

Article

A Machine Learning-Driven Approach to Uncover the Influencing Factors Resulting in Soil Mass Displacement

Apostolos Parasyris ^{*}, Lina Stankovic  and Vladimir Stankovic 

Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XW, UK; lina.stankovic@strath.ac.uk (L.S.); vladimir.stankovic@strath.ac.uk (V.S.)

* Correspondence: apostolos.parasyris@strath.ac.uk

Abstract: For most landslides, several destabilising processes act simultaneously, leading to relative sliding along the soil or rock mass surface over time. A number of machine learning approaches have been proposed recently for accurate relative and cumulative landside displacement prediction, but researchers have limited their studies to only a few indicators of displacement. Determining which influencing factors are the most important in predicting different stages of failure is an ongoing challenge due to the many influencing factors and their inter-relationships. In this study, we take a data-driven approach to explore correlations between various influencing factors triggering slope movement to perform dimensionality reduction, then feature selection and extraction to identify which measured factors have the strongest influence in predicting slope movements via a supervised regression approach. Further, through hierarchical clustering of the aforementioned selected features, we identify distinct types of displacement. By selecting only the most effective measurands, this in turn informs the subset of sensors needed for deployment on slopes prone to failure to predict imminent failures. Visualisation of the important features garnered from correlation analysis and feature selection in relation to displacement show that no one feature can be effectively used in isolation to predict and characterise types of displacement. In particular, analysis of 18 different sensors on the active and heavily instrumented Hollin Hill Landslide Observatory in the north west UK, which is several hundred metres wide and extends two hundred metres downslope, indicates that precipitation, atmospheric pressure and soil moisture should be considered jointly to provide accurate landslide prediction. Additionally, we show that the above features from Random Forest-embedded feature selection and Variational Inflation Factor features (Soil heat flux, Net radiation, Wind Speed and Precipitation) are effective in characterising intermittent and explosive displacement.

Keywords: feature selection; feature extraction; hierarchical clustering; supervised learning; dendrogram; landslides; shear failure; slope; monitoring



Citation: Parasyris, A.; Stankovic, L.; Stankovic, V. A Machine Learning-Driven Approach to Uncover the Influencing Factors Resulting in Soil Mass Displacement. *Geosciences* **2024**, *14*, 220. <https://doi.org/10.3390/geosciences14080220>

Academic Editor: Hans-Balder Havenith

Received: 28 June 2024

Revised: 7 August 2024

Accepted: 15 August 2024

Published: 18 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Moisture-induced landslides, activated by prolonged and heavy rainfall periods, are an increasing threat to humans, especially around train tracks, major roads, dam reservoirs, canals and densely populated areas [1]. With the growing intensity of landslides, especially those linked to precipitation and climate change, it is important to understand the underlying processes leading up to a slope failure. Geophysical monitoring of moisture-induced landslides can provide knowledge about spatial and temporal subsurface variations, while also enhancing and guiding the deployment of effective monitoring technologies [2]. However, subsurface analyses are more limited due to the cost of monitoring and maintenance, access to high-risk slopes and slow emergence of data-driven approaches from the community. The timely prediction of imminent landslides at the slope scale still remains a challenging problem, because the relationship between indicators (such as displacement) and influencing factors (such as temperature, soil heat flux, net radiation, moisture content), including their connection with stabilising forces (friction, gravity) and destabilising forces

(rainfall, gravity in higher angles of inclination and earthquakes), that govern a slope's stability are as yet not fully understood.

Over the years, many landslide models have been developed based either on limit equilibrium analysis methods [3], or on numerical simulation methods [4–6] to perform slope stability analyses, taking into account slope geometry, influencing physical and mechanical geomaterial properties contributing to a slope failure, as extracted from costly laboratory tests. For example, the Factor of Safety (FOS) of a slope, i.e., the ratio of shear strength stress to acting shear stress, is sensitive to slope angle, slope height, unit weight, friction angle and cohesion of soil, while it is least sensitive to the deformation parameters of soil and the depth of foundation layer or the choice of the constitutive models of the material response to different loads [7].

Besides these physical model-driven approaches, developed to model the influencing physical factors contributing to slope instabilities, there has been a recent trend to investigate data-driven approaches to better understand the spatial and temporal relationships between the influencing factors and landslide deformation. This has been enabled by a large amount of sensor measurements that have been collected and made available for slope stability analysis. In an attempt to quantify the link between identified landslides and meteorological data (i.e., rainfall, maximum–minimum temperature, wind speed, relative humidity and net solar radiation) through the use of Self-Organizing Map and clustering, it was concluded that 15-day accumulated precipitation is the most influential factor for landslides under observation [8]. A similar observation was reached during hierarchical and K-means clustering of rainfall data at different time scales (1 h, 3 h, 6 h, 24 h, 48 h and 72 h before the landslide event) against historical landslide data from the Metropolitan Region of Recife at Pernambuco State in Brazil, obtained from six gauges and three geotechnical stations for a period from 2005 to 2021, showing that rainfall accumulation thresholds are critical for issuing landslide warnings [9]. Ref. [10] focused on environmental factors related to heat exchange, such as thermoelasticity, permafrost and snow insulation, that were identified as triggering factors for landslide failures, concluding that a range of meteorological observations can be linked to and used to predict slope failure.

