Performance analysis of Bayesian optimised gradient-boosted decision trees for digital elevation model (DEM) error correction: interim results

Chukwuma Okolie[1,2,3*], Adedayo Adeleke[4], Julian Smit[5], Jon Mills[3], Caleb Ogbeta[6] and Iyke Maduako[7]

[1]Division of Geomatics, University of Cape Town, South Africa; chukwuma.okolie@alumni.uct.ac.za
[2]Department of Surveying & Geoinformatics, University of Lagos, Nigeria
[3]School of Engineering, Newcastle University, United Kingdom; jon.mills@newcastle.ac.uk
[4]Department of Geography, Geoinformatics and Meteorology, University of Pretoria, South Africa; adedayo.adeleke@up.ac.za
[5]Department of Civil Engineering and Geomatics, Cape Peninsula University of Technology, South Africa; drjlsmit@gmail.com
[6]Geomatics Lab, School of Civil and Construction Engineering, Oregon State University, USA; ogbetac@oregonstate.edu
[7]Department of Geoinformatics and Surveying, University of Nigeria, Nsukka, Nigeria; iykemadu84@gmail.com

Commission II/WG4

Keywords: Digital Elevation Model, Copernicus, Bayesian optimisation, Gradient boosted decision trees, Machine learning, Hyperparameter tuning.

Abstract

Gradient-Boosted Decision Trees (GBDTs), particularly when tuned with Bayesian optimisation, are powerful machine learning techniques known for their effectiveness in handling complex, non-linear data. However, the performance of these models can be significantly influenced by the characteristics of the terrain being analysed. In this study, we assess the performance of three Bayesian-optimised GBDTs (XGBoost, LightGBM and CatBoost) using digital elevation model (DEM) error correction as a case study. The performance of the models is investigated across five landscapes in Cape Town South Africa: urban/industrial, agricultural, mountain, peninsula and grassland/shrubland. The models were trained using a selection of datasets (elevation, terrain parameters and land cover). The comparison entailed an analysis of the model execution times, regression error metrics, and level of improvement in the corrected DEMs. Generally, the optimised models performed considerably well and demonstrated excellent predictive capability. CatBoost emerged with the best results in the level of improvement recorded in the corrected DEMs, while LightGBM was the fastest of all models in the execution time for Bayesian optimisation and model training. These findings offer valuable insights for applying machine learning and hyperparameter tuning in remote sensing.

1. Introduction

In recent years, Digital Elevation Models (DEMs) have become indispensable tools in a wide range of applications, from geographic information systems and environmental modelling to urban planning and civil engineering (Musa et al., 2015; Muhadi et al., 2020). DEMs provide a simplified representation of the Earth's varying topography, but they are not without errors (Wechsler, 2007; Elkhrachy, 2018). These errors can arise from various sources, including data collection methods, processing algorithms, or the inherent limitations of the technologies used (Hugonnet et al., 2022). Accurate prediction and correction of these errors is crucial for the reliability and usefulness of DEMs in environmental and hydrological applications.

The advent of machine learning (ML) has opened new avenues for enhancing the accuracy of DEMs. Among various ML techniques, Gradient Boosted Decision Trees (GBDTs) have emerged as a powerful tool for handling complex, non-linear spatial data (Wang et al., 2021; Wen et al., 2022). GBDT models are particularly adept at capturing intricate patterns in data, making them well-suited for modelling the errors in DEMs. However, the performance of GBDT models can be significantly influenced by their hyperparameters, which necessitates sophisticated optimization techniques (Ke et al., 2017).

Bayesian Optimisation (BO) has gained prominence as an effective method for hyperparameter tuning of ML models (Yang & Shami, 2020; Kavzoglu & Teke, 2022). By employing a probabilistic model, Bayesian Optimization efficiently navigates the hyperparameter space to find optimal settings. This optimisation process is crucial for enhancing the performance of GBDT models in predicting DEM errors. However, the performance of BO-optimised models, especially when dealing with diverse and complex terrain is still open to investigation.

Performance assessments are important for measuring the computational efficiency of machine leaning algorithms. Scientists, researchers, and industry practitioners are usually interested in the computational efficiency of machine learning algorithms before deploying them on a large scale. For example, an algorithm with faster training time is more efficient, while algorithms with faster prediction times are preferable (e.g. in real-time applications) (Pushp, 2023). Moreover, an algorithm with better accuracy is preferable for many applications in remote sensing. Bayesian optimisation has been proven as an effective technique for improving the prediction accuracy of GBDTs, for example in DEM correction (e.g. Okolie et al. 2023a; 2023b). However, to our knowledge, the relative performances of Bayesian-optimised models when applied to DEM correction in diverse landscapes has not yet been investigated.

