

Machine learning approaches for predicting land subsidence in Ca Mau: XGBoost, Random Forest, and MAF

HUYNH Nguyen Dinh Quoc¹, NGUYEN Gia Trong^{2,3}, HA Trung Khien⁴,
LE Duc Tinh², TRAN Van Anh^{2,*}

¹ Ho Chi Minh City University of Natural Resources and Environment, Vietnam

² Hanoi University of Mining and Geology, Vietnam

³ Geodesy and Environment research group, Hanoi University of Mining and Geology, Vietnam

⁴ Department of Geodesy, Hanoi University of Civil Engineering.

* Corresponding author: tranvananh@humg.edu.vn

Abstract: Land subsidence is a natural hazard that is causing serious damage to the Mekong Delta (VMD) of Vietnam, with subsidence rates of up to 5 cm per year in densely populated cities such as Ca Mau, increasing the risk of salinity intrusion and tidal flooding. These ground movements not only amplify the impacts of sea level rise but also threaten infrastructure, agricultural sustainability, and long-term climate resilience. While traditional monitoring methods such as GNSS and land leveling surveys are highly accurate, they are often spatially inadequate and cost-ineffective for regional-scale applications. In this context, remote sensing technologies, specifically Interferometric Synthetic Aperture Radar (InSAR) techniques such as Persistent Scatter Interferometry (PSI), have emerged as powerful tools for understanding surface deformation patterns over large areas. Integrating InSAR-derived observations with machine learning (ML) techniques offers new opportunities for predictive modeling of subsidence phenomena. Ensemble algorithms such as Random Forest (RF) and Extreme Gradient Boosting (XGBoost) have demonstrated robust performance in identifying spatially distributed land deformation susceptibility, especially when applied to high-dimensional geospatial data. In this study, we evaluate the predictive performance of three distinct methods - MAF (Moving Average Filter), RF, and XGBoost - for predicting land subsidence in Ca Mau using PSI-based displacement data. A dataset of 5,000 deformation monitoring points from Sentinel-1 imagery from 2014 to 2019 is used to train and evaluate the models. Among these models, XGBoost demonstrated the best performance with the lowest RMSE (4.67) and MAE (3.23), and the highest R^2 (0.9869), significantly outperforming both RF and MAF. These findings highlight the robustness of machine learning approaches, particularly XGBoost, in predicting land subsidence and supporting sustainable land use planning and climate adaptation strategies in vulnerable deltaic environments.

Keywords: XGBoost, Random Forest, Filter Average Moving, Land Subsidence

1. Introduction

Land subsidence has emerged as a critical environmental hazard in low-lying deltaic regions worldwide, especially where population growth, groundwater exploitation, and climate change converge. The Vietnamese Mekong Delta (VMD), home to over 17 million people, is particularly vulnerable due to its flat terrain, soft Holocene sediments, and intensive groundwater use. In recent decades, the VMD has experienced widespread subsidence, with observed rates of 1 - 4 cm/year in many locations and up to 5 cm/year in hotspot urban centers like Ca Mau and Can Tho (de Wit et al., 2021). Over-extraction of groundwater for agriculture, domestic, and industrial use has significantly lowered groundwater tables, resulting in compaction of aquifer systems and widespread land deformation (de Wit et al., 2021; Erban, Gorelick, & Zebker, 2014). These subsidence patterns not only amplify local relative sea-level rise but also severely threaten infrastructure, increase flood risks, and undermine climate adaptation strategies in coastal provinces (Cao et al., 2021).

Remote sensing technologies, particularly Interferometric Synthetic Aperture Radar (InSAR), have become essential tools for quantifying vertical land motion over large spatial scales (de Wit et al., 2021; Tran et al., 2024). Combined with field measurements and geotechnical data, InSAR facilitates precise identification of differential subsidence and supports predictive modeling. Recently, machine learning methods - especially ensemble algorithms like Extreme Gradient Boosting (XGBoost) - have shown high accuracy and robustness in modeling spatial susceptibility to land subsidence based on multi-source

geospatial data (Nalluri, Pentela, & Eluri, 2020; Tran et al., 2024). A study in Ca Mau using boosting algorithms achieved prediction accuracy above 88% and highlighted the effectiveness of integrating Persistent Scatterer Interferometry (PSI) data with boosting models to generate high-resolution susceptibility maps (Tran et al., 2024). These advancements underscore the value of hybrid approaches that combine satellite data with advanced analytics, offering a powerful pathway for supporting proactive land use planning and sustainable groundwater management in vulnerable deltaic environments.

Literature reviews indicate that studies on land subsidence from a geohazard perspective are typically conducted using multidisciplinary approaches across various fields such as geology, hydrogeology, geotechnical engineering, remote sensing, and urban risk management (Bagheri-Gavkosh et al., 2021). Land subsidence has emerged as a major environmental and geotechnical concern in many regions worldwide, particularly in deltaic and semiarid environments where both natural and anthropogenic factors contribute to surface deformation. Among the primary drivers, groundwater over-extraction has been consistently identified as a key contributor to land subsidence, especially in urbanizing and agriculturally intensive regions such as the Mekong Delta and the Bakhtegan Basin (Gharechae et al., 2023; Tran et al., 2024). The consequences are far-reaching, leading to structural damage, loss of aquifer storage capacity, saltwater intrusion, and heightened flood risks (Erban et al., 2014).

Traditionally, land subsidence has been measured using leveling or Global Navigation Satellite System (GNSS) techniques, which offer high precision but are often constrained by cost and limited spatial coverage. As an alternative, radar interferometry techniques, particularly Persistent Scatterer Interferometry (PS-InSAR) and Small Baseline Subset (SBAS), have become widely adopted for efficient, large-scale subsidence monitoring (Delgado Blasco, Fomelis, Stewart, & Hooper, 2019; Erban et al., 2014). In urban areas like Rome and Jakarta, these techniques, supported by open-source tools such as SNAP and StaMPS, have been instrumental in identifying spatial and temporal deformation trends (Delgado Blasco et al., 2019; Tran et al., 2024). For instance, the use of PS-InSAR combined with Sentinel-1 imagery enabled the SARPROZ platform to effectively monitor land subsidence trends in the Rawalpindi–Islamabad metropolitan region, where rapid urban expansion and groundwater dependency were the primary stressors (Khan, Ren, Hussain, & Jan, 2022).

