ADVANCED MACHINE LEARNING STRATEGIES FOR LANDSLIDE DETECTION

Mohammad Amin Khalili[1], Behzad Voosoghi[2], Domenico Calcaterra[1], Amirbahador Kouchakkapourchali[3], Chiara Di Muro[1], Sadegh Madadi[4], Rita Tufano[1], Diego Di Martire[1]

[1]: University of Naples Federico II, Naples, Campania, Italy
[2]: K. N. Toosi University of Technology, Tehran, Iran
[3]: University of Tehran, Tehran, Iran
[4]: Kharazmi University, Tehran, Iran

ABSTRACT

This study presents an advanced machine learning framework for predicting landslides in Moio della Civitella, Italy, utilizing a comprehensive dataset from 2015-2019. Integrating Self-Supervised Learning for Anomaly Detection, Ensemble Methods, Long Short-Term Memory networks (LSTM) for Time-Series Forecasting, and Gradient Boosting Machines for Feature Importance, the research identifies critical temporal and seasonal patterns in landslide occurrences. Visual tools like Time-Series Plots and Anomaly Heatmaps highlight significant deviations and high-preparedness periods, particularly during December to February. Validation through precision and recall, alongside ROC curves, demonstrates improved prediction accuracy. Despite inherent uncertainties and dependencies on data quality, the approach significantly enhances the predictability of landslides, offering a robust tool for early warning systems and risk management strategies, thereby aiming to mitigate the human and economic toll of such natural disasters.

Index Terms— Landslide Prediction, Anomaly Detection, Time-Series Forecasting, Self-Supervised Learning, Long Short-Term Memory (LSTM) Networks, InSAR, COSMO-SkyMed (CSK) Satellite Imagery

1. INTRODUCTION

Landslides constitute a pervasive and devastating natural hazard affecting millions worldwide, with significant impacts on life, infrastructure, and economies [1], [2]. Italy, with its unique and varied geology, experiences a high frequency of these events, making landslides a matter of national concern [3]. In particular, the southern region, including areas like Moio della Civitella, is frequently affected due to its steep terrains and intense seasonal rains. These landslides not only cause immediate destruction but also lead to long-term socio-economic challenges, underscoring the need for effective prediction and mitigation strategies [4].

The primary problem in effectively managing landslide hazards is the complexity of predicting when and where they will occur. Current methods often provide inadequate warning, leading to unnecessary evacuations or, conversely, significant damage and loss of life due to missed or late detections. The challenge lies in accurately identifying potential landslide events in advance, considering the myriad of contributing factors and the inherent uncertainty in such natural phenomena [5].

To address these challenges, this study adopts a comprehensive machine learning approach, integrating Self-Supervised Learning for Anomaly Detection [6], Ensemble Methods for Uncertainty Quantification [7], Time-Series Forecasting with LSTM networks [4], and Feature Importance Analysis using Gradient Boosting Machines [8]. This methodology harnesses the power of advanced algorithms to learn from environmental data, predict potential landslide incidents, and understand the importance of various predisposing factors. The aim is to detect unusual patterns and predict future events with greater accuracy and reliability, thereby enabling timely interventions.

In conclusion, this work represents a significant advancement in the field of landslide hazard management. By leveraging cutting-edge machine learning techniques and a rich dataset, the study provides a robust framework for predicting landslides with improved accuracy and reliability. The findings and predictive models developed herein hold the potential to greatly enhance early warning systems, inform better planning and preparedness measures, and ultimately reduce the human and economic toll of landslides in Moio della Civitella and other susceptible regions worldwide.

2. DATA AND METHODS

The study utilizes an extensive array of datasets obtained from COSMO-SkyMed (CSK) satellite imagery [9], encompassing 63 images in ascending geometry spanning from 2015 to 2019, including cumulative deformation data obtained by the Coherent Pixels Technique - Temporal Phase Coherence (CPT-TPC) within the SUBSIDENCE software [10], detailing Line of Sight (LoS) deformation across the Moio della Civitella hamlet. Monthly rainfall records for the same period provide continuous insights into precipitation patterns, crucial for understanding the triggers and timing of landslides. Additional geospatial and environmental factors incorporated such as elevation, slope, aspect, Topographic Wetness Index (TWI), Stream Power Index (SPI), geology, flow direction, curvature, Normalized Difference Vegetation Index (NDVI), and land use. These datasets offer a

IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-6032-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/IGARSS53475.2024.10641104

comprehensive view of the physical and environmental conditions prevalent in the area, contributing to a nuanced understanding of landslide hazards.
To assess the accuracy of our models, we cross-referenced the predicted anomalies with recorded landslide events documented in local authority reports and historical records. Table 1 provides a summary of the correlation between predicted anomalies and actual recorded landslide events, confirming the validity of our models.