In this study, we compare the performance of three Bayesian-optimised gradient boosted trees: the Extreme gradient boosting (XGBoost), Light boosting machine (LightGBM) and Categorical boosting (CatBoost) in a DEM correction use case. The test sites are drawn from various landscapes and this could reveal new insights on the influence of terrain complexity on model performance. This comparison could also facilitate the selection of the most efficient model for similar applications by other researchers, and also contributes to the ever-expanding machine learning applications in the remote sensing body of knowledge.

*Corresponding author

2. Methodology

2.1 Study Area and Datasets

Cape Town is South Africa’s most south-western city, and is situated on the south-western coast of the Western Cape Province. The town is topographically and geomorphologically diverse. Five different landscapes were selected for this study: urban/industrial, agricultural, mountain, peninsula and grassland/shrubland. The topography of a section of Cape Town is shown in Figure 1. The Copernicus GLO-30 DEM (Airbus, 2020a, 2020b) is the main dataset for this study, while the City of Cape Town (CCT) airborne LiDAR-derived DEM is used as the reference dataset. The CCT DEM was derived from a point cloud with height accuracy of 0.15 m. The Copernicus DEM and CCT DEM were harmonised to the same vertical datum (EGM2008), and the elevation differences (ΔH) between both DEMs were calculated. To characterise the influence of the terrain on the elevation error, the following additional input variables were generated from the Copernicus DEM elevations: slope, aspect, surface roughness, topographic position index (TPI), terrain ruggedness index (TRI), terrain surface texture (TST), and vector ruggedness measure (VRM). Additionally, the percentage forest canopy and percentage bare ground cover data were acquired from the Global Land Analysis and Discovery (GLAD) data portal (Hansen et al., 2013), while building footprints were extracted from the Global Urban Footprints (GUF) dataset (Esch et al., 2010, 2013).

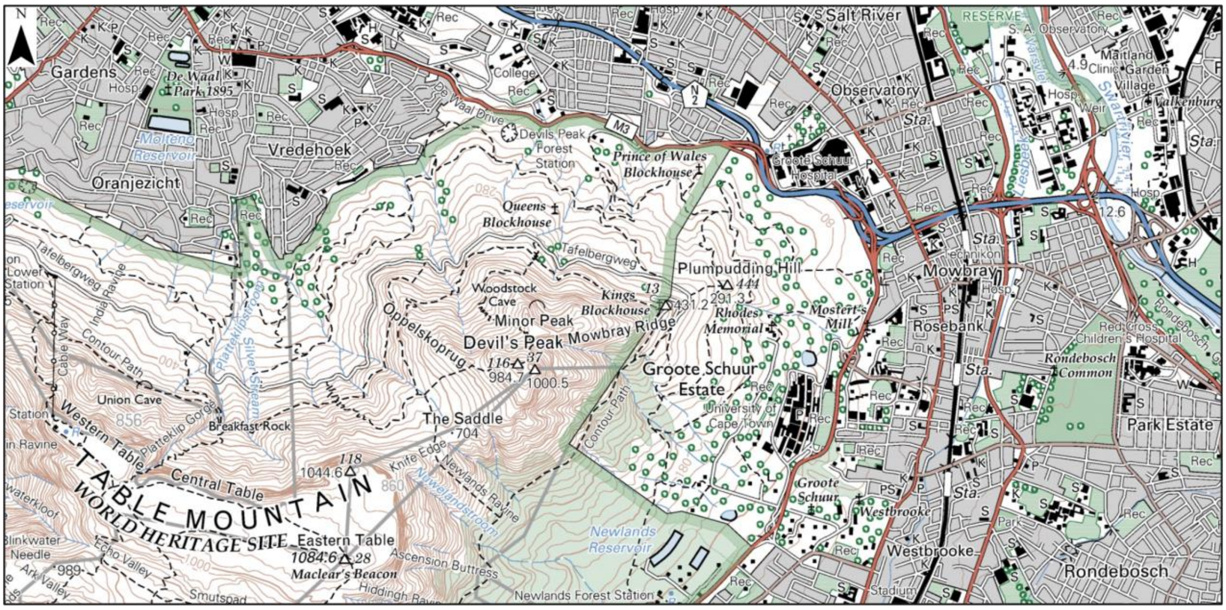

Figure 1. A topographic map of a section of Cape Town showing urban areas, vegetation and the imposing Table Mountain. (Source: 1:50,000 South Africa topographic map series, published by the Chief Directorate: National Geo-spatial Information, © 2015)

