performance remains dependent on terrain characteristics, sampling density, and data distribution (Arun, 2013; Li & Heap, 2014). Then the respective DEM values for validation points were extracted and the subsequent difference between the ground survey point elevation and refined DEM values were obtained through the GeoAI model and validation statistics were calculated..

Results

Overall DEM correction performance

The proposed DEM correction framework demonstrated significant improvement in elevation accuracy with respect to different sampling methods and sampling sizes. The baseline RMSE of the original SRTM DEM was 5.010 m. Following GeoAI-based error correction, the mean validation RMSE ranged from 2.173 m to 3.332 m across the 21 sampling configurations, corresponding to improvements of approximately 28.62%–56.62% relative to the original DEM (Table 1). These results demonstrate that the magnitude of improvement varied according to both the number and spatial sampling strategy of the training observations.

Each configuration was evaluated using 30 Monte Carlo runs. The relatively narrow confidence intervals indicate that the estimated performance was reasonably stable across repeated sampling realizations. The SD and 95% CI provide an estimate of the uncertainty related with the repeated sampling process.

Effect of sample size on DEM correction accuracy

The mean value of validation RMSE decreased with an increase in the sample size for both types of random and stratified sampling. Thus, for random sampling, RMSE changed from 3.067 m at 50 points of sampling operations to 2.211 m at 500 points of sampling operations. Stratified sampling showed RMSE changing from 3.140 m to 2.173 m over the same increase in the size of the sample. In contrast, the condition of clustered sampling became a less consistent response while there was an increase in the size of the sample. Specifically, the average value of RMSE showed a decline from 3.332 m at 50 points of sampling operation to 2.751 m at 300 points of sampling operation. After that, there was an increase of RMSE values reaching the values of 2.654 m and 2.632 at 400 and 500 points of sampling operation correspondingly. Hence, it can be said

that increase of the number of observations does not always lead to the improvement when the observations were concentrated within the low-slope regions.

Table 1: Performance comparison of different sampling strategies and sample sizes for GeoAI-based DEM error correction

Strategy	Sample Size	Mean RMSE (m)	SD (m)	95% CI (m)	Mean MAE (m)	Mean R ²	Improvement (%)
Random	50	3.067	0.254	2.972–3.162	2.345	0.125	34.29
Random	100	2.914	0.225	2.830–2.998	2.204	0.211	37.57
Random	150	2.746	0.182	2.678–2.814	2.053	0.3	41.16
Random	200	2.598	0.201	2.523–2.673	1.929	0.373	44.33
Random	300	2.433	0.153	2.376–2.490	1.793	0.451	47.88
Random	400	2.25	0.215	2.170–2.330	1.646	0.528	51.79
Random	500	2.211	0.17	2.147–2.274	1.569	0.95	55.87
Clustered	50	3.332	0.347	3.202–3.461	2.575	-0.037	28.62
Clustered	100	3.04	0.245	2.949–3.132	2.354	0.141	34.86
Clustered	150	2.87	0.203	2.794–2.945	2.228	0.235	38.52
Clustered	200	2.784	0.179	2.717–2.851	2.165	0.281	40.35
Clustered	300	2.751	0.13	2.702–2.799	2.128	0.299	41.06
Clustered	400	2.654	0.133	2.604–2.704	2.055	0.347	43.13
Clustered	500	2.632	0.143	2.579–2.686	2.038	0.358	43.61
Stratified	50	3.14	0.199	3.066–3.214	2.413	0.085	32.72
Stratified	100	2.878	0.207	2.801–2.956	2.174	0.231	38.33
Stratified	150	2.746	0.163	2.686–2.807	2.045	0.301	41.16
Stratified	200	2.64	0.166	2.578–2.702	1.946	0.354	43.45
Stratified	300	2.431	0.141	2.378–2.483	1.776	0.452	47.92
Stratified	400	2.28	0.141	2.228–2.333	1.651	0.518	51.14
Stratified	500	2.173	0.154	2.116–2.231	1.541	0.952	56.62