Machine learning approaches have been also used for landslide zonation mapping generation, for either susceptibility [11] or hazard assessment. A multivariate learning approach, taking advantage of XGBoost, incorporating parameters such as rainfall intensity, soil moisture, temperature and snowfall, was proposed for generating a unified Landslide Hazard Indicator to describe the seasonality of landslides based on National Climate Assessment—Land Data Assimilation System and Pacific Northwest Landslide Inventory data, where it was concluded that rainfall, soil moisture and temperature are the most important predictors of landslides [12].

One of the most critical aspects in the attempt to predict landslide failure is to select predictors according to their relative importance. The choice of parameters used as predictors according to their relative importance can vary greatly, since factors that have a high contribution for one prediction model may be useless for another [13]. At the same time, not all the selected factors have good predictive ability and in several cases can create noise and reduce prediction quality [14], and so the choice of unrepresentative variables in the model can lead to poor prediction capabilities [15]. Several machine learning-based studies highlight the importance of feature selection for landslide susceptibility map generation through ML methods, since the use of important factors such as rainfall, slope degree and elevation lead to prediction accuracy [16]. In another study, the selection of elevation, lithology, Normalised Difference Vegetation Index (NDVI), slope degree, solar radiation, Terrain Ruggedness Index (TRI) and distance to roads among 15 conditioning factors resulted in accepted results for susceptibility mapping via an ML approach [17].

1.1. Literature Review

Machine learning has been increasingly used for landslide displacement prediction to provide early warning of landslide failure. In this subsection, we review machine learning-driven landslide displacement prediction studies most relevant to our work.

In [15], a combination of groundwater level (GWL)-derived features and precipitation measurements with a climatological index for only two years of landslide displacement data were used for the prediction of rainfall-induced landslide movements. Using RF, the maximum absolute prediction error of 0.68 mm/day is achieved for daily relative displacement prediction and less than 5.5 mm daily cumulative prediction within periods up to 30 days. In [18], seven deep learning architectures were examined for the prediction of relative displacement on four landslides with different geographic locations, geological settings, time step dimensions and measurement instruments. The results obtained using 3, 4, 5 and 13 years of continuous recordings of displacement, precipitation and, in some cases, GWL fluctuation measurements, show that the Multiple Layer Perceptron (MLP), long short-term memory (LSTM) and gated recurrent unit (GRU) architectures achieved similar relative displacement prediction, ranging from an RMSE of 0.706 mm and $R^2 = 0.5928$ to RMSE = 13.555 mm and $R^2 = 0.6562$. In [19], landslide movement prediction was implemented via the decomposition of cumulative displacement and separate prediction of trend and periodic parts; polynomial approximation was then used for predicting the trend, and a Two-stage Combined Deep Learning Dynamic Prediction Model (TC-DLDPM) for the periodic part. The dataset contained 5 years of recorded displacement used for training and 1 year for testing, while rainfall and water level on various accumulating periods were used as influencing factors. The prediction of cumulative displacement resulted in an MAE of 8.93 mm. Ref. [20] compares five machine learning methods on three case studies of landslides for cumulative displacement prediction, based on GWL and rainfall, for six years of continuous monitoring (5 years training and 1 testing). The best mean prediction accuracy and most stable results were obtained by particle swarm optimisation–support vector machine (PSO–SVM) and particle swarm optimisation–least squares support vector machine (PSO–LSSVM), that led to mean RMSE = 12.4420mm- $R^2 = 0.9483$; RMSE = 45.9456 mm- $R^2 = 0.9710$ and RMSE = 17.2830 mm, $R^2 = 0.9750$. Ref. [21] compared Support Vector Machine Regression, XGBoost and deep learning-based RNN models for displacement prediction on a landslide region located in China, where the authors recorded monitoring data of precipitation, soil moisture and slope displacement during and after rainfall events. XGBoost algorithm outperformed the other two regression models due to XGBoost's ability to better capture the nonlinear information with a small number of data samples provided for the prediction of large short-term displacements (time history prediction for approximately 6.5 unseen hours).

In summary, the above studies have demonstrated that ensemble algorithms perform best with relatively small training datasets for daily relative and cumulative landslide displacement prediction, with a good performance of up to 30 days. This motivates our approach to explore ensemble algorithms in more detail with up to 30-day accumulation time windows for displacement prediction. However, most of the above studies are limited in that they use only precipitation and GWL measurements as indicators of displacement for prediction.

1.2. Summary of Contributions

While the above reviewed studies demonstrate the value of machine learning in predicting landslide displacement, they do not investigate the optimal set of physical indicators needed to provide accurate prediction while minimising measurement, data collection and processing effort. In this study, we use a data-driven machine learning approach to explore the nonlinear relationships between the large range of near-surface (including meteorological) and subsurface measurements taken at the active and heavily instrumented Hollin Hill Landslide Observatory (HHLO), which experiences ongoing slope movements. We

propose a methodology that tackles multi-modal instrumentation measurements, collected at relatively low spatial and temporal resolution to shed light on our currently limited understanding of temporal and spatial causalities between precipitation and displacement and enable development of robust data-driven complex engineering solutions to mitigate the devastating effect of slope instabilities. Unlike [15,18–20], we make predictions of landslide movements through the exploration of a wide variety of 18 influencing factors, including rarely measured ground parameters, such as soil moisture from multiple sensors, soil temperature at multiple depths, soil heat flux and solar net radiation, but also air pressure, air temperature, wind speed and wind direction, in an attempt to discover the optimal subset of parameters that lead to high prediction accuracy. Unlike [8], our work includes ground parameters, such as soil moisture, soil temperature and soil heat flux. Our data-driven contribution towards understanding the relationship between influencing factors of slope stability with respect to displacement and slip explosiveness leverages upon feature selection providing a physics-based explanation of the influencing factors associated with indicators measured on a landslide zone.