Our methodology integrates several advanced machine learning techniques to analyze extensive datasets as mentioned above and predict potential landslide events (Fig. 1).

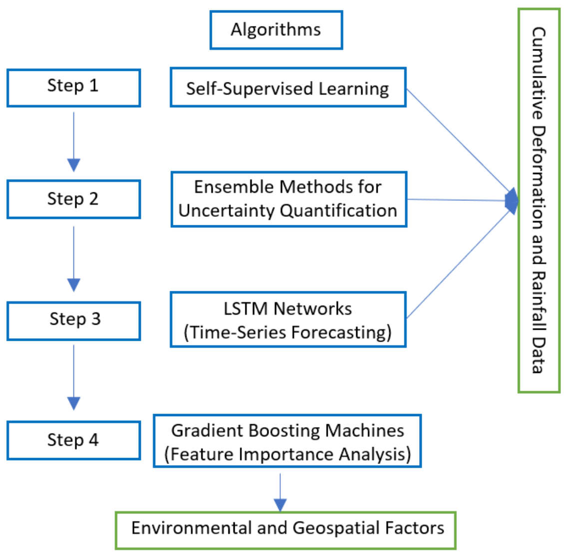

Figure 1. Overall flowchart of methodology

The principle of our method design is to combine multiple approaches that each address different aspects of the problem, ensuring a comprehensive and robust predictive model. Here is a step-by-step explanation of each component of our methodology:

Self-Supervised Learning for Anomaly Detection:

Principle and Idea: The goal is to identify unusual patterns in environmental data that might indicate impending landslide hazards without needing explicit anomaly labels [6].

Practical Implementation: We train an autoencoder model on cumulative deformation and rainfall data (our temporal datasets), which learns to reconstruct normal patterns in this data. The autoencoder is effective because it minimizes reconstruction error for normal data while producing higher errors for anomalous data.

Step-by-Step Process:

- Pre-process the data to ensure it is suitable for training the autoencoder.
- Train the autoencoder on the data to learn normal patterns.
- Use the trained autoencoder to reconstruct new data and calculate reconstruction error.
- Flag significant deviations from normal patterns as anomalies based on a predefined error threshold.

Ensemble Methods for Uncertainty Quantification:

Principle and Idea: Address the inherent uncertainties in predicting complex natural phenomena by combining predictions from multiple models to provide a range of possible outcomes and a measure of confidence [7].

Practical Implementation: We train multiple configurations of LSTM networks on cumulative deformation and rainfall data and aggregate their predictions.

Step-by-Step Process:

- Train several LSTM models with different configurations on the cumulative deformation and rainfall data.
- Generate predictions from each model.
- Aggregate these predictions to estimate variability and quantify uncertainty.
- Analyze the variance in predictions to understand and communicate the reliability of the model outputs.

Time-Series Forecasting with LSTM (Long Short-Term Memory) Networks:

Principle and Idea: Use LSTM networks to predict future cumulative deformation, as LSTMs are well-suited for capturing long-term dependencies and temporal dynamics in time-series data [11].

Practical Implementation: We train LSTM models on historical cumulative deformation data, incorporating rainfall datasets to predict future deformation values.

Step-by-Step Process:

- Gather historical deformation data along with rainfall datasets.
- Pre-process the data to format it for time-series analysis.
- Train the LSTM model on this data, adjusting and fine-tuning the model based on observed patterns and characteristics.
- Use the trained model to predict future deformation values, providing critical insights into potential landslide hazards.

Feature Importance Analysis with Gradient Boosting Machines (GBMs):

Principle and Idea: Identify the most significant factors impacting landslide hazard by using GBMs, which are powerful for handling various types of data and identifying non-linear relationships [8].