Comparison of sampling strategies

The results of the study were also evident in the mean R² values shown in Table 1. The mean R² value for random sampling improved steadily from 0.125 at 50 points to 0.950 at 500 points. Likewise, mean R² for stratified sampling increased from 0.085 at 50 points to 0.952 at 500 points. In contrast, the mean R² values for clustered sampling remained lower, growing from -0.037 at 50 points to 0.358 at 500 points. The results demonstrate that random and stratified sampling yielded better performance than clustered sampling when it comes to correctly predicting DEM elevations, particularly at a larger sample size.

The best performances occurred at 500 stratified points and showed a mean RMSE, MAE, and R value of 2.173 m, 1.541 m and 0.952, respectively, the mean improvement was 56.62%. At a similar sample size random samples gave slightly better RMSE of 2.211 m and R² of 0.950,

whereas clustered samples gave a higher mean RMSE of 2.632 m and an R^2 of 0.358, and it can be clearly observed that only when paired with stratified sampling is it consistently the best for mean RMSE at all evaluated sample size, yet pairwise comparison between random and stratified not statistically significant after Holms correction.

Statistical comparison of sampling strategies and sample sizes

Statistical comparisons were conducted to test whether observed differences between sampling strategies were significant. (See below Table 2 for the comparison results).

Table 2: Pairwise statistical comparison of sampling strategies at each sample size

Sample Size	Strategy A	Strategy B	Mean RMSE A (m)	Mean RMSE B (m)	Difference A-B (m)	Welch t	p-value	Holm-adjusted p	Significant
50	Random	Clustered	3.067	3.332	-0.265	-3.375	0.00138	0.01938	Yes
50	Random	Stratified	3.067	3.14	-0.073	-1.239	0.22057	1	No
50	Clustered	Stratified	3.332	3.14	0.192	2.629	0.01159	0.12749	No
100	Random	Clustered	2.914	3.04	-0.126	-2.075	0.04249	0.33995	No
100	Random	Stratified	2.914	2.878	0.036	0.645	0.52153	1	No
100	Clustered	Stratified	3.04	2.878	0.162	2.766	0.00765	0.09177	No
150	Random	Clustered	2.746	2.87	-0.124	-2.491	0.01565	0.14088	No
150	Random	Stratified	2.746	2.746	0	0	1	1	No
150	Clustered	Stratified	2.87	2.746	0.124	2.609	0.01166	0.12749	No
200	Random	Clustered	2.598	2.784	-0.186	-3.785	0.00037	0.00555	Yes
200	Random	Stratified	2.598	2.64	-0.042	-0.882	0.3813	1	No
200	Clustered	Stratified	2.784	2.64	0.144	3.231	0.00204	0.02653	Yes
300	Random	Clustered	2.433	2.751	-0.318	-8.675	<0.000001	<0.000001	Yes
300	Random	Stratified	2.433	2.431	0.002	0.053	0.95819	1	No
300	Clustered	Stratified	2.751	2.431	0.32	9.139	<0.000001	<0.000001	Yes
400	Random	Clustered	2.25	2.654	-0.404	-8.753	<0.000001	<0.000001	Yes
400	Random	Stratified	2.25	2.28	-0.030	-0.639	0.52568	1	No
400	Clustered	Stratified	2.654	2.28	0.374	10.568	<0.000001	<0.000001	Yes
500	Random	Clustered	2.211	2.632	-0.421	-10.380	<0.000001	<0.000001	Yes
500	Random	Stratified	2.211	2.173	0.038	0.907	0.368	1	No
500	Clustered	Stratified	2.838	2.173	0.665	18.569	<0.0001	<0.0001	Yes

The analysis of statistics has revealed a notable difference between the perceived ranking of sampling methods and the statistical relevance of their discrepancies. Although stratified sampling maintained the minimum mean RMSE across all sample sizes, the difference between it and random sampling is not statistically significant through all of the sample sizes after applying the Holm correction. The case is different for clustered sampling, which performs significantly worse than stratified sampling across all sample sizes and at random sampling starting from sample size 100.