2.2 Gradient Boosted Decision Trees

Gradient Boosted Decision Trees (GBDTs) are very powerful for handling a myriad of classification and regression problems and can attain excellent results in a variety of applications. The extreme gradient boosting (XGBoost), light boosting machine (LightGBM) and categorical boosting (CatBoost) are recent implementations of gradient boosting that have revolutionised the machine learning community.

Extreme gradient boosting (XGBoost) is a “scalable end-to-end tree boosting system” that has excelled in several machine learning tasks (Chen & Guestrin, 2016). LightGBM (Ke et al., 2017) overcomes several limitations of previous engineering optimisations used in GBDTs. LightGBM is “mainly featured by the decision tree algorithm based on gradient-based one-side sampling (GOSS), exclusive feature bundling (EFB), and a histogram and leaf-wise growth strategy with a depth limit” (Li et al., 2022). With GOSS and EFB, the speed of training is increased (Safaei et al., 2022). Categorical Boosting (CatBoost) was introduced in 2018, and is well-suited for handling categorical and heterogenous data (Hancock & Khoshgoftaar, 2020). Two important innovations were introduced in CatBoost: ordered boosting, and the processing of categorical features using a novel algorithm (Prokhorenkova et al., 2018).

2.3 Bayesian Optimisation

Bayesian optimisation (BO) enables the construction of optimal programs for computing better solutions (Hoos, 2010; Shahriari et al., 2016). With BO, the performance of ML algorithms can be improved through hyperparameter tuning. A hyperparameter is “a parameter whose value is given by the user and used to control the learning process” (Mariani & Sipper, 2022). Their values “control the learning process and determine the values of model parameters that a learning algorithm ends up learning” (Nyuytiymbiy, 2020). Essentially, BO “builds a probability model of the objective function to determine the optimal hyperparameters in an informed manner, reducing the number of times the objective function needs to be run by choosing only the most promising set of hyperparameters” (Habtemariam et al., 2023).

2.4 Data Processing and Analysis

The three GBDT models were trained using the elevation, slope, aspect, surface roughness, topographic position index, terrain ruggedness index, terrain surface texture, vector ruggedness measure, percentage forest canopy, percentage bare ground cover and urban footprints as input parameters, and the elevation error as the target variable. The attributes of the urban footprints had numeric codes representing the urban and non-urban areas. The data points used for model training, validation and testing encompass a total area of approximately 1579.6km^2 (the area of each landscape is shown in Table 1). The data points were split into separate subsets: 80% for training and validation, and 20%

Landscape	Training/ validation/ test area (km^2)	No. of iterations for BO	Execution time for BO (s)		
			XGBoost	LightGBM	CatBoost
Urban/ industrial	532.9	50	17959.282	1033.409	6157.880
Agricultural	509.0	40	26080.685	1691.126	14862.721
Mountain	280.6	50	2911.410	295.583	4548.114
Peninsula	135.2	50	2486.832	167.138	4102.972
Grassland/ shrubland	121.9	50	2976.456	151.293	4267.270

Table 1. Execution time for Bayesian optimisation (BO) of XGBoost, LightGBM and CatBoost

Landscape	XGBoost		LightGBM		CatBoost	
	Training (s)	Prediction (s)	Training (s)	Prediction (s)	Training (s)	Prediction (s)
Urban/ industrial	291.885	0.399	20.045	2.603	916.761	0.457
Agricultural	1371.493	3.900	74.576	5.545	1860.448	0.911
Mountain	106.034	0.221	4.877	0.164	1824.249	0.156
Peninsula	135.030	0.268	1.601	0.084	290.231	0.059
Grassland/ shrubland	88.611	0.163	1.208	0.108	489.858	0.101

Table 2. Execution time for training and prediction of XGBoost, LightGBM and CatBoost

Landscape	RMSE (m)		
	XGBoost	LightGBM	CatBoost
Urban/ industrial	0.859	0.935	0.857
Agricultural	0.601	0.860	0.602
Mountain	1.883	1.934	1.856
Peninsula	1.138	1.195	1.145
Grassland/ shrubland	0.838	0.968	0.842

Table 3. Comparison of XGBoost, LightGBM and CatBoost test error

for testing. For the hyperparameter tuning, Bayesian optimisation was adopted.