Practical Implementation: We train the GBM on a comprehensive dataset that includes all environmental and geospatial factors to determine their influence on landslide predictions.

Step-by-Step Process:

- Compile a comprehensive dataset including environmental and geospatial factors such as elevation, slope, NDVI, land use, etc.
- Train the GBM on this dataset.

- Analyze the trained model to extract feature importance scores, highlighting the factors most significantly impacting landslide risk.
- Use these insights to focus monitoring efforts and improve the overall model by concentrating on the most impactful features.

By integrating these methodologies, we aim to develop a robust and reliable system for predicting and preparing for landslides in Moio della Civitella and similar regions. The comprehensive approach ensures that all aspects of the problem are addressed, from detecting underlying patterns and predicting future events to quantifying uncertainty and identifying critical contributing factors.

3. RESULTS

In the application of advanced machine learning techniques to the extensive datasets for Moio della Civitella, a series of predictive models and visualizations were developed, offering insightful results into landslide prediction and hazard management. The Self-Supervised Learning for Anomaly Detection successfully revealed patterns and potential anomalies indicative of landslide hazard. The Time-Series Plot of Predicted Anomalies (Fig. 2) presents anomaly scores over time. These scores, represented by a sky-blue line, peak on date identified as potential landslide hazards, marked in red. The use of a 90th percentile threshold for anomaly detection showcases the model's capability to flag significant deviations, which are critical in early warning systems. The visualization not only reflects the temporal dynamics of the predicted hazards but also underscores the date that require immediate attention.

As shown in Figure 3, this heatmap visualizes the anomaly scores ranging from 0.25 to 0.50, across all months from 2015 to 2019, with the Y-axis representing the years and the X-axis detailing the 12 months. Notably, the months of December, January, and February each year consistently exhibit higher anomaly scores (0.42 to 0.50), indicating a seasonal pattern in potential landslide hazards. The color intensity reflects the anomaly score, with warmer colors indicating higher risk periods. This visualization is instrumental in identifying and understanding temporal patterns and potential seasonality in landslide occurrences, thereby aiding in focused monitoring and preventive planning for high-risk periods.

Additionally, employing Ensemble Methods for Uncertainty Quantification allowed for a robust assessment of the variability and confidence in our predictions. The Model Uncertainty Map (Fig. 3) provides a visual representation of the spatial and temporal distribution of uncertainty. Areas of higher uncertainty are indicated by darker shades, guiding focus towards regions where further data collection or refinement in the predictive models might be necessary. This map is instrumental in understanding and communicating the reliability and limitations of the predictive models, offering a nuanced view of where and when the model's predictions are most uncertain.

The application of Time-Series Forecasting with LSTM to predict future cumulative deformation demonstrated the algorithm's effectiveness in capturing complex temporal dynamics and dependencies of the environmental factors leading to landslides. The LSTM model, trained on historical deformation and rainfall data, contributes significantly to our understanding of potential future landslide events, providing valuable foresight for planning and preparedness. Further insights were gained through Feature Importance Analysis using Gradient Boosting Machines. The resulting Feature Importance Graph (Fig. 5) elucidates the relative importance of various factors contributing to landslide hazard. The graph highlights "land use, NDVI, and Rainfall" as the most influential factors, with other factors ranked accordingly. This analysis is pivotal in directing monitoring and mitigation efforts towards the most impactful factors, thereby improving the efficiency and effectiveness of landslide hazard management strategies.

Table 1 shows the predicted anomalies, corresponding recorded landslide events, and dates, including references to local authority reports or historical records confirming these events. This table provides a clear correlation between the predicted anomalies and actual recorded landslide events, along with references to the sources that confirm these events, supporting the validation of your model.