The analysis conducted on sample sizes indicated that an increase in the number of observations generally improved the performance in the DEM correction process, although the magnitude of improvement depended on the method of sampling. In particular, sequence of increases in the number of samples led to statistically significant lower values of RMSE in the cases of both random sampling and stratified sampling. On the contrary, the method of cluster sampling yielded statistically significant improvement only until 300 sample points. After which further increases sample sizes to 400 and 500, statistically significant reduction in RMSE was not produced by the points.

Spatial variability of DEM errors

The results of the variogram analysis indicate differences in the spatial variability of DEM errors associated with different sampling strategies. The original sill value for DEM error was 22.759, while the sill value decreased when the error surfaces were corrected, especially in terms of stratified sampling. In this last case, the sill value at 500 points was equal to 16.400 for random sampling, 28.272 for clustered sampling, and 13.067 for stratified sampling. Lower sill achieved for stratified sampling indicates that the residual spatial variance was lower than for the original DEM error surface, while relatively high sill achieved for clustered sampling indicates considerable residual spatial variability (Figure 3).

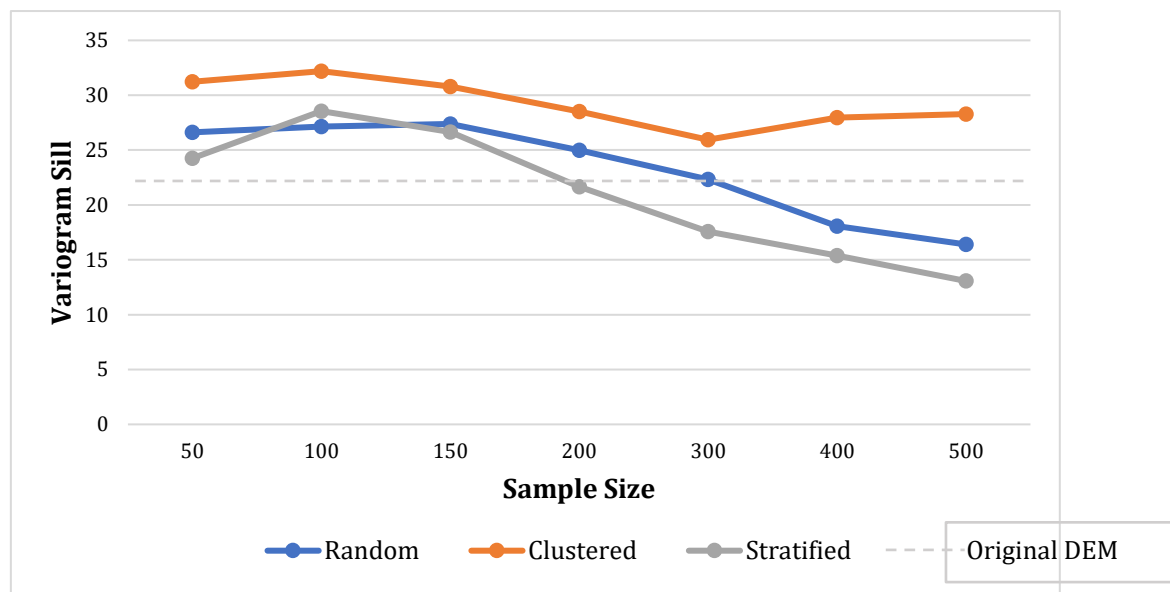


Figure 3. Variogram sill comparison under different sampling strategies

Validation and spatial refinement of the corrected DEM

The scatter plot showed that the corrected elevation values of DEM were strongly correlated with actual survey elevations. The corrected elevation of DEM closely followed the 1:1 reference line with regression analysis showing a high $R^2 = 0.976$ (Figure 4).