The contributions of this paper can be summarised as follows:

- Statistical analysis of a comprehensive database of 18 influencing factors in the form of multivariate time series recordings, exploring the correlation between pairs of recordings and removing multicollinearity from multiple correlated recordings. The objective is to identify a unique set of distinct features (Section 3).
- Feature extraction and embedded feature selection of the 18 influencing factors in order to determine which subset of recordings are most important in predicting time series displacement via three types of regression: Lasso, Random Forest and XGBoost. Regression performance is compared with features obtained from statistical correlation analysis above (Section 4).
- Unsupervised predictive agglomerative clustering to identify distinct types of displacement from the features identified above and visualised via a dendrogram. Clustering also explains visually why no one feature in isolation (inc. precipitation) is sufficient to characterise types of displacement (Section 5).

This paper is organised as follows. In Section 2, we introduce the dataset and data pre-processing steps needed for continuous data analysis. This is followed by our three contribution sections, as described above, before discussing key findings in Section 6 and concluding in Section 7.

2. Dataset from Hollin Hill Landslide Observatory

Hollin Hill is a moisture-induced landslide zone [2] that lies to the north of York in UK. It is several hundred metres wide and extends two hundred metres downslope. Located on the south-facing side of a degraded Devensian ice-margin drainage channel, the slope has an angle of approximately 12° . The slope at HHLO consists of Redcar Mudstone and Whitby Mudstone at the base, with an outcrop of the Staithes Sandstone Formation ('Middle Lias') running across the middle section of the slope. See [22,23] for a more detailed description of the site and the map.

Table 1 lists the full set of sensors deployed at the site together with the resolution of recordings provided. Placed heat flux plates, G1 and G2, measure soil heat flux at a depth of 3 cm (Model: Hukseflux HFP01SC self-calibrating heat flux plate). Near-surface soil temperature (STP) is measured at five depths (2, 5, 10, 20 and 50 cm) using a profile of thermocouples (Model: Hukseflux STP01, self-calibrating heat flux plate). Soil moisture sensors (Model: Acclima Digital TDT Soil Moisture Sensor) at depth of 10 cm use the time domain transmissometry (TDT) technique and provide absolute volumetric water content (TDT1VWC and TDT2VWC) and soil temperature (TDT1SOIL and TDT2SOIL). The soil moisture data are not calibrated to the site specific soil type, but rely on generic calibration information. Automatic weather station measures air temperature and relative humidity by a probe situated within a naturally aspirated radiation shield (Model: Rotronic HC2A-S3 within the Gill MetPak Pro Base Station). Precipitation is measured through Digital

weighing rain gauge (Model: OTT Pluvio), which provides data on the amount and intensity of solid and liquid precipitation, Tipping bucket rain (TBR) gauge (Model: EML SBS 500), which gives data on the amount of liquid precipitation at 0.2 mm resolution and Tipping weighing rain gauge (Model: Lambrecht Raine), which provides greater data reliability when the Pluvio rain gauge data is offline. Wind speed and wind direction are measured through a 3D sonic anemometer (Model: Gill WindMaster 3D Sonic Anemometer), which monitors wind speeds of 0–50 m/s (0–100 mph) while an integrated sonic anemometer is used for high-accuracy wind speed and direction measurement with automatic weather station (Model: Gill Integrated WindSonic). Finally, a four-component radiometer measures the individual radiation components using upward and downward facing pyranometers and pyrgeometers (Model: Hukseflux four-component radiometer).

Table 1. Time series recordings considered from the Hollin Hill Landslide Observatory. Refer to [24] for a detailed description of each of the sensors 1 to 18 and [22,25] for Displacement.

No.	Variable	Physical Quantity	Units	Sensor	Resolution
1	PRECIP	Precipitation	mm	Rain gauge	30 min
2	RN	Net Radiation	Wm ⁻²	Radiometer	30 min
3	G1	Soil Heat flux 1	Wm ⁻²	Soil heat flux plate	30 min
4	G2	Soil Heat flux 2	Wm ⁻²	Soil heat flux plate	30 min
5	PA	Atm. Pressure	hPa	Cosmic-Ray Neutron Sensor (CRNS)	30 min
6	TA	Air Temperature	°C	Automatic weather station	30 min
7	WS	Wind Speed	ms ⁻¹	Integrated 2D sonic anemometer	30 min
8	WD	Wind Direction	deg	Integrated 2D sonic anemometer	30 min
9	RH	Relative Humidity	%	Automatic weather station	30 min
10	TDT1TSOIL	Soil Temperature	°C	Soil temperature sensor at 10 cm	30 min
11	TDT1VWC	Soil Moisture	%	Point soil moisture sensor	30 min
12	TDT2TSOIL	Soil Temperature	°C	Soil temperature sensor at 10 cm	30 min
13	TDT2VWC	Soil Moisture	%	Point soil moisture sensor	30 min
14	STPTSOIL2	Soil Temperature	°C	Soil temperature sensor at 2 cm	30 min
15	STPTSOIL5	Soil Temperature	°C	Soil temperature sensor at 5 cm	30 min
16	STPTSOIL10	Soil Temperature	°C	Soil temperature sensor at 10 cm	30 min
17	STPTSOIL20	Soil Temperature	°C	Soil temperature sensor	30 min
18	STPTSOIL50	Soil Temperature	°C	Soil temperature sensor at 50 cm	30 min
19	DISP	Displacement	mm	Leica System 1200 RTK	60 min

In this paper, we used all timestamped recordings during the period from 25 March 2014 to 9 March 2022, during which there were two catastrophic landslides presenting explosive landslide movement, as shown in Figure 1 (at the start of 2016 and 2018) obtained from DISP measurements (Table 1). The Leica System measuring displacement [22,25], consists of a grid of sensors, and in this paper, only “sensors-9”, placed at the eastern lobe of the hill were used, since they showed the stages of failure most prominently.