The optimisation for XGBoost, LightGBM and CatBoost was meticulously implemented using 10 steps of random exploration, and depending on the available computing power, 40 - 50 steps of optimisation. For Copernicus DEM in the urban/industrial, mountain, peninsula and grassland/shrubland landscapes, 50 iterations were executed for the Bayesian optimisations, while 40 iterations were executed in the agricultural landscape. Nine hyperparameters were selected for each model, for example:

- The number of decision trees/boosting rounds: range (1, 50000)
- Learning rate: range (0.0001, 1)
- The depth of the tree: XGBoost and LightGBM (range (1, 50)); CatBoost (range (1, 16))

After training and testing, the models were implemented for predicting the DEM errors at two separate implementation sites with similar terrain characteristics. The predicted elevation errors were applied for deriving corrected DEMs (i.e., $DEM_{Corrected} = DEM_{Original} - \Delta H_{Predicted}$) at the implementation sites. The assessed models were compared based on the execution time and accuracy of predictions.

3. Results and Discussion

Table 1 shows the Bayesian optimisation execution time (in seconds) of XGBoost, LightGBM and CatBoost for Copernicus DEM. LightGBM was several times faster than XGBoost and CatBoost. For example, in the Cape Peninsula, LightGBM was 15x faster than XGBoost (167.138 s vs 2486.832 s) and 25x faster than CatBoost (167.138 s vs 4102.972 s) respectively. Similarly, in the grasslands/shrublands, it was 20x faster than XGBoost (151.293 s vs 2976.456 s), and 28x faster than CatBoost (151.293 s vs 4267.270 s) respectively. The optimisation process consumed a considerable amount of time and system resources, especially in the urban/industrial areas and agricultural lands. XGBoost was slower than CatBoost in the optimisation time in the urban/industrial and agricultural lands, but was faster in the mountain, peninsula and grassland/shrubland areas. However, the number of training samples in the urban and agricultural lands exceeded that of other landscapes by several hundreds of thousands.

Table 2 shows the execution time comparisons for model training and prediction of Copernicus DEM. LightGBM is known for its faster training speed, high efficiency, lower memory usage and capability to handle large-scale data (LightGBM Documentation, 2024). This explains its very fast training speed in all landscapes, far ahead of XGBoost and CatBoost.

Table 3 compares the test root mean square error (RMSE) for XGBoost, LightGBM and CatBoost respectively. Both XGBoost and CatBoost had lower RMSEs than LightGBM. Generally, the optimised models performed considerably well and demonstrated excellent predictive capability. Test RMSEs were lowest in the agricultural lands (0.601 – 0.860 m) and highest in the mountains (1.856 – 1.934 m). In a related study, Bentéjac et al compared the training speed, generalisation performance and hyperparameter setup of several GBDTs. In their results, CatBoost emerged with the best results in generalisation accuracy while LightGBM was the fastest of all models, although not the most accurate.

Finally, the predicted errors by XGBoost, LightGBM and CatBoost were applied to correct the original Copernicus DEM at two external implementation sites. Figures 2 and 3 show the percentage reduction in the vertical RMSE of the original Copernicus DEM at the first and second implementation sites