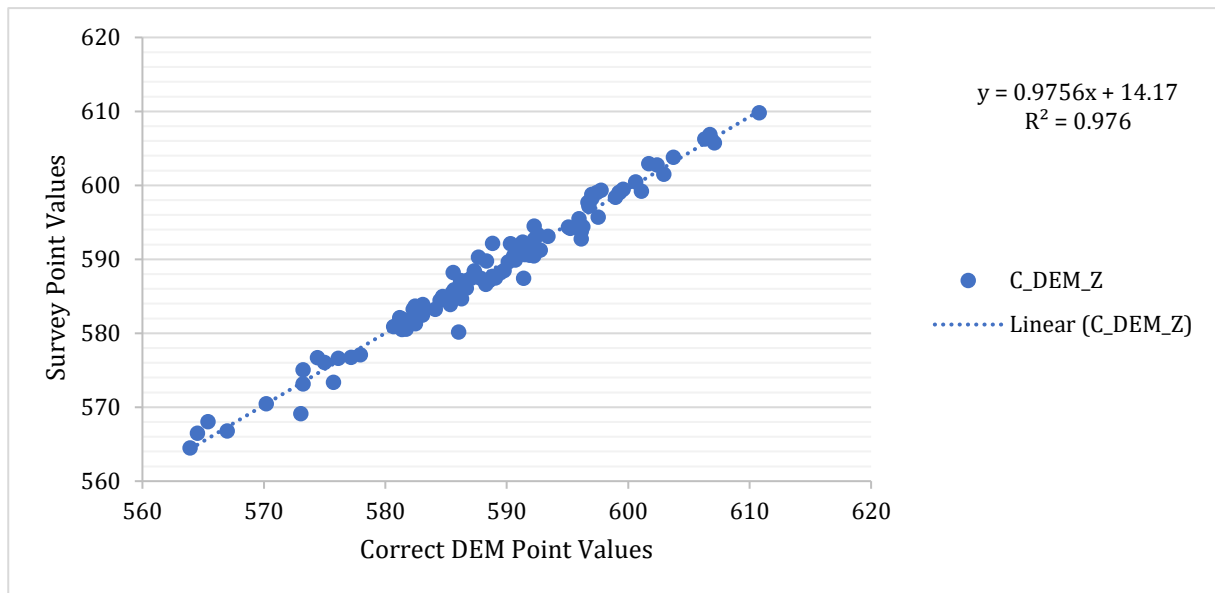
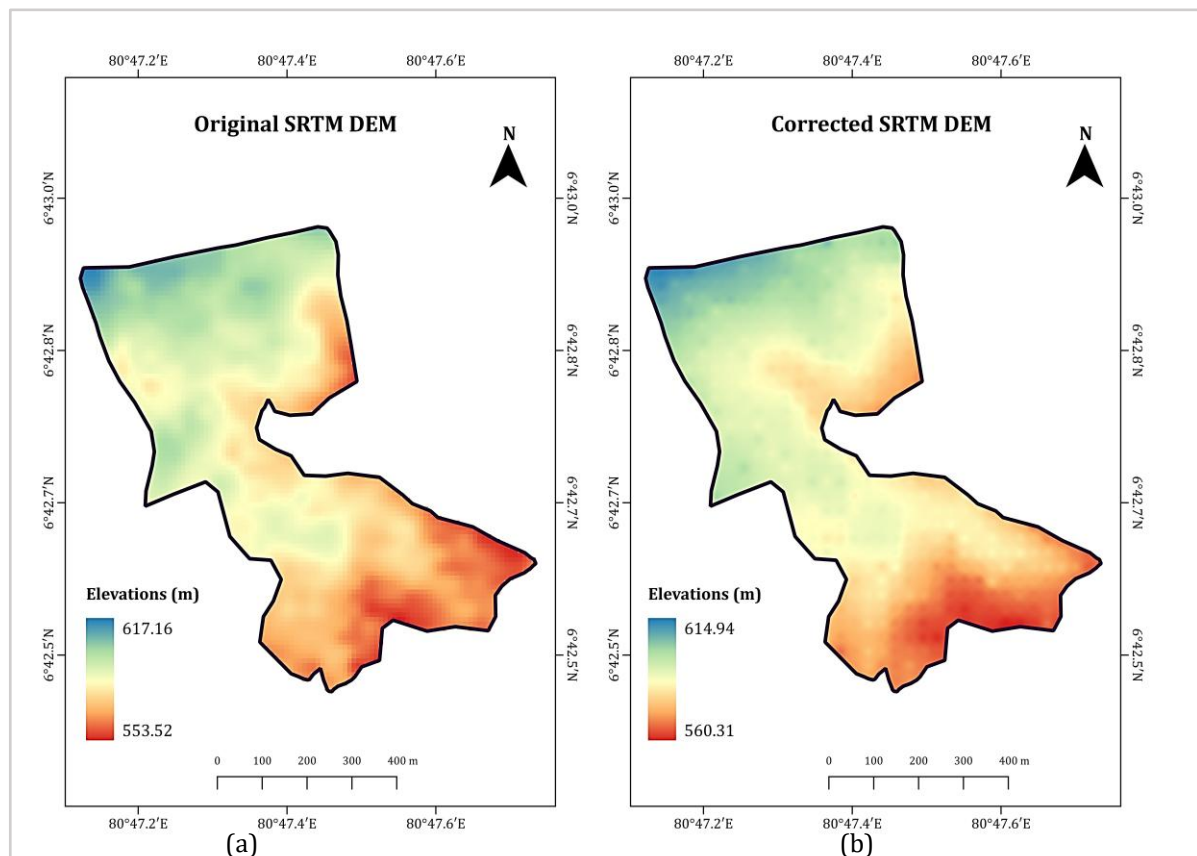