As seen in Figure 1, two periods of mass or explosive movement and three periods of intermittent movement can be identified through the displacement recordings. Indeed, the first period of explosive movement lasted from mid December 2015 to mid April 2016, while the second period of explosive movement lasted from the end of November 2017 to the end of April 2018.

Pre-Processing: Data Cleaning, Gaps, Interpolation and Downsampling of the Data

All the non-cumulative weather data were downsampled from half-hour recordings to mean values per day, while for displacement, we used the daily cumulative displacement value obtained by summing all recordings collected hourly in a day. As per [18], we downsampled the data to one measurement per day, to reduce noise and smooth short-term fluctuations, as well as achieve computational efficiency, while also being able to obtain a higher-level overview of data patterns.

The displacement recordings were transformed to absolute values by substitution of the first recorded point (reference) from the Leica System and then interpolated (where small data gaps were present in the recordings) to capture continuous and differentiable stages of failure. After the absolute displacement was interpolated, the relative velocity time history (or daily differential displacement) is extracted through numerical differentiation, and is a generated indicator feature that we consider in addition to absolute displacement.

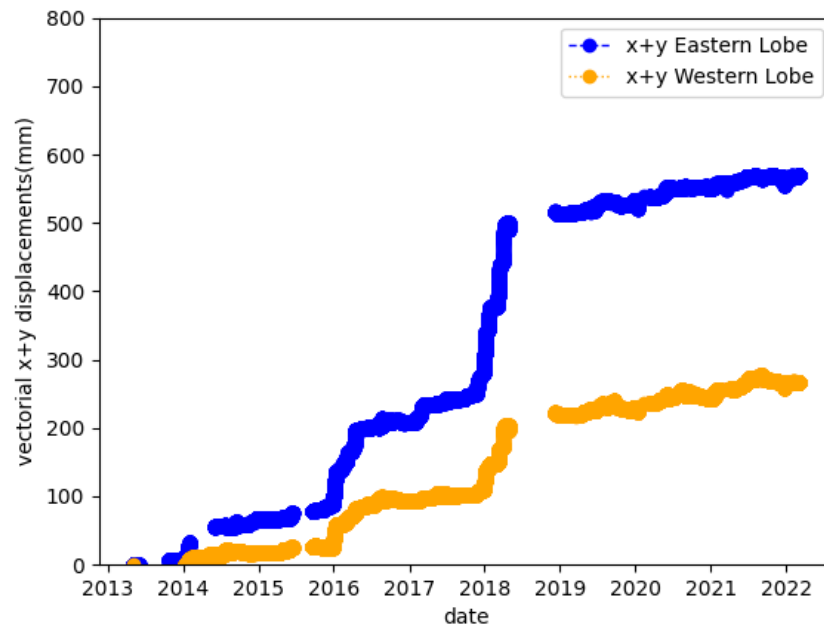


Figure 1. Displacement recordings transformed to absolute plane vectorial displacements (x-coordinate + y-coordinate) for eastern and western lobes of the slope.

Before performing feature selection of the data points in Section 4, we normalised recorded values to zero mean and unit variance. Before performing agglomerative clustering in Section 5, for visualisation, we scaled the data, so that all features belong to the same range of values.

3. Methodology for Exploring Statistics of Multivariate Time Series Recordings

The first approach towards understanding the multivariate measurements is to perform statistical analysis across these measurements. We explore: (1) Correlation Heat Map, which quantifies the correlation values between pairs of time series measurements, and (2) Variational Inflation Factor, which provides a global view across all multivariate time series measurements, removing multicollinearity from multiple correlated variables.

3.1. Correlation Analysis

The correlation between pairs of all influencing factors is shown in Table 1, and the correlation between each influencing factor and the displacement and velocity indicators (daily differential displacement) is calculated using the correlation coefficients, and is shown in Figure 2. The correlation matrix shows the strength (closer to magnitude 1) and direction of the correlation as a value between -1 and 1 , where a negative value indicates that as one variable increases, the other decreases, whereas a positive value indicates positive correlation.

It is worth noting that, firstly, the displacement (disp) and velocity (vel) indicators are only weakly correlated to the influencing factors, hinting that these indicators are functions of multiple influencing factors that should be considered jointly. Secondly, as expected, there are many subsets of highly correlated influencing factors, e.g., all the variables related to temperature or all the variables related to energy (e.g., net radiation and soil heat flux) are highly mutually correlated. Hence, the dimensionality of the influencing factors to

be measured could potentially be reduced without losing relevant information. Note that we can identify six distinct, less correlated with others, groups of influencing factors. These are: (1) precipitation, atmospheric pressure, (2) wind speed, (3) wind direction, (4) relative humidity, (5) net radiation, soil heat flux, air temperature and soil temperature and (6) soil moisture.