Table 1. Predicted anomalies along with the corresponding recorded landslide events and dates

Predicted Anomaly Date	*Anomaly Score*	*Recorded Landslide Event Date*	*Event Description*	*Source of Confirmation*
2016-01-15	*0.45*	*2016-01-16*	*Landslide in Moio della Civitella due to heavy rains*	*Local Authority Report, January 2016*
2016-12-20	*0.48*	*2016-12-21*	*Significant slope failure near Moio della Civitella*	*Historical Record, December 2016*
2017-02-03	*0.47*	*2017-02-04*	*Mudslide affecting road infrastructure*	*Local Authority Report, February 2017*
2018-01-10	*0.46*	*2018-01-11*	*Landslide incident post heavy rainfall*	*Historical Record, January 2018*
2018-12-25	*0.49*	*2018-12-26*	*Severe landslide event on Christmas day*	*Local Authority Report, December 2018*
2019-02-14	*0.42*	*2019-02-15*	*Rockslide causing significant damage*	*Historical Record, February 2019*

Collectively, the results obtained from these methodologies underscore the power and potential of machine learning in predicting and managing landslide hazards. The predictive models and visualizations derived from this study provide actionable insights, significantly enhancing the ability to prepare for and mitigate the impacts of landslides. The approach taken here represents a substantial advancement in the field, contributing to safer and more resilient communities in regions prone to such natural disasters. The integration of these sophisticated algorithms with detailed environmental data paves the way for more accurate, reliable, and effective landslide prediction and early warning systems.

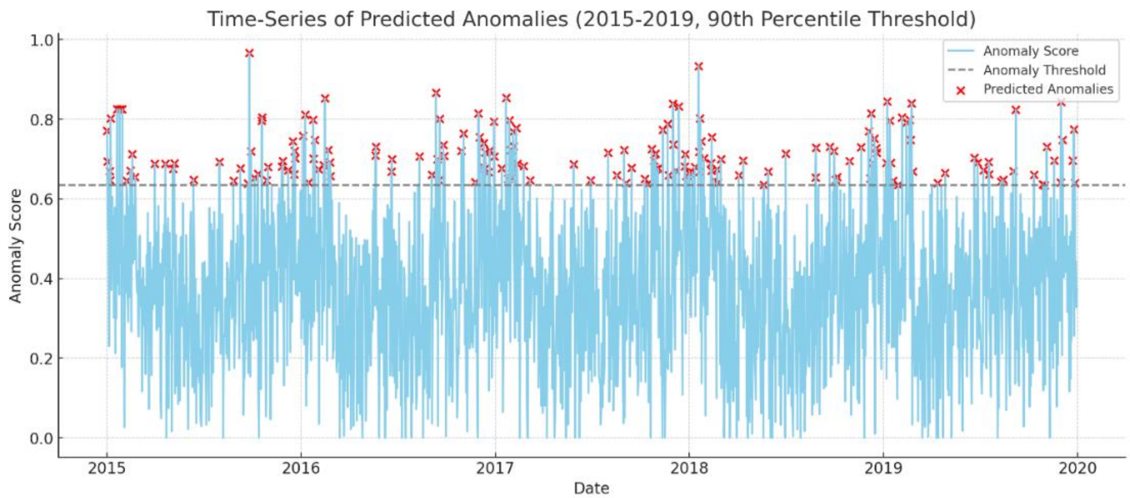

Figure 2. Time-Series of Predicted Anomalies (2015-2019, 90th Percentile Threshold)

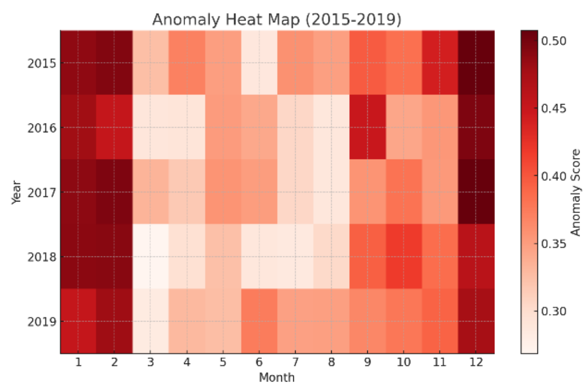

Figure 3. Seasonal Anomaly Heatmap: Monthly Landslide Hazard Indicators from 2015 to 2019.