Figure 4. Scatter plot diagram of DEM validation

Figure 5 shows the Original SRTM DEM and the corrected DEM prepared using the Kriging interpolation method using the data derived from the developed GeoAI model. The visual comparison shows changes in spatial distribution of elevation values after correction.



Discussion

The findings of the present study indicate the significant potential of the proposed GeoAI model for improving the accuracy of DEMs with a limited number of ground survey points. As illustrated

by the significant reduction in RMSE for all the examined sampling strategies, the RF model is indeed capable of successfully identifying the links between various properties of the terrain and DEM error values. The obtained accuracy improvement demonstrate that reliable DEM correction can be done even under limited ground survey points.

The significance of the sampling strategy used for collecting ground data was one of the key findings of this study. Stratified sampling showed the highest accuracy rates in relation to all other types of samples considered, demonstrating the lowest RMSE and the highest R^2 . This can be explained by the necessity to create the training dataset that would cover the diversity of the study area and reflect its varied terrain characteristics. Clustered sampling resulted in relatively low and inconsistent accuracy due to the high clustering effect that was observed for small-sized samples while random sampling showed stable improvement with increasing sample size which indicates a reasonable representative training performance.

The difference between clustered sampling and the other strategies was more pronounced. The low-slope concentrated strategy produced the highest RMSE at most sample sizes and showed a clear performance plateau after approximately 300 observations. At 500 points, its RMSE remained at 2.632m, compared with 2.211m for random sampling and 2.173m for stratified sampling. The statistical tests further confirmed these differences, with clustered sampling significantly underperforming stratified sampling at all sample sizes and random sampling from 100 points onward. This suggests that concentrating observations within a restricted terrain condition can limit the ability of the RF model to learn error relationships across the broader terrain variability of the study area (Wang et al., 2012; Li & Heap, 2014).