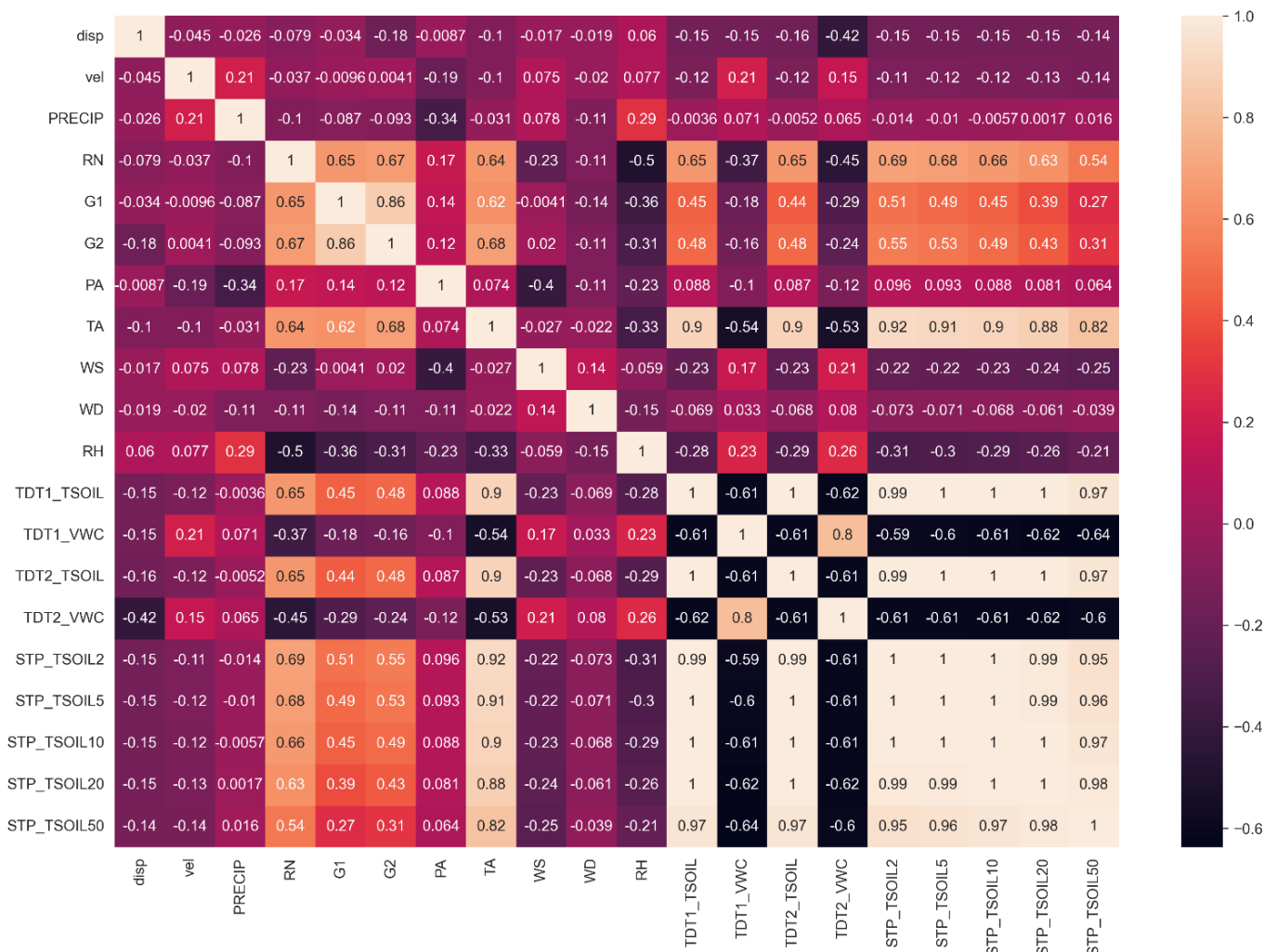


Figure 2. Heat map of correlation values between pairs of influencing factors. The color bars' values on the right side of the Figure indicate how strongly the factors are correlated.

Variational Inflation Factor (VIF)

While the correlation matrix provides an indication of correlation between pairs of variables, VIF takes a more global view across variables and removes the multicollinearity that arises from multiple correlated variables [26]. Given n independent variables (influencing factors—the first 18 rows of Table 1), X_i , the VIF algorithm in each iteration, sets one independent variable as a target, and builds a predictor as a weighted linear combination of all other independent variables

$$X_i = \beta_0 + \sum_{k,k \neq i} \beta_k X_k. \tag{1}$$

Finally, the amount of multicollinearity is quantified by calculating VIF for the independent variable i as

$$VIF_i = 1 / (1 - R_i^2) \tag{2}$$

where R_i^2 represents the coefficient of determination for regressing the i th variable on all other independent variables, as shown in (1). If variable X_i is uncorrelated to other variables, $R_i^2 = 0$ and $VIF_i = 1$. VIF below 5 is usually accepted as small-to-moderate multicollinearity, which is how we selected the five features that VIF considers the most distinct globally, as shown in Table 2. These are precipitation (PRECIP), wind speed (WS), net radiation (RN) and soil heat flux for eastern (G1) and western lobes (G2). Whilst precipitation and wind speed were identified as unique in Figure 2, VIF stresses the additional uniqueness of soil temperature/humidity in the forms of net radiation and soil heat flux. However, the VIF analysis does not indicate correlation between these factors and displacement, causing a danger that some of the 13 variables that are removed could be more correlated to displacement than the 5 retained features.