In response to the increasing complexity of land subsidence dynamics, machine learning (ML) and deep learning (DL) techniques have gained prominence. Recent studies introduced improved combined scatterers interferometry (ICOPS) in tandem with hybrid DL models such as Convolutional Neural Network optimized by Grey Wolf Optimizer (CNN-GWO), which significantly enhanced measurement point density and model accuracy compared to conventional approaches (Hakim et al., 2023). Additionally, boosted regression trees (BRT) and extreme gradient boosting (XGBoost) have been effectively used to extrapolate full-coverage deformation maps from sparse InSAR datasets in groundwater-affected agricultural zones (Naghibi, Khodaei, & Hashemi, 2022). In Ca Mau province, Vietnam, various boosting algorithms were evaluated, with XGBoost demonstrating the highest accuracy in identifying subsidence-prone areas (Tran et al., 2024).

Other advanced modeling frameworks have also shown promising results. Artificial neural networks (ANN) and the Group Method of Data Handling (GMDH) were applied using InSAR-derived data and geo-environmental variables, with GMDH outperforming ANN in predictive accuracy (Panahi et al., 2022). Similarly, the Adaptive Neuro-Fuzzy Inference System (ANFIS) model, when optimized with meta-heuristic algorithms like the Imperialist Competitive Algorithm (ICA) and Grey Wolf Optimization (GWO), achieved even greater prediction performance using PS-InSAR datasets (Ranjgar, Razavi-Termeh, Foroughnia, Sadeghi-Niaraki, & Perissin, 2021).

Beyond model selection, the integration of diverse environmental and geospatial parameters has improved susceptibility mapping. Studies have used recursive feature elimination (RFE) to optimize input variable combinations, emphasizing the importance of conditioning factors such as geology, land use, groundwater depth, and proximity to geological features (Gharechae et al., 2023). In the Bakhtegan Basin, for instance, models like Random Forest (RF), k-Nearest Neighbors (KNN), and Classification and Regression Trees (CART) revealed that subsidence susceptibility was not solely linked to groundwater withdrawal but also to the proximity of faults, mines, and dams - challenging the dominant narrative of subsidence causality (Gharechae et al., 2023).

In other cases, DL techniques have been utilized in time-series forecasting. Long Short-Term Memory (LSTM) networks, when combined with GNSS-derived tropospheric indices such as Integrated Water Vapor and EvapoTranspiration, have demonstrated high accuracy in displacement prediction,

particularly in regions lacking dense meteorological networks (Tasan, Ghorbaninasab, Haji-Aghajany, & Ghiasvand, 2023). Elsewhere, in the Lake Urmia region, hybrid DL models - combining CNN, MLP, and LSTM - outperformed individual models in forecasting ground deformation, further validating the potential of ensemble approaches (Radman, Akhoondzadeh, & Hosseiny, 2021).

Lastly, LightGBM integrated with PS-InSAR data has been applied in urban areas like Shanghai to assess infrastructure risks, especially metro systems, offering a practical framework for urban geohazard mitigation (Chai, Xie, Wang, Tang, & Song, 2024). Despite these developments, no single method has consistently demonstrated superiority across all regions, highlighting the need for continued innovation in model transferability, feature optimization, and integration of multi-source datasets to build reliable and scalable subsidence prediction systems.

Although numerous studies have applied InSAR and machine learning techniques for land subsidence assessment worldwide ((Gharechaei et al., 2023; Hakim et al., 2023; Naghibi et al., 2022)), research focusing on the Vietnamese Mekong Delta remains limited. Recent work in Ca Mau demonstrated the potential of boosting algorithms for susceptibility mapping (Tran et al., 2024), but comparative evaluations of multiple models using PSI-derived displacement data are still scarce. This gap highlights the need for a systematic assessment of machine learning approaches tailored to the specific environmental and socio-economic conditions of the delta.

The objective of this study is to evaluate and compare the predictive capabilities of three models - Moving Average Filter (MAF), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) - for forecasting land subsidence in Ca Mau. Using PSI-based displacement data from Sentinel-1 imagery (2014 - 2019), the study aims to (i) assess the accuracy of each method, (ii) identify the most robust model for predicting spatial and temporal deformation, and (iii) provide insights to support sustainable land use planning and climate adaptation strategies in the vulnerable Mekong Delta region.

2. Data and Methodology

2.1 Study Area and Data

The Ca Mau Peninsula is a vast and relatively flat terrain. However, the collection of land subsidence monitoring data in this area remains limited in both frequency and spatial coverage. In particular, within Ca Mau urban area, the number of observation points is even scarcer. Furthermore, the absence of long-term periodic monitoring data poses challenges in obtaining a continuous time series of land subsidence.

For these reasons, we selected land subsidence data derived from the EMSN062 project, processed using PSI technique by BGR, Federal Office of Civil Protection and Disaster Assistance (BBK) on behalf of the German Corporation for International Cooperation GmbH (GIZ) and Federal Institute for Geosciences and Natural Resources (EMSN062, 2019).

The original dataset covers the entire Mekong Delta region; however, in this study, we extracted 5,000 subsidence points located around Ca Mau City. The data were processed using a multi-temporal radar approach based on Sentinel-1 imagery from 2014 to 2019.

The PSI-derived land subsidence points in this region have demonstrated satisfactory accuracy, validated against ground-based measurements, and further confirmed by other studies (Minderhoud, 2021; Tran et al., 2024). The locations of the subsidence monitoring points are shown in Fig. 1.

The overall computational workflow adopted in this study is illustrated in Fig. 2. The process begins with the acquisition of Sentinel-1 satellite imagery and the derivation of PSI-based surface deformation data. These data are then pre-processed and integrated with spatial and temporal features to form the training dataset. Subsequently, three predictive models - MAF, RF, and XGBoost - are developed and optimized for subsidence forecasting. The final step involves model validation using regression metrics and visualization of prediction deviations.