The sample-size analysis also provides evidence that increasing the number of observations generally improves DEM error prediction. Both random and stratified sampling showed progressively lower RMSE values as sample size increased. Nevertheless, the magnitude of improvement became progressively smaller at higher sample sizes. For example, stratified sampling improved RMSE from 2.431 m at 300 points to 2.280 m at 400 points and then to 2.173 m at 500 points**. Thus, although additional observations continued to improve the mean performance, the practical gain from each additional group of observations became smaller (Li & Heap, 2014).

Clustered sampling behaved differently. Its RMSE improved from 3.332 m at 50 points to 2.751 m at 300 points, followed by smaller improvements to 2.654 m at 400 points and 2.632 m at 500 points. This indicates a clear reduction in the rate of performance improvement at larger sample sizes. Increasing the number of observations within the same restricted low-slope environment therefore appears insufficient to achieve the performance obtained through random or stratified sampling, highlighting the importance of representing broader terrain variability in the training dataset.

This study further identified the sampling threshold between 300 and 400 data points after which further increases in sample size provided minor improvements in accuracy. Thus, even moderately sized but properly distributed training sets could be considered adequate for DEM correction, which makes it possible to reduce the costs associated with field data acquisition.

Variogram analysis confirmed the effectiveness of the proposed correction framework by showing reduced spatial autocorrelation of DEM errors after correction. Lower sill values observed in the corrected DEM indicate better spatial consistency and reduced systematic elevation bias. Furthermore, the strong agreement between corrected DEM elevations and independent validation elevations showed the reliability of proposed framework.

The visual comparison between the original and corrected SRTM DEM further illustrated the effectiveness of proposed framework. The corrected DEM showed smoother elevation changes and fewer localized elevation issues than the original DEM. These improvements confirm the capability of the GeoAI-based framework to enhance the DEM accuracy under limited ground data condition while these findings indicate that spatial representativeness of training data plays a critical role in DEM correction accuracy.

Conclusions

This study evaluated the influence of sampling strategies and sample sizes on Geo AI based DEM error correction. The proposed RF based error correction model improved the SRTM DEM accuracy under all evaluated sampling conditions. The results clearly demonstrated the significant influence of the sample size and sampling strategy for acquiring ground elevations on the DEM error correction. It was clearly identified that increasing the number of ground elevation points generally improved the model accuracy. Among the evaluated sampling strategies, stratified sampling produced optimal results for all sample sizes highlighting the importance of terrain representative data. In the meantime, clustered sampling showed lower predictive accuracy due to the limited spatial representation of terrain variability. Geo-statistical analysis further confirmed the effectiveness of the developed GeoAI model by showing the reduction in spatially auto correlated SRTM DEM errors after the correction. Independent validation using the reserved ground survey points further showed a strong agreement between the corrected DEM elevations by the model and the ground truth data.

Overall, the findings highlight the significant role of spatial sampling design in ML based DEM error correction and demonstrate the applicability of GeoAI approaches for improving the accuracy of the freely available DEMs with limited ground data conditions. This framework was evaluated in a relatively small area using terrain derived variables only. Therefore, it is recommended to explore the applicability of proposed framework across larger and more complex landscapes while incorporating with more environmental and terrain related variables..

Acknowledgements

Authors extend their gratitude to the Department of Surveying and Geodesy, Faculty of Geomatics, Sabaragamuwa University of Sri Lanka for providing required ground survey data and to all online platforms and organizations that provide open access geospatial data for this research.

Declaration of Funding Sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest statement

The authors declare no conflict of interest.

Data Availability Statement

The data used in this study are available from the corresponding author upon reasonable request.

Author Contribution

All authors accepted responsibility for the content of this paper, reviewed the findings, and approved the final version for submission.