Table 2. VIF iterations showing influencing factors selected. “X” indicates the factors that are removed at each iteration.

Feature	VIF 1st it.	VIF 2nd it.	VIF 3rd it.	VIF 4th it.	VIF 5th it.	VIF 6th it.	VIF 7th it.	VIF 8th it.	VIF 9th it.	VIF 10th it.	VIF 11th it.	VIF 12th it.	VIF 13th it.	VIF 14th it.
PRECIP	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.30	1.30	1.30	1.27	1.24	1.23	1.21
RN	6.22	6.22	6.2	6.0	6.0	5.9	5.8	5.19	3.96	3.94	3.94	1.24	2.45	1.91
G1	7.6	6.7	6.6	6.3	5.4	5.0	4.5	4.36	4.32	4.24	4.21	4.17	4.17	4.12
G2	9.2	9.0	8.9	6.3	7.0	6.9	5.6	5.17	4.79	4.75	4.46	4.33	4.29	4.18
PA	233.2	233.0	230.4	6.3	213.5	213.5	212.1	X	X	X	X	X	X	X
TA	68.8	68.8	68.5	65.5	63.8	63.8	63.6	63.57	14.26	14.25	8.85	7.09	X	X
WS	6.0	5.9	5.9	5.9	5.8	5.8	5.8	5.44	4.95	4.88	4.88	4.72	4.04	1.67
WD	12.3	12.3	12.2	12.2	12.2	12.1	12.1	10.97	10.96	10.85	10.62	X	X	X
RH	131.5	131.3	128.1	123.8	123.5	123.3	119.5	46.77	39.64	37.42	X	X	X	X
TDT1TISOIL	4333.2	4300.5	4134.7	3553.5	X	X	X	X	X	X	X	X	X	X
TDT1VWC	50.1	49.8	49.7	49.6	49.0	49.0	45.9	44.72	44.64	20.36	8.17	5.90	5.33	X
TDT2TISOIL	7541.2	7500.5	7027.5	3278.0	2669.4	X	X	X	X	X	X	X	X	X
TDT2VWC	50.0	49.5	50.0	50.0	49.6	49.5	48.5	47.05	46.47	X	X	X	X	X
STPTSOIL2	287,868.1	13,947.1	2804.8	2560.9	1760.6	488.5	90.5	85.92	X	X	X	X	X	X
STPTSOIL5	1,301,313	X	X	X	X	X	X	X	X	X	X	X	X	X
STPTSOIL10	827,317	110,355.7	X	X	X	X	X	X	X	X	X	X	X	X
STPTSOIL20	130,527	73,352.3	15,947.3	X	X	X	X	X	X	X	X	X	X	X
STPTSOIL50	4538.0	4048.9	1666.8	853.1	771.9	500.2	X	X	X	X	X	X	X	X

4. Methodology for Feature Extraction and Selection for Predicting Landslide Movements

Whilst the correlations between pairs of influencing factors and across influencing factors have shown unique variables in the form of precipitation and wind speed (meteorological measurements) and soil heat flux, and the VIF analysis identified the five most distinct factors, these findings do not consider the importance of each of these factors in relation to displacement.

The objective of our study is to identify, via feature extraction and embedded feature selection for regression, which sensor recordings have the strongest influence on relative displacement prediction. In particular, Linear Discriminant Analysis, as a supervised dimensionality reduction approach, is used for feature extraction, since it can transform the feature space in relation to displacement. Since our aim is to effectively predict time series displacement, we leverage on popular Lasso, RF and XGBoost embedded feature selection from the 18 time series recordings and demonstrate their effectiveness during displacement prediction. RF and XGBoost are widely adopted in the literature for landslide displacement prediction, as per the review in Section 1.1. Both are generally popular ensemble regression algorithms that are robust to relatively smaller training sets compared to deep learning neural networks. This makes them ideal for our study. We also included Lasso regression because it is a relatively less complex model with fewer parameters and shorter execution time.

4.1. Predictive Performance Evaluation Metrics

The quantified predictive performance analysis has been performed using the following metrics: root mean squared error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2) as presented in Equations (3)–(5), respectively.

$$RMSE = \sqrt{\frac{\sum_{i=1,N}(x_i - x'_i)^2}{N}} \quad (3)$$

$$MAE = \frac{\sum_{i=1,N} |x_i - x'_i|}{N} \quad (4)$$

$$R^2 = \frac{(\sum_{i=1,N}((x_i - x_{im})(x'_i - x'_{im})))^2}{\sum_{i=1,N}(x_i - x_{im})^2 \sum_{i=1,N}(x'_i - x'_{im})^2} \quad (5)$$

where N is the number of samples, x_i and x'_i are the measured and predicted value, respectively, x_{im} and x'_{im} are the average measured and predicted value, respectively.

4.2. Linear Discriminant Analysis for Feature Extraction

Linear Discriminant Analysis (LDA) is a method mostly used in statistics that searches linear combinations of features to better explain a large dataset, and is often used for dimensionality reduction purposes. LDA is a supervised approach that exploits eigenvalue decomposition to find the projection of the data that minimises the inter-class variance and maximises the distance between the projected means of the classes [27].