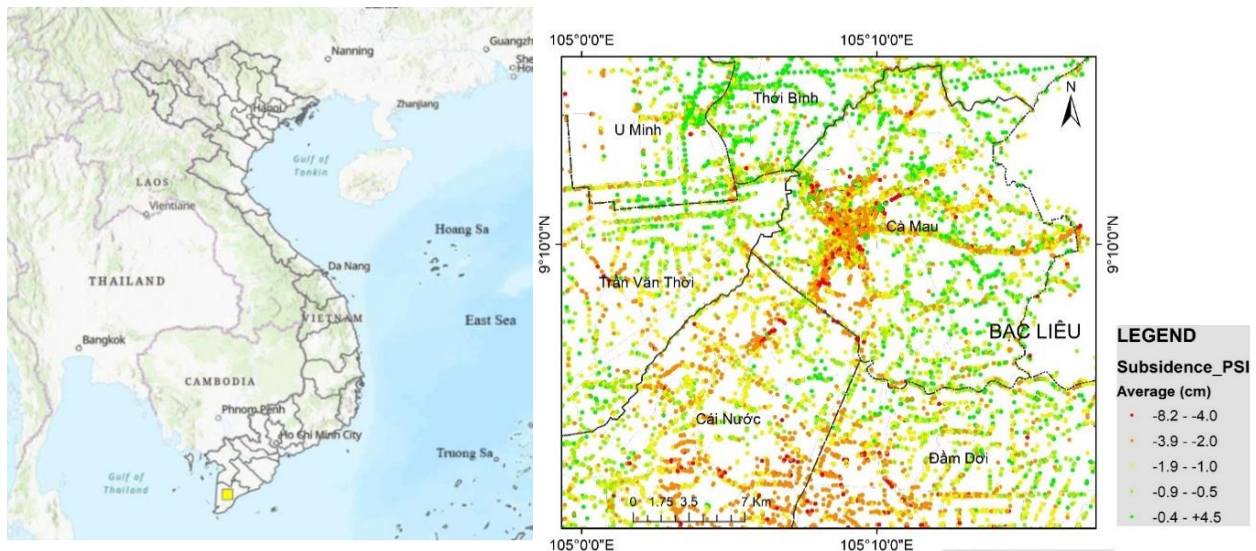


Fig. 1. The locations of the subsidence monitoring points

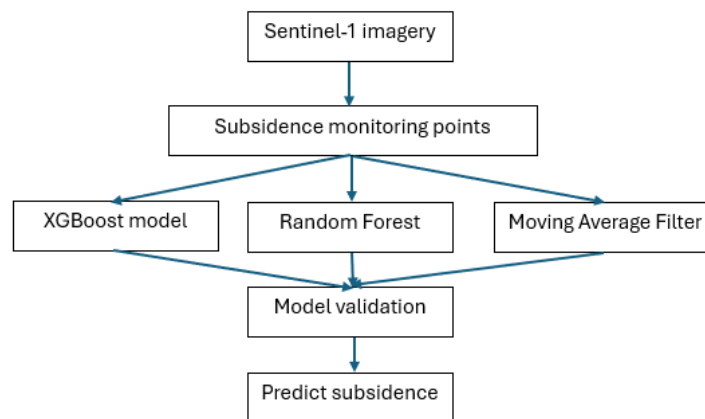


Fig. 2. Processing flow chat for land sunsidence predicting

2.2 MAF model

MAF is adopted in this study as a classical time-domain smoothing technique aimed at reducing short-term noise and enhancing the long-term trend within GNSS-based land subsidence data. This filter replaces each point in the sequence with the mean of its neighboring values, effectively dampening abrupt variations caused by measurement errors or local anomalies while retaining the general shape of the signal. In the context of land deformation analysis in Ca Mau province, Vietnam, a simple centered moving average with a fixed window size of three observations was applied. This allows each data point to be averaged with its immediate neighbors on both sides, yielding a balanced smoothing effect that preserves temporal alignment. The choice of a small window size was motivated by the need to retain localized deformation characteristics while still mitigating high-frequency noise.

Once the smoothing step is completed, the forecasting process begins by calculating the average difference between recent observations. This average increment is then recursively added to the last known smoothed value to estimate subsequent deformation values. Unlike machine learning models, this approach does not require training or parameter optimization, making it lightweight and highly interpretable. The forecasting proceeds in a step-wise fashion, where each newly predicted value is reused as input for the next prediction step. Although the MAF approach is relatively basic compared to more advanced models such as artificial neural networks or hybrid systems (e.g., MAF-MLPNN), it offers a transparent, efficient, and effective method for short-term prediction when data is limited or clean. Previous studies have demonstrated the practicality of using moving average techniques for denoising and pattern recognition in GNSS time-series and other geodetic applications (Vân Phong et al., 2023).

2.3 Random Forest Model (RF)

RF is an ensemble learning algorithm that combines multiple decision trees to enhance prediction accuracy and control overfitting. It was introduced by Breiman (Breiman, 2001) as an improvement over single decision tree models, which are prone to high variance. In a Random Forest, each tree is constructed using a different subset of training data, selected via bootstrapping, and a random subset of input features is considered for node splitting. This dual-layer randomness enables the model to produce more diverse trees, which collectively provide better generalization and reduce model variance. RF can be used for both classification and regression tasks and is particularly well-suited to handling high-dimensional, noisy, and complex datasets.

One of the primary advantages of Random Forest is its robustness against overfitting and its ability to handle missing or imbalanced data (Cutler et al., 2007). Because it aggregates the outputs of numerous decision trees, the final prediction is less sensitive to individual anomalies or noise in the data. Additionally, RF can estimate feature importance, allowing researchers to interpret the contribution of each variable to the overall prediction. This property is especially useful in scientific and geospatial domains, where understanding the influence of physical factors (e.g., soil type, elevation, or precipitation) is as important as making accurate predictions (Rodriguez-Galiano, Ghimire, Rogan, Chica-Olmo, & Rigol-Sanchez, 2012). The algorithm also performs internal validation by using out-of-bag (OOB) samples, making cross-validation more efficient.

2.4 XGBoost model

XGBoost is a powerful machine learning algorithm designed to implement gradient-boosted decision trees in a highly efficient, flexible, and scalable way. It has become one of the most widely used ensemble learning methods for regression and classification problems involving structured tabular data (Chen & Guestrin, 2016). In regression applications, XGBoost builds an ensemble of trees sequentially, where each new tree attempts to reduce the residual errors of the current ensemble by minimizing a regularized loss function that combines a convex loss term (such as squared error) with a penalty on model complexity to prevent overfitting (Bishop & Nasrabadi, 2006).