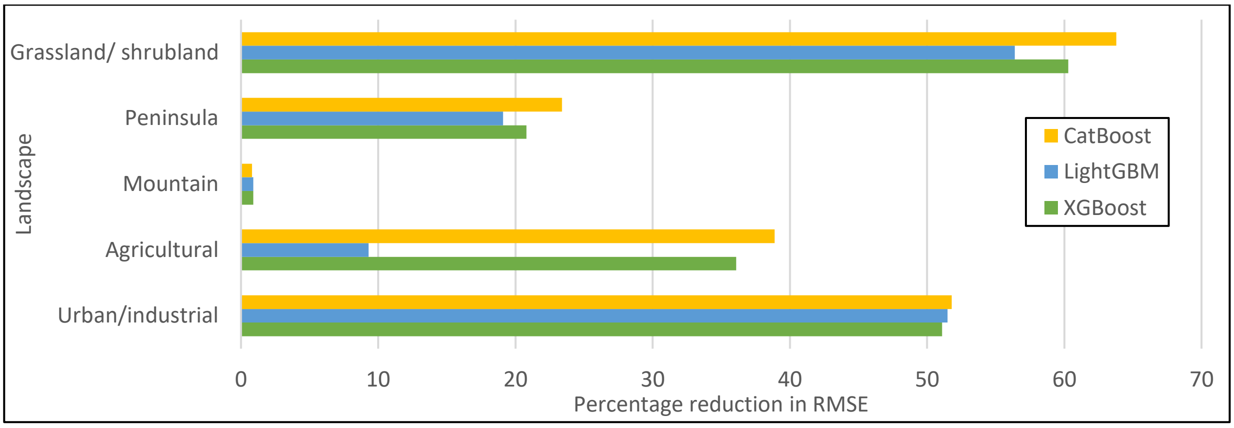

Figure 2. Percentage reduction in vertical RMSE of the original Copernicus DEM at the first implementation site

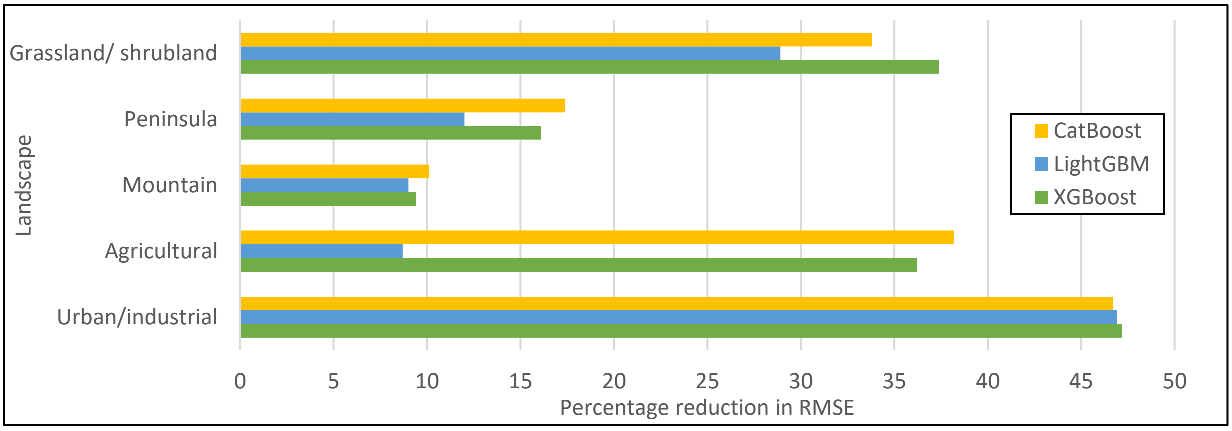

Figure 3. Percentage reduction in vertical RMSE of the original Copernicus DEM at the second implementation site

respectively. Generally, CatBoost achieved the highest reduction in RMSE of original DEM (the greatest improvement in DEM accuracy) in most landscapes, followed by XGBoost and LightGBM respectively. The highest improvements in DEM accuracy occurred in the urban/industrial, agricultural and grassland/shrubland landscapes.

The results suggest that while Bayesian optimisation significantly enhances the accuracy of these models, it also imposes a considerable computational burden, especially in data-rich environments like urban/industrial and agricultural lands.

4. Conclusion

The aim of this study was to compare the performance of three Bayesian-optimised gradient boosted trees: XGBoost, LightGBM and CatBoost in a DEM correction use case, with test sites spread across five landscapes in Cape Town, South Africa (urban/industrial, agricultural, mountain, peninsula and grassland/shrubland). The results have yielded critical insights into the balance between computational efficiency and model accuracy. The comparative analysis of XGBoost, LightGBM and CatBoost under various terrain conditions has highlighted significant differences in performance. This finding is crucial for applications where time efficiency is as important as model accuracy, such as real-time environmental monitoring and rapid response scenarios.

In practical terms, the study suggests a necessary trade-off between the expected accuracy improvements from hyperparameter optimisation and the associated computational time and resources. For instance, while the optimised CatBoost model shows a marked slowdown across various landscapes, the accuracy gains might justify this trade-off in scenarios where precision is paramount. Conversely, in situations where speed is critical, the relatively more efficient LightGBM, even in its optimised form, might be the preferred choice. The accuracy improvements achieved through hyperparameter optimisation could have tremendous benefits for the use of the DEMs in hydrological modelling and other environmental applications. This finding is particularly relevant for researchers and practitioners in the field of geospatial analysis, where the choice of an appropriate machine learning model can significantly impact the efficiency and effectiveness of their work.