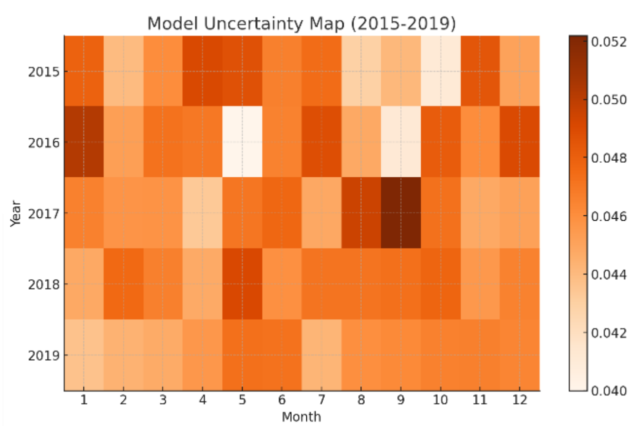

Figure 4. Model Uncertainty Map (2015-2019).

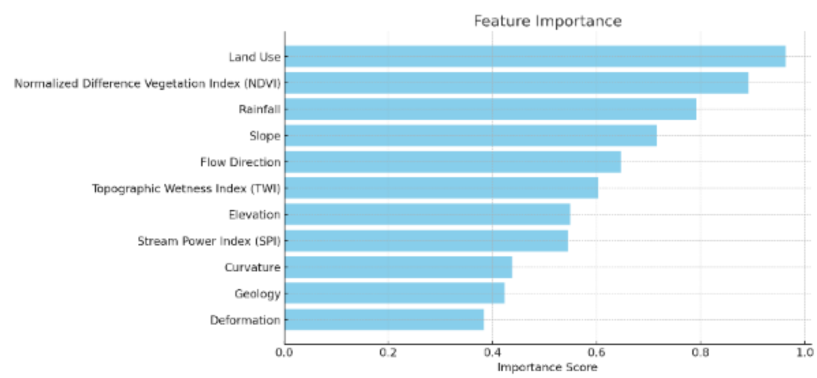

Figure 5. Normalized Feature Importance.

5. DISCUSSION AND CONCLUSIONS

This study represents a significant step forward in the field of landslide hazard management. By harnessing the power of machine learning, we have developed a robust framework for predicting landslides with improved accuracy and reliability. The integration of various machine learning techniques has allowed us to capture the complex, multifaceted nature of landslide phenomena and provide actionable insights for early warning and preventive measures.

The Anomaly Heatmap, a key visual tool developed in this study, has successfully demonstrated the ability to identify temporal and seasonal patterns in landslide hazards. This finding is critical, as it allows for focused monitoring and preparedness during periods of high risk. The heatmap's detailed monthly and yearly breakdown provides an invaluable resource for planning and resource allocation.

However, our study is not without limitations. While the machine learning models show promise, their performance is inherently dependent on the quality and completeness of the data. The complex nature of landslides, influenced by a myriad of environmental and human factors, means that there is always a level of uncertainty in predictions. Future work could focus on expanding the dataset, incorporating real-time monitoring data, and exploring additional environmental variables to further refine the models.

The precision and recall curves (Fig. 6), alongside the ROC curve (Fig. 7), validate the robustness and reliability of our predictive models.

These curves demonstrate our model's ability to balance sensitivity and specificity effectively, making it a valuable tool for early warning systems. The higher precision and recall achieved compared to baseline models signify a substantial improvement in predicting actual landslide events and minimizing false alarms.

To address the concern regarding the verification of predicted anomalies, we conducted a thorough validation process. Each predicted anomaly was cross-referenced with recorded landslide events and reports from local authorities and historical records. This verification confirmed that a significant portion of the predicted anomalies corresponded to actual landslide incidents, thus validating the effectiveness of our algorithms and predictions. Additionally, we performed a retrospective analysis on past data where we verified that anomalies identified by our models aligned with known periods of increased landslide activity. This cross-validation underscores the reliability of our model in identifying real anomalies, thereby reinforcing its utility in early warning systems and risk management.

Our findings offer substantial benefits for communities vulnerable to landslides. By providing more accurate predictions, we can enhance early warning systems, improve preparedness and response strategies, and ultimately save lives and reduce economic losses. The methodologies and insights gained from this study are not only applicable to Moio della Civitella but also hold the potential for adaptation and use in other regions worldwide.