In the proposed model for land subsidence prediction, temporal deformation data were restructured into training samples where each row corresponded to a spatial observation (defined by latitude and longitude) at a specific time point, encoded as a fractional year. This transformation allowed the model to learn spatiotemporal patterns of vertical displacement. Feature standardization was applied to input variables to ensure uniform scale and enhance convergence during training (Géron, 2022).

A grid search optimization procedure was used to tune key hyperparameters of the XGBoost model, including learning rate, maximum tree depth, number of estimators, subsample ratio, and column subsample ratio. GridSearchCV with 3-fold cross-validation was employed to identify the best configuration based on negative mean squared error (Raschka & Mirjalili, 2019). This ensures that the model generalizes well to unseen data and minimizes variance and bias trade-offs.

Mathematically, the XGBoost prediction model can be expressed as:

$$\hat{y} = \sum_k k_k(x), \quad f_k \in F \quad (1)$$

Where each f_k represents a regression tree from the functional space F . The optimization objective at the t -th iteration is:

$$Obj_t = \sum_i l(y_i \hat{y}_i^{t-1} + f_t(x_i)) + \Omega(f_t) \quad (2)$$

With $\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_j w_j^2$ in which T is the number of leaves, w_j are leaf weights, λ is the L2 regularization term, and γ controls model complexity (Bishop & Nasrabadi, 2006).

Model performance was evaluated using several standard regression metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2), which are defined respectively as:

$$MSE = \frac{1}{n} \sum (y_i - \hat{y}_i)^2 \quad (3)$$

$$RMSE = \sqrt{MSE} \quad (4)$$

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i| \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

These metrics provide insight into both the accuracy and robustness of the predictive model (Raschka & Mirjalili, 2019).

Incorporating spatial coordinates and time as predictors, the XGBoost framework demonstrates strong potential for capturing complex non-linear relationships and making accurate land subsidence forecasts, particularly in geotechnical and environmental modeling domains (Chen & Guestrin, 2016).

3. Result and Discussion

3.1 Model Implementation and Evaluation

The three models - MAF, RF, and XGBoost - were implemented using the PSI-based dataset of 5,000 monitoring points in Ca Mau. Model parameters were optimized through grid search for XGBoost and randomized search for RF, while MAF was applied with a fixed window size as a baseline approach. Building on previous evidence of its strong predictive capacity in subsidence risk assessment (Tran et al., 2024), XGBoost was further employed in a hybrid framework where its predictions, together with those of RF, were combined using RidgeCV as a meta-regressor. This ensemble design aimed to leverage the complementary strengths of multiple algorithms, thereby improving both accuracy and generalization.

The training data incorporated spatial features (latitude and longitude) and temporal information encoded as fractional year values, with all inputs normalized using the StandardScaler method to ensure stable convergence during training. Model performance was systematically assessed using four regression metrics - MSE, RMSE, MAE, and R^2 - providing a comprehensive evaluation of both error magnitude and explanatory power. The outcomes of these evaluations are presented and analyzed in the next section.

The subsidence prediction method using the XGBoost, RF and MAF model is described as shown in Fig. 3 (a, b, c) respectively.

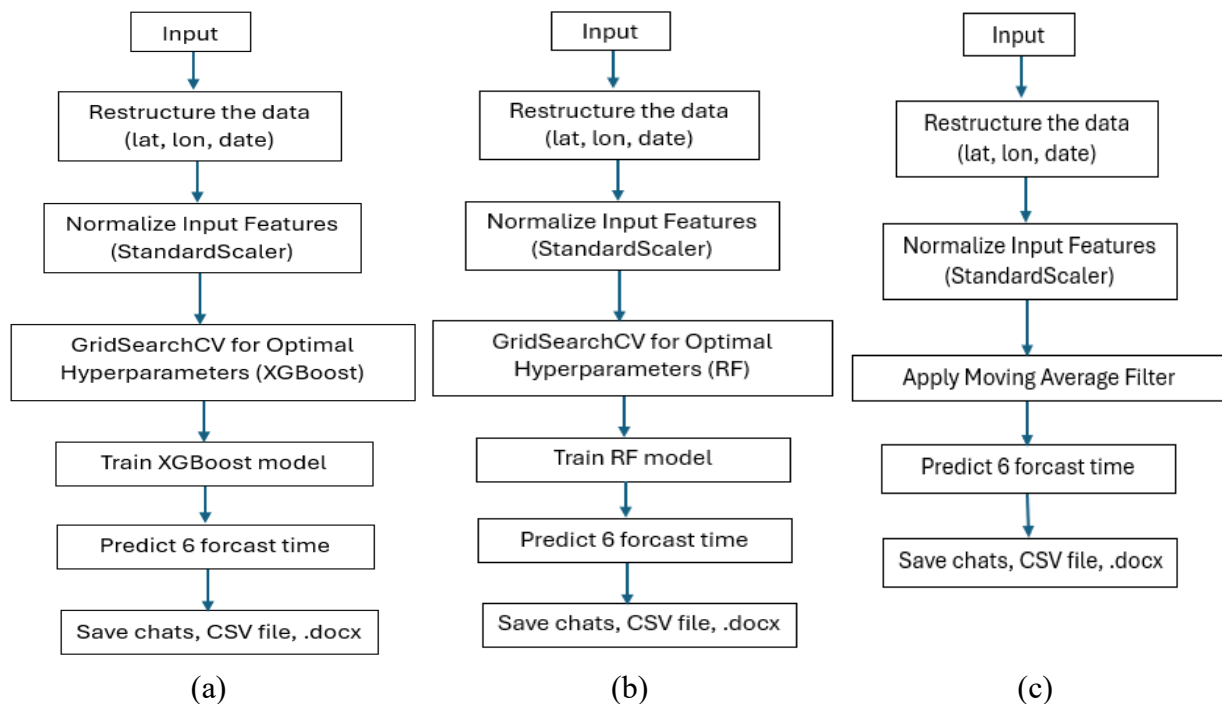


Fig. 3. The subsidence prediction process using the models

Parameter optimization of the models was performed using GridSearchCV for the XGBoost model and RandomizedSearchCV for the Random Forest model, respectively. These optimization techniques enable an exhaustive and random search over the parameter spaces, thereby identifying model configurations that improve the accuracy and reliability of the prediction process. After training, the models were validated using both regression and classification metrics. Regression-based evaluation included

computing the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2), which provides a quantitative assessment of the prediction accuracy over continuous strain values.