As in earlier work [28] where feature extraction was studied in depth for this dataset, each measured data point is labelled into the following 5 classes of stages of failure: 1st intermittent (from 25 March 2014 until 29 December 2015), 1st explosive (from 30 December 2015 until 15 April 2016), 2nd intermittent (from 16 April 2016 until 28 December 2017), 2nd explosive (from 29 December 2017 until 2 April 2018) and 3rd intermittent (from 3 April 2018 until 9 March 2022) (as per Figure 1). In [28], three dimensionality reduction methods were compared, namely, LDA, 2-dimensional Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE), in their ability to separate the 5 classes. It was concluded that LDA better differentiated the data points (see Figure 3) compared to other two methods. Furthermore, the two LDA components led to the best prediction performance for the residual part of the cumulative displacement time series after decomposition of the initial signal into periodic, trend and random parts and using XGBoost regression [28].

Note that LDA uses displacement for class labelling, hence capturing the correlation between the measurements and the target, but by projecting the data onto a new coordinate system, loses the information about initial sensor recordings.

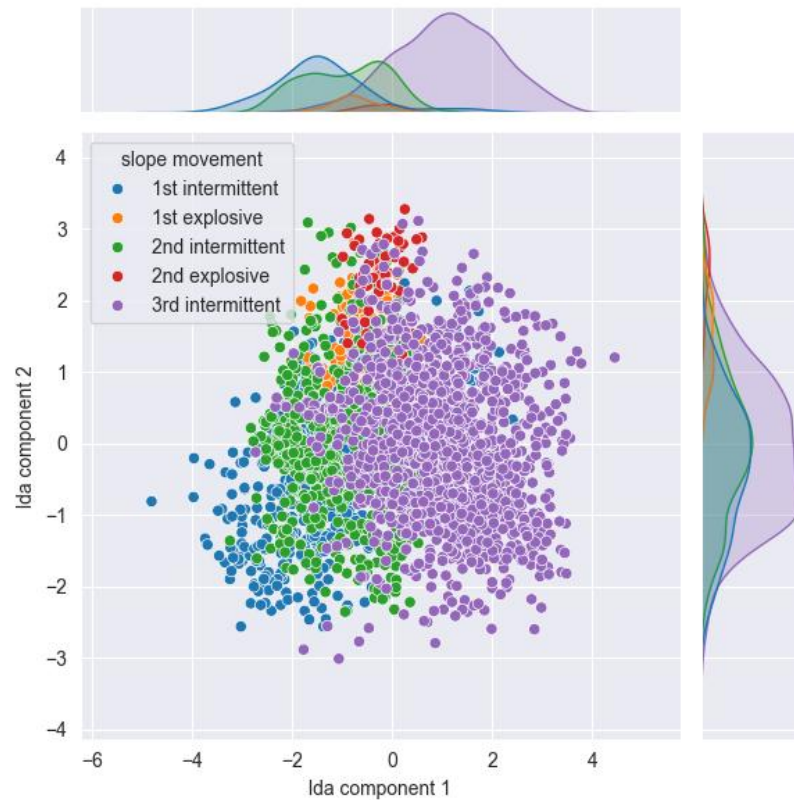


Figure 3. A 2D representation with 2 LDA components of 5 patterns of failure movement [28].

4.3. Feature Selection

Lasso is a popular embedded feature selection method, widely used to improve predictions of regression algorithms. It uses a control parameter, α , in the L1 penaliser to control the number of selected features, whereby the higher the value of the control parameter, the fewer features selected [29]. For our model, the α parameter was selected to be equal to 0.00001, extracted through GridSearchCV that was used for hyperparameter tuning. RF constructs and fits a number of decision trees on various sub-samples of the dataset and uses mean average prediction of the individual trees to improve the predictive accuracy and control over-fitting [30,31]. For our RF model, the following parameters were set: $n_estimators = 1000$, $random_state = 42$, $criterion = squared_error$, $min_samples_split = 2$, $min_samples_leaf = 1$. XGBoost is another ensemble learning algorithm that is particularly suited for efficient performance for regression tasks for large datasets. For our XGBoost model, the following parameters were set: $objective = reg:squarederror$, $n_estimators = 1000$, $nthread = 24$. Figure 4 shows the obtained relative feature importance scores. Since there is a relatively large importance gap between the 4th and the 5th most important features for Lasso, we draw a line at 0.35, where the influencing factors TDT1TSOIL, TDT2TSOIL, STPTSOIL2 and STPTSOIL10 are selected for relative displacement prediction. For RF, similarly to the Lasso case, we draw the line at 0.05, selecting the influencing factors PRECIP, PA, TDT1VWC, TDT2VWC as the most important for daily differential displacement. These represent precipitation, atmospheric pressure and soil moisture for daily differential displacement. With XGBoost, we draw the line at 0.05, selecting the influencing factors PRECIP, PA, WD, TDT1VWC and TDT2VWC. These represent precipitation, air pressure, wind direction and soil moisture. These selected features together with those selected by VIF are summarised in Table 3, and are used next for displacement prediction. It is interesting to note that precipitation was selected by VIF, RF and XGBoost. This is inline with previous studies. As expected, both ensemble methods selected the same set of fea-

tures (precipitation, atmospheric pressure and soil moisture) as important except for wind direction. Lasso and VIF do not have any features in common.

Table 3. Features selected by each method represented by black dots.