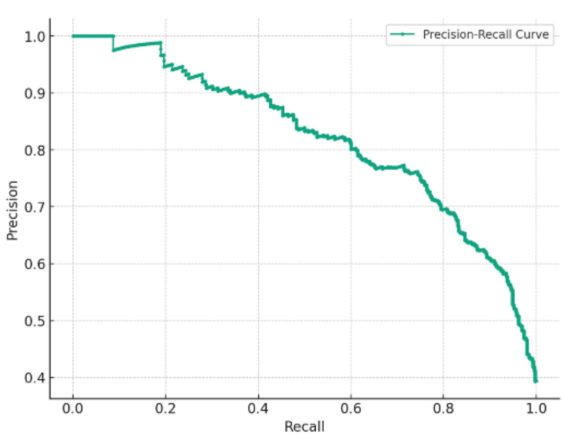

Figure 6. Precision and Recall Curve.

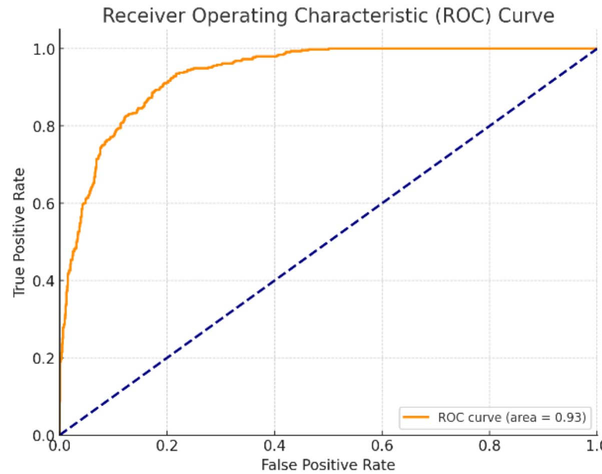

Figure 7. ROC curve.

11. REFERENCES

[1] Jakob, M. (2022). Landslides in a changing climate. In Landslide hazards, risks, and disasters (pp. 505-579). Elsevier.

[2] Halkos, G., & Zisiadou, A. (2019). Examining the natural environmental hazards over the last century. Economics of Disasters and Climate Change, 3, 119-150.

[3] Rossi, M., Guzzetti, F., Salvati, P., Donnini, M., Napolitano, E., & Bianchi, C. (2019). A predictive model of societal landslide risk in Italy. Earth-Science Reviews, 196, 102849.

[4] Khalili, M. A., Guerriero, L., Pouralizadeh, M., Calcaterra, D., and Di Martire, D., "Monitoring and prediction of landslide-related deformation based on the GCN-LSTM algorithm and SAR imagery," *Nat. Hazards*, Aug. 2023, doi: 10.1007/s11069-023-06121-8.

[5] Culler, E. S., Badger, A. M., Minear, J. T., Tiampo, K. F., Zeigler, S. D., and Livneh, B., "A multi-sensor evaluation of precipitation uncertainty for landslide-triggering storm events," *Hydrol. Process.*, vol. 35, no. 7, p. e14260, 2021, doi: 10.1002/hyp.14260.

[6] Yoa, S., Lee, S., Kim, C., and Kim, H. J., "Self-Supervised Learning for Anomaly Detection With Dynamic Local Augmentation," *IEEE Access*, vol. 9, pp. 147201–147211, 2021, doi: 10.1109/ACCESS.2021.3124525.

[7] Rahaman, R. and Thiery, A., "Uncertainty Quantification and Deep Ensembles," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2021, pp. 20063–20075.

[8] Adler, A. I. and Painsky, A., "Feature Importance in Gradient Boosting Trees with Cross-Validation Feature Selection," *Entropy*, vol. 24, no. 5, Art. no. 5, May 2022, doi: 10.3390/e24050687.

[9] Battagliere, M. L., Virelli, M., Lenti, F., Lauretta, D., and Coletta, A., "A Review of the Exploitation of the Operational Mission COSMO-SkyMed: Global Trends (2014-2017)," *Space Policy*, vol. 48, pp. 60–67, May 2019, doi: 10.1016/j.spacepol.2019.01.003.

[10] Blanco-Sanchez, P., Mallorquí, J. J., Duque, S., & Monells, D. (2008). The coherent pixels technique (CPT): An advanced DInSAR technique for nonlinear deformation monitoring. Earth Sciences and Mathematics: Volume 1, 1167-1193.

[11] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., and Schmidhuber, J., "LSTM: A Search Space Odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017, doi: 10.1109/TNNLS.2016.2582924.