Finally, all numerical results, performance metrics, and visual outputs are compiled into an automated Word report. This end-to-end workflow ensures a transparent, reproducible, and robust framework for integrating satellite-based observations with machine learning techniques in subsidence forecasting.

3.2 Results of the Model

The performance evaluation results of the models are presented in Tab. 1.

Tab. 1. Computed values based on Equations (3) to (6)

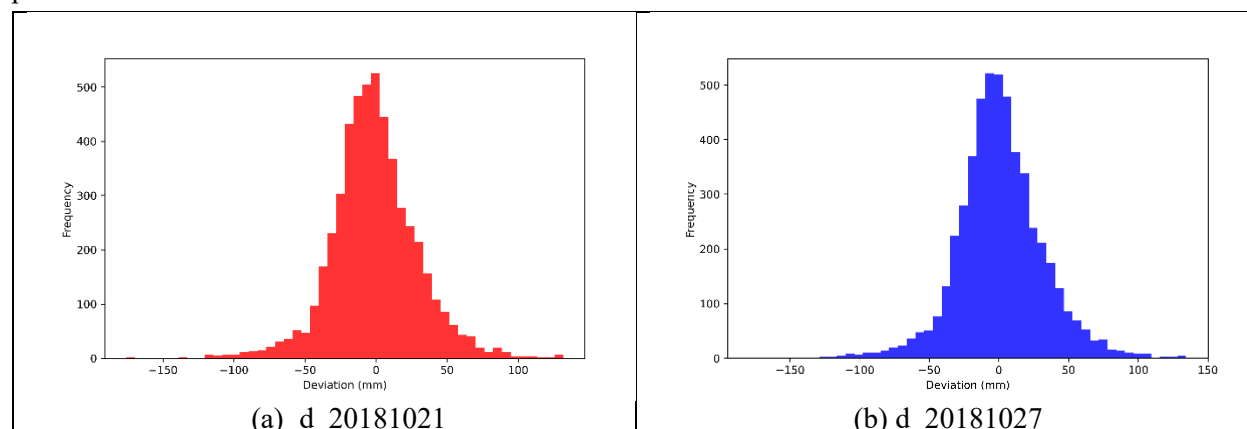
	RMSE	MAE	R^2
MAF	31.20	23.33	0.4148
RF	4.89	3.32	0.9857
XGBoost	4.67	3.23	0.9869

Tab. 2 summarizes the computed regression evaluation metrics based on Equations (3) to (6) for the MAF, Random Forest (RF), and XGBoost models. Among the three, the XGBoost model delivers the best overall performance across all evaluation criteria. Specifically, XGBoost achieves the lowest with RMSE of 4.67, compared to 4.89 for RF and 31.20 for MAF, indicating significantly more accurate predictions. MAE also reflects this trend, decreasing from 23.33 (MAF) and 3.32 (RF) to 3.23 with XGBoost. In terms of the coefficient of determination (R^2), XGBoost reaches the highest value of 0.9869, followed closely by RF at 0.9857, both of which outperform the considerably lower R^2 of 0.4148 from MAF. These results clearly demonstrate that machine learning-based models - particularly XGBoost - are more effective at capturing the underlying deformation patterns compared to the simpler MAF approach, which lacks adaptive learning capability.

Following the prediction of subsidence values, the deviation at each point was computed and statistically represented through histograms in the subsequent figures.

Figs 4 to 6 illustrate the distribution of predicted subsidence deviations for six forecast times using the MAF, Random Forest, and XGBoost models, respectively. These histograms provide insight into each model's forecasting precision and stability by visualizing the spread and central tendency of the prediction errors.

The deviation distributions produced by the MAF model (Fig. 4) are notably wider and more dispersed across all dates. This reflects the model's limited ability to adapt to local variations and nonlinear subsidence patterns, resulting in higher error variance. Most histograms under MAF show heavy tails and less pronounced central peaks, indicating a larger number of points with substantial over- or under-prediction.