In conclusion, this research contributes valuable knowledge to the field of DEM analysis, particularly in the application of machine learning models for DEM error correction. It underscores the importance of carefully considering the specific requirements of each application, balancing the need for accuracy with the constraints of computational resources. The insights gained from this study are expected to guide future research and practical applications in DEM analysis and terrain modelling, and other environmental studies. This study not only contributes to the advancement of geospatial analysis using machine learning but also could aid in optimising DEM applications in various fields such as urban planning, environmental monitoring, and navigation systems. The conclusions drawn from this research highlight the potential of Bayesian-Optimized GBDT models in handling geospatial data efficiently. Future research can compare the performance of gradient boosted trees with deep learning approaches.

Acknowledgements

Special thanks to the Commonwealth Scholarship Commission UK, and the University of Cape Town Postgraduate Funding Office for funding support for this research. LIDAR data for the City of Cape Town was provided by the Information and Knowledge Management Department, City of Cape Town. Also, we appreciate the assistance of Professor Jennifer Whittal, Dr Hossein Bagheri, Tom Komar and Chima Iheaturu.

References

Airbus. 2020a. *Copernicus DEM Copernicus Digital Elevation Model Product Handbook*.

Airbus. 2020b. *Copernicus DEM Copernicus Digital Elevation Model Validation Report*.

Chen, T., Guestrin, C. 2016. XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-August-2016, 785–794. https://doi.org/10.1145/2939672.2939785

Habtemariam, E. T., Kekeba, K., Martínez-Ballesteros, M., Martínez-Álvarez, F. 2023. A Bayesian Optimization-Based LSTM Model for Wind Power Forecasting in the Adama District, Ethiopia. *Energies*, 16(5), 2317. https://doi.org/10.3390/EN16052317

Hancock, J. T., Khoshgoftaar, T. M. 2020. CatBoost for big data: an interdisciplinary review. *Journal of Big Data*, 7(1), 1–45. https://doi.org/10.1186/S40537-020-00369-8/FIGURES/9

Hoos, H. H. 2010. *Programming by Optimisation*. University of British Columbia Department of Computer Science Technical Report TR-2010-14. https://www.cs.ubc.ca/~hoos/Publ/Hoos10.pdf

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y. 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems*, 30. https://github.com/Microsoft/LightGBM.

Li, K., Xu, H., Liu, X. 2022. Analysis and visualization of accidents severity based on LightGBM-TPE. *Chaos, Solitons & Fractals*, 157, 111987. https://doi.org/10.1016/J.CHAOS.2022.111987

Mariani, S., Sipper, M. 2022. High Per Parameter: A Large-Scale Study of Hyperparameter Tuning for Machine Learning Algorithms. *Algorithms 2022, Vol. 15, Page 315*, 15(9), 315. https://doi.org/10.3390/A15090315

Nyuytiymbiy, K. 2020. *Parameters and Hyperparameters in Machine Learning and Deep Learning*. Towards Data Science. https://towardsdatascience.com/parameters-and-hyperparameters-aa609601a9ac

Okolie, C., Mills, J., Adeleke, A., Smit, J. 2023. *Digital elevation model correction in urban areas using extreme gradient boosting, land cover and terrain parameters*. https://arxiv.org/abs/2308.06545v1

Okolie, C., Mills, J., Adeleke, A., Smit, J., Maduako, I. 2023. The explainability of gradient-boosted decision trees for digital elevation model (DEM) error prediction. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-M-3–2023(M-3–2023), 161–168. https://doi.org/10.5194/ISPRS-ARCHIVES-XLVIII-M-3-2023-161-2023

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., Gulin, A. 2018. Catboost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 2018-Decem, 6638–6648. https://github.com/catboost/catboost

Safaei, N., Safaei, B., Seyedekrami, S., Talafidaryani, M., Masoud, A., Wang, S., Li, Q., Moqri, M. 2022. E-CatBoost: An efficient machine learning framework for predicting ICU mortality using the eICU Collaborative Research Database. *PLOS ONE*, 17(5), e0262895. https://doi.org/10.1371/JOURNAL.PONE.0262895

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., De Freitas, N. 2016. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1), 148–175. https://doi.org/10.1109/JPROC.2015.2494218