References

- Arun, P. V. (2013). A comparative analysis of different DEM interpolation methods. *Geodesy and Cartography*, 39(4), 171–177. <https://doi.org/10.3846/20296991.2013.859821>
- Azaad, S. (2026). *Computing Cohen's dz from commonly reported statistics: A practical guide for the meta-analysis of paired samples mean differences* [Preprint]. OSF Preprints. https://doi.org/10.31234/osf.io/7gnj6_v2
- Chen, C., Liu, Y., Li, Y., & Chen, D. (2024). Explainable artificial intelligence framework for urban global digital elevation model correction based on the SHapley additive explanation-random forest algorithm considering spatial heterogeneity and factor optimization. *International Journal of Applied Earth Observation and Geoinformation*, 129, 103843. <https://doi.org/10.1016/j.jag.2024.103843>
- Guenther, E. J., Neuenschwander, A. L., Magruder, L. A., & Maze-England, D. (2024). Correcting SAR-derived DEMs with ICESat-2 using deep learning. In M. D. Turner, G. W. Kamerman, & L. A. Magruder (Eds.), *Laser Radar Technology and Applications XXIX* (Vol. 13049, Article 1304907). SPIE. <https://doi.org/10.1117/12.3013897>
- Hermans, L. D., Davis, P. T., McCord, M. J., & Bidanset, P. E. (2025). The effects of Monte Carlo sampling on automated valuation model performance in real estate assessment. *Journal of Property Tax Assessment & Administration*, 22(1), 29–50. <https://doi.org/10.63642/3067-4816.1273>
- Kandil, W. M., Zarzoura, F. H., Salah, M. M., Shaaban, M. M., & El-Mewafi, M. M. (2025). Machine learning algorithms with enhanced features for reconstructing high-resolution urban DEMs. *Journal of Integrated Engineering and Technology*. <https://ijet.journals.ekb.eg/article/410009>
- Li, J., & Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, 53, 173–189. <https://doi.org/10.1016/j.envsoft.2013.12.008>

- Midway, S., Robertson, M., Flinn, S., & Kaller, M. (2020). Comparing multiple comparisons: Practical guidance for choosing the best multiple comparisons test. *PeerJ*, 8, e10387. <https://doi.org/10.7717/peerj.10387>
- Okolie, C., Adeleke, A., Smit, J., Mills, J., Ogbeta, C., & Maduako, I. (2024). Performance analysis of Bayesian optimised gradient-boosted decision trees for digital elevation model (DEM) error correction: Interim results. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-2-2024, 179–183. <https://doi.org/10.5194/isprs-annals-X-2-2024-179-2024>
- Preety, K., Prasad, A. K., Varma, A. K., & El-Askary, H. (2022). Accuracy assessment, comparative performance, and enhancement of public domain digital elevation models (ASTER 30 m, SRTM 30 m, CARTOSAT 30 m, SRTM 90 m, MERIT 90 m, and TanDEM-X 90 m) using DGPS. *Remote Sensing*, 14(6), 1334. <https://doi.org/10.3390/rs14061334>
- Rogerson, P. A. (2025). Testing hypotheses when you have more than a few. *Geographical Analysis*, 57(2), 175–190. <https://doi.org/10.1111/gean.12412>
- Tian, X., & Shan, J. (2025). Learning-based spatial interpolation for sparse space altimetry measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 1–14. <https://doi.org/10.1109/TGRS.2025.3643406>
- Wang, J., Ge, Y., Heuvelink, G. B. M., Zhou, C., & Brus, D. J. (2012). Effect of the sampling design of ground control points on the geometric correction of remotely sensed imagery. *International Journal of Applied Earth Observation and Geoinformation*, 18, 91–100. <https://doi.org/10.1016/j.jag.2012.01.001>
- Zhu, C., Tang, F., Yang, Q., Li, J., Xue, J., Su, Y., & Jia, Y. (2025). An adaptive ground-point-cloud misclassification detection and correction algorithm based on seed-point guidance for high-precision DEM construction in complex terrain. *IEEE Sensors Journal*, 25(21), 40399–40412. <https://doi.org/10.1109/JSEN.2025.3615605>