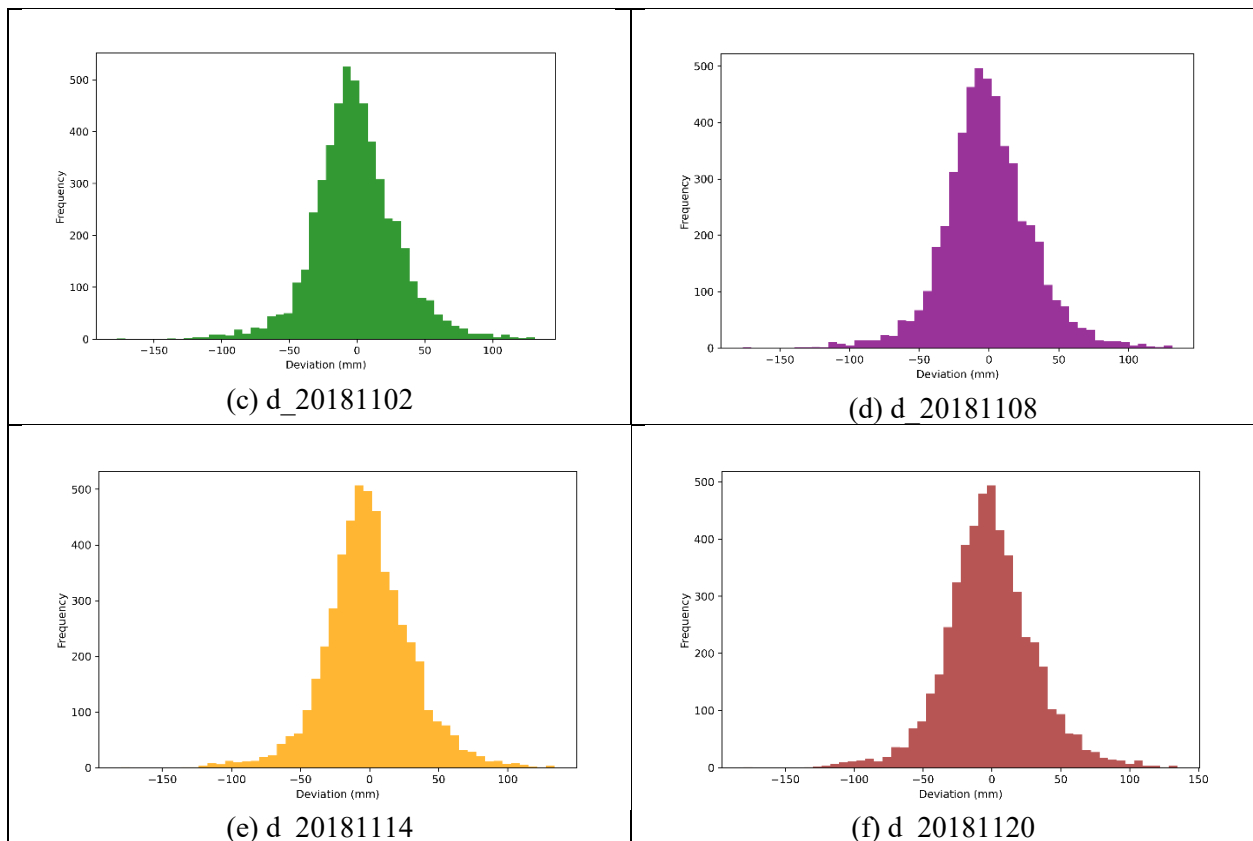
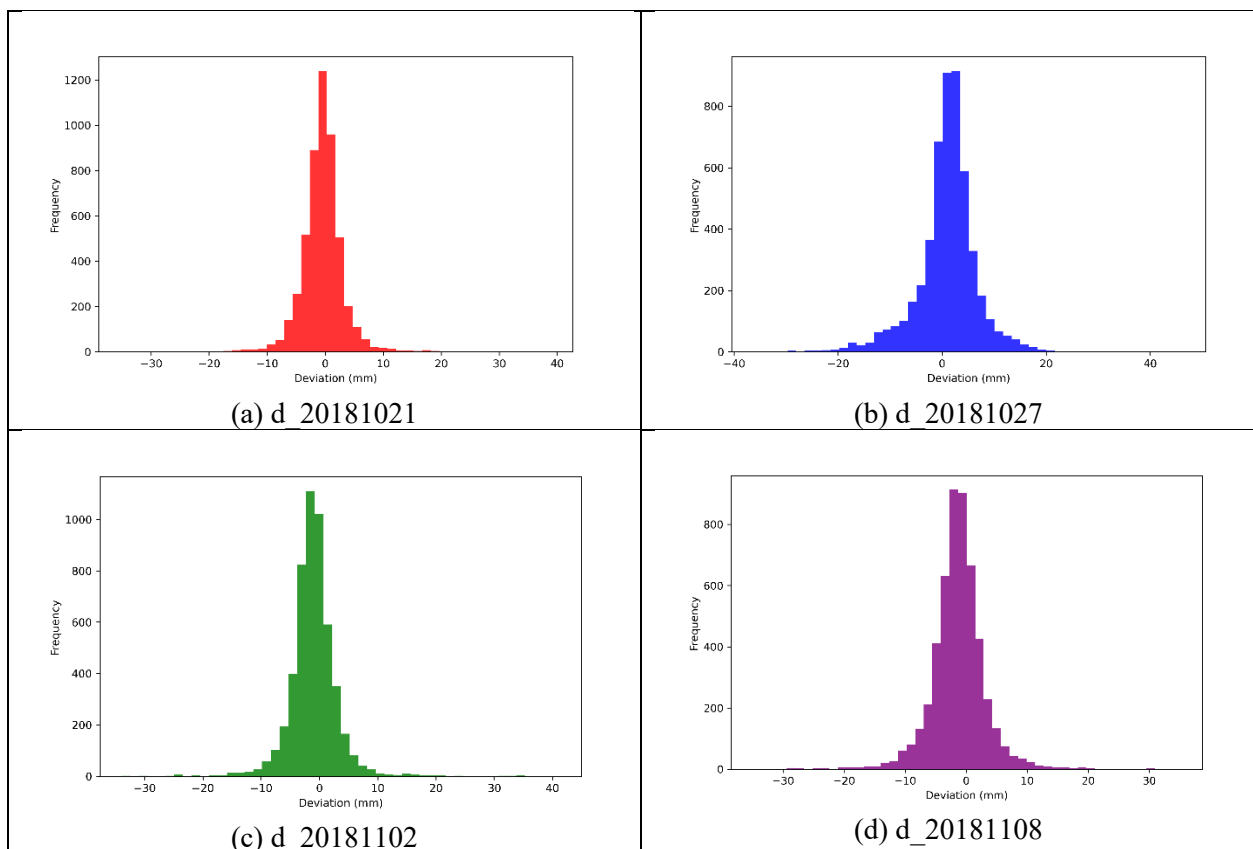


Fig. 4. Histogram of Predicted Subsidence Deviations Using MAF



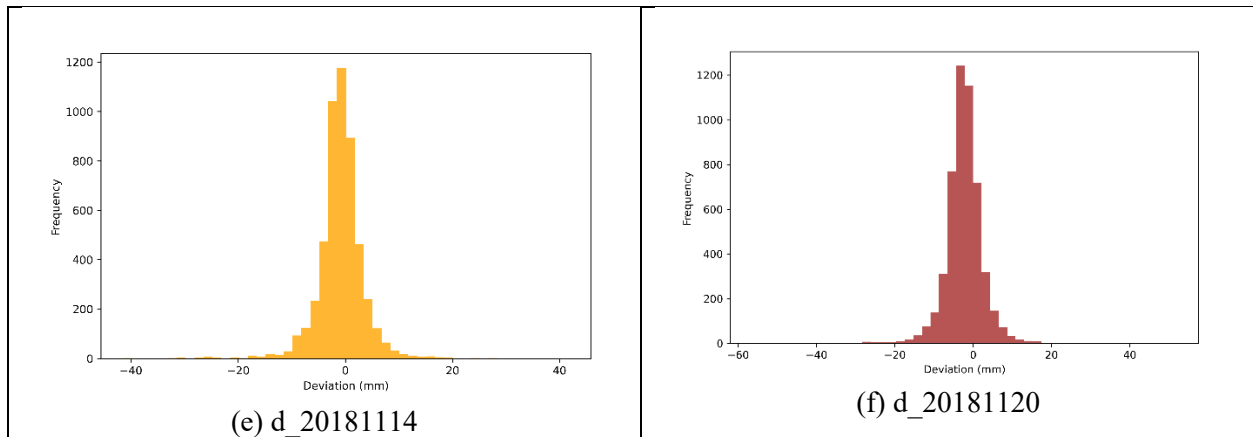


Fig. 5. Histogram of Predicted Subsidence Deviations Using RF Model

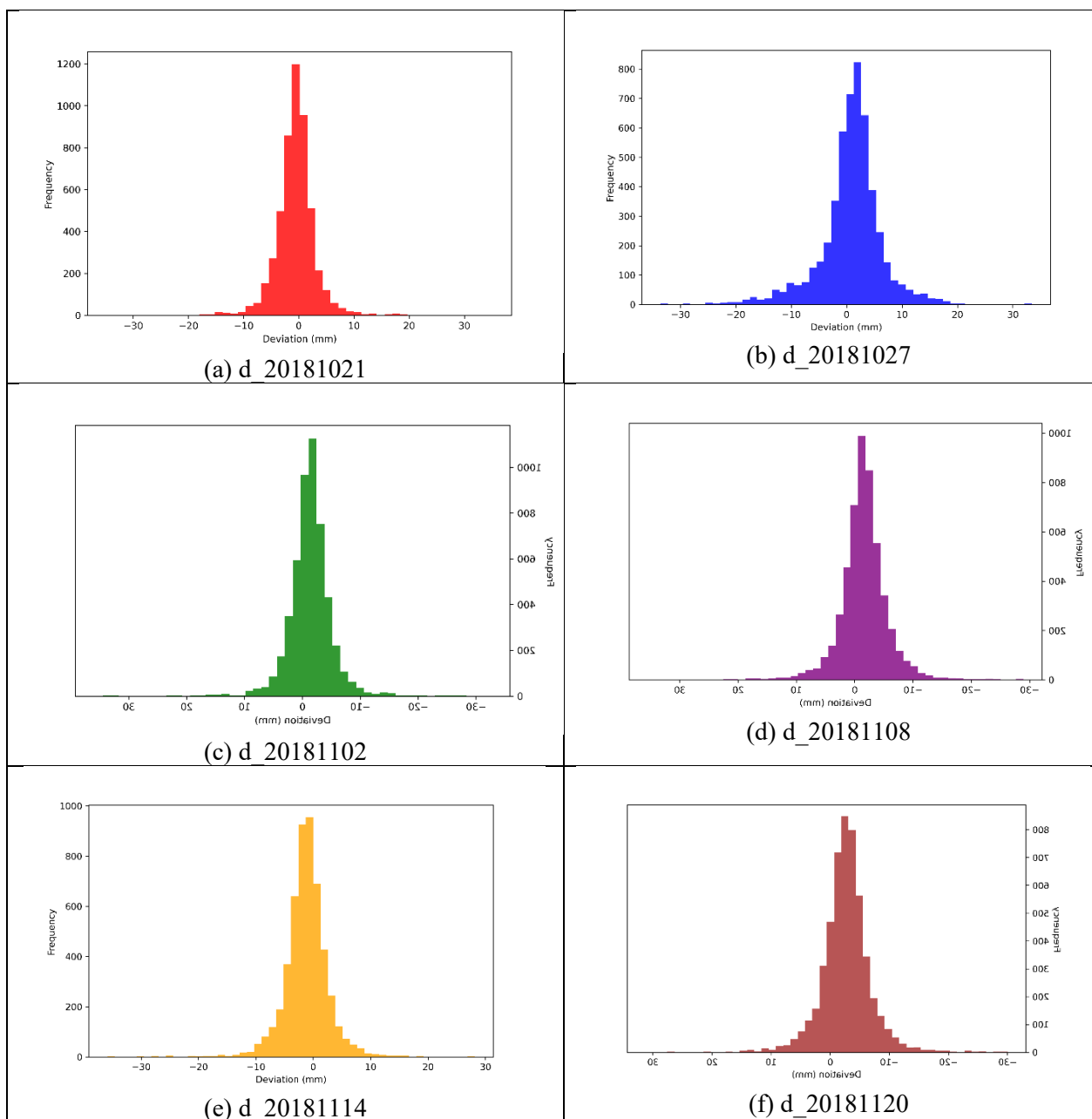


Fig. 6. Histogram of Predicted Subsidence Deviations Using XGBoost Model

In contrast, the Random Forest model (Fig. 5) demonstrates significantly improved performance. The deviation histograms are more narrowly concentrated around zero, with taller peaks and thinner tails. This suggests that RF is more capable of capturing spatial and temporal variability in the data, leading to more consistent and reliable predictions. Although slight skewness appears in a few time steps, the overall distribution is symmetric and tighter compared to MAF.

The XGBoost model (Fig. 6) further improves on the performance seen in RF. The histograms are the most centered and compact, with minimal deviation from zero. This indicates that XGBoost not only maintains high consistency across dates but also minimizes systematic bias in its predictions. The symmetry and sharpness of the distributions imply that the model effectively generalizes across the forecasted intervals, reducing both random and structural errors.

The R-squared values obtained from both the Random Forest and XGBoost models approach 100%, indicating an excellent fit to the input data. However, further efforts are still necessary to reduce RMSE in order to enhance the accuracy of the subsidence forecasting process.

Overall, the histogram analysis confirms the quantitative metrics: MAF produces the largest prediction deviations, while Random Forest and especially XGBoost offer more accurate and stable forecasts of land subsidence. These results reaffirm the superiority of machine learning-based approaches, particularly gradient-boosted ensembles, in modeling complex geospatial phenomena like vertical ground displacement.

4. Conclusion

This study presents the land subsidence prediction results of three prediction models, namely Moving Average Filter (MAF), Random Forest (RF), and XGBoost, for the Ca Mau Peninsula, Vietnam. The input data here consists of 5,000 monitoring points derived from PSI and incorporating spatiotemporal features, each model is evaluated using standard regression metrics. The results show that the machine learning-based models, especially XGBoost, significantly outperform the simple MAF method in terms of accuracy and stability. XGBoost achieves the lowest prediction error and the highest R^2 value, indicating a close fit to the deformation data. The Random Forest model also demonstrates high performance and consistency, reinforcing the value of ensemble learning methods in geospatial prediction.

In addition to the numerical data, the histogram analysis of the prediction bias further highlights the superior predictive ability of the XGBoost model, as its error distribution is concentrated and narrowest across all forecast time steps. Although the MAF model provides a lightweight and interpretable approach, its limited adaptability to complex spatial patterns limits its practical utility in high-resolution subsidence prediction. Meanwhile, the R^2 values of up to nearly 100% achieved by RF and XGBoost demonstrate the excellent generalization ability of the two models; however, further reduction in RMSE is still needed to enhance the predictive reliability.

Overall, the findings highlight the advantages of integrating InSAR-derived observations with advanced machine learning techniques. The XGBoost model, in particular, offers a promising and scalable solution for monitoring land subsidence at regional scales, supporting data-driven decision making in land use planning and climate adaptation strategies across vulnerable deltas.

References

- Bagheri-Gavkosh, M., Hosseini, S. M., Ataie-Ashtiani, B., Sohani, Y., Ebrahimian, H., Morovat, F., & Ashrafi, S. (2021). Land subsidence: A global challenge. *Science of The Total Environment*, 778, 146193.
- Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4): Springer.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- Cao, A., Esteban, M., Valenzuela, V. P. B., Onuki, M., Takagi, H., Thao, N. D., & Tsuchiya, N. (2021). Future of Asian Deltaic Megacities under sea level rise and land subsidence: current adaptation pathways for Tokyo, Jakarta, Manila, and Ho Chi Minh City. *Current opinion in environmental sustainability*, 50, 87-97.
- Chai, L., Xie, X., Wang, C., Tang, G., & Song, Z. (2024). Ground subsidence risk assessment method using PS-InSAR and LightGBM: A case study of Shanghai metro network. *International Journal of Digital Earth*, 17(1), 2297842.

- Chen, T., & Guestrin, C. (2016). *Xgboost: A scalable tree boosting system*. Paper presented at the Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.
- Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783-2792.
- de Wit, K., Lexmond, B. R., Stouthamer, E., Neussner, O., Dörr, N., Schenk, A., & Minderhoud, P. S. (2021). Identifying causes of urban differential subsidence in the Vietnamese Mekong Delta by combining InSAR and field observations. *Remote Sensing*, 13(2), 189.
- Delgado Blasco, J. M., Foumelis, M., Stewart, C., & Hooper, A. (2019). Measuring urban subsidence in the Rome metropolitan area (Italy) with Sentinel-1 SNAP-StaMPS persistent scatterer interferometry. *Remote Sensing*, 11(2), 129.
- EMSN062. (2019). *EMSN062: Assessing changes in ground subsidence rates, Mekong Delta, Vietnam*. Retrieved from
- Erban, L. E., Gorelick, S. M., & Zebker, H. A. (2014). Groundwater extraction, land subsidence, and sea-level rise in the Mekong Delta, Vietnam. *Environmental Research Letters*, 9(8), 084010.
- Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*: " O'Reilly Media, Inc."
- Gharechae, H., Samani, A. N., Sigaroodi, S. K., Baloochiyan, A., Moosavi, M. S., Hubbard, J. A., & Sadeghi, S. M. M. J. L. (2023). Land subsidence susceptibility mapping using Interferometric Synthetic Aperture Radar (InSAR) and machine learning models in a semiarid region of Iran. *Remote Sensing*, 12(4), 843.
- Hakim, W. L., Fadhillah, M. F., Park, S., Pradhan, B., Won, J.-S., & Lee, C.-W. (2023). InSAR time-series analysis and susceptibility mapping for land subsidence in Semarang, Indonesia using convolutional neural network and support vector regression. *Remote Sensing of Environment*, 287, 113453.
- Khan, J., Ren, X., Hussain, M. A., & Jan, M. Q. (2022). Monitoring land subsidence using PS-InSAR technique in Rawalpindi and islamabad, Pakistan. *Remote Sensing*, 14(15), 3722.
- Minderhoud, P. (2021). *Modelling Mekong Delta Subsidence, Challenges and How to Improve Quantifications*. Paper presented at the Conference proceedings: 4th Asia Pacific Meeting on Near Surface Geoscience & Engineering.
- Naghibi, S. A., Khodaei, B., & Hashemi, H. (2022). An integrated InSAR-machine learning approach for ground deformation rate modeling in arid areas. *Journal of hydrology*, 608, 127627.
- Nalluri, M., Pentela, M., & Eluri, N. R. (2020). A scalable tree boosting system: XG boost. *Int. J. Res. Stud. Sci. Eng. Technol*, 7(12), 36-51.
- Panahi, M., Khosravi, K., Golkarian, A., Roostaei, M., Barzegar, R., Omidvar, E., . . . Jun, C. (2022). A country-wide assessment of Iran's land subsidence susceptibility using satellite-based InSAR and machine learning. *Geocarto International*, 37(26), 14065-14087.
- Radman, A., Akhoondzadeh, M., & Hosseiny, B. (2021). Integrating InSAR and deep-learning for modeling and predicting subsidence over the adjacent area of Lake Urmia, Iran. *GIScience Remote Sensing*, 58(8), 1413-1433.
- Ranjgar, B., Razavi-Termeh, S. V., Foroughnia, F., Sadeghi-Niaraki, A., & Perissin, D. (2021). Land subsidence susceptibility mapping using persistent scatterer SAR interferometry technique and optimized hybrid machine learning algorithms. *Remote Sensing*, 13(7), 1326.
- Raschka, S., & Mirjalili, V. (2019). *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2*: Packt publishing ltd.
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS journal of photogrammetry remote sensing*, 67, 93-104.
- Tasan, M., Ghorbaninasab, Z., Haji-Aghajany, S., & Ghiasvand, A. (2023). Leveraging GNSS tropospheric products for machine learning-based land subsidence prediction. *Earth Science Informatics*, 16(4), 3039-3056.

- Tran, A. V., Brovelli, M. A., Ha, K. T., Khuc, D. T., Tran, D. N., Tran, H. H., & Le, N. T. (2024). Land subsidence susceptibility mapping in Ca Mau Province, Vietnam, using boosting models. *ISPRS International Journal of Geo-Information*, 13(5), 161.
- Van Phong, D., Trong, N. G., Chien, N. V., Thanh, N. H., Ha, L. L., Quan, N. V., & Quang, P. N. (2023). Analysis of vertical crustal displacement using ANN function based on time-series GNSS data processing results. *Journal of Meteorology and Hydrology*, 752, 41–50.