Feature	VIF	Lasso	RF	XGBoost
PRECIP	•		•	•
RN	•			
G1	•			
G2	•			
PA			•	•
TA				
WS	•			
WD				•
RH				
TDT1TSOIL		•		
TDT1VWC			•	•
TDT2TSOIL		•		
TDT2VWC			•	•
STPTSOIL2		•		•
STPTSOIL5				
STPTSOIL10		•		
STPTSOIL20				
STPTSOIL50				

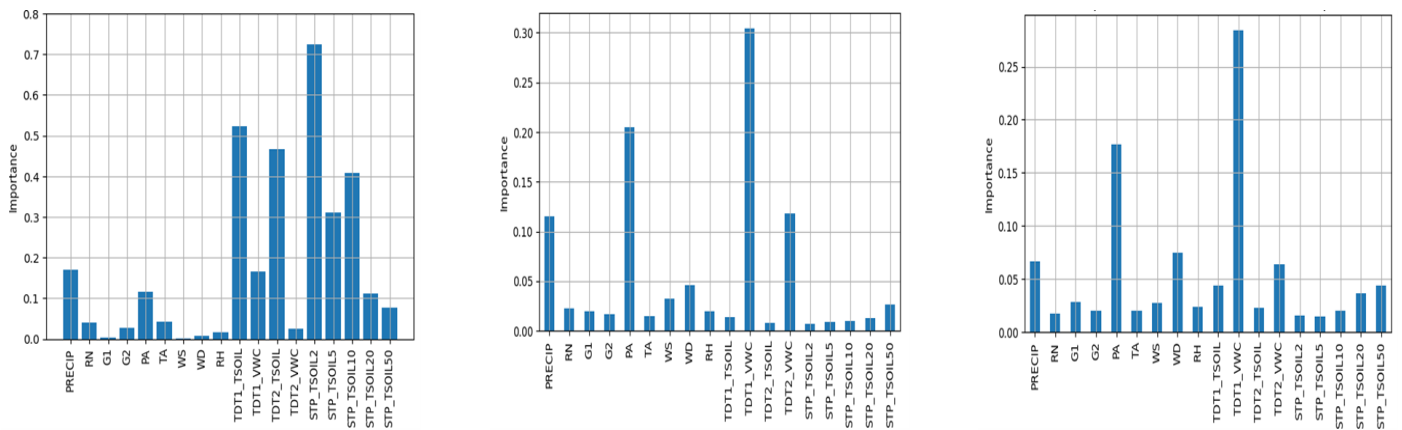


Figure 4. Left to right: Lasso feature importance ranking, RF regression feature importance ranking and XGBoost regression feature importance ranking, with respect to relative displacement.

4.4. Prediction Performance

In order to validate the effectiveness of the above feature extraction and feature selection methods, the following experiments are performed during displacement prediction:

1. Regression using Lasso, RF and XGBoost with training/testing split ratio of 70/30%. We output feature importance scores and select only the most important, i.e., the highest scoring features; we compare the accuracy of landslide movements with the selected features vs. the case when all 18 features are used for daily prediction on unseen movements of the last intermittent failure only. The results are shown in Table 4 along the 70/30 rows.
2. We predict unseen landslide movements, in the form of relative displacement points, of the last intermittent failure and second major failure of 2018 by reducing the training/testing split ratio to 50/50%, with and without feature selection as per Table 4 for the three regression methods, and in the case of RF-LDA, with 4 and 2 extracted features. The results are shown in Table 4 along the 50/50 rows.

- Based on the accuracy of the predictability of the models trained on relative displacement, we attempt to indirectly predict the absolute displacements on various time windows (i.e., 1 days, 5 days, 10 days, 15 days and 30 days) by training on daily resolution and summing the predicted daily differential displacements and comparing them across the 3 regression methods. The results are shown in Table 5 and Figure 5.

The results of predicting relative displacement with and without feature selection for all 3 regression methods for 70/30% and 50/50% train/test set split ratios can be found in Table 4. The most significant observation is that, for all feature selection and regression methods, performance was improved with feature selection ('selected') compared to using all 18 features. Multicollinearity is known to limit the accuracy of predictive models by increasing model complexity and causing overfitting. Results for all 3 methods are similar, with RF negligibly better than the other methods.

Table 4. Daily prediction of relative displacement: performance of Lasso, RF and XGBoost with embedded feature selection, LDA feature extraction and displacement-agnostic VIF feature selection (denoted as 'selected') vs. all 18 features for regression (denoted as 'all'). The units of RMSE and MAE are both mm/day. Values in bold represent best performance.

Features	Lasso	RF	XGBoost	RF-LDA	RF-VIF
SPLIT RATIO	RMSE MAE	RMSE MAE	RMSE MAE	RMSE MAE	RMSE MAE
all 70/30	0.75 0.51	0.57 0.29	0.57 0.31	0.57 0.30	- -
selected 70/30	0.60 0.37	0.56 0.30	0.57 0.33	0.54 0.29	0.66 0.37
all 50/50	0.87 0.52	0.89 0.49	0.97 0.52	0.78 0.35	- -
selected 50/50	0.82 0.42	0.81 0.44	0.91 0.48	0.78 0.34	0.86 0.41

Table 5. Performance and complexity in terms of training and test time of Lasso, RF and XGBoost for prediction of cumulative displacement over $t = 1, 5, 10, 15, 30$ days of accumulation. The units of RMSE and MAE metrics are mm/($t * \text{days}$) for both metrics. The indication 70|50 on the 2nd row of the Table refers to the size of the training set (of relative displacements) that is used in order the model to be fitted and be able to predict the cumulative displacement time series in the various time windows. Values in bold represent best performance.

Time Window	Metric	Lasso	RF	XGBoost
		train time: 2154 ms test time: 1583 ms train sizes: 70 50	train time: 28,080 ms test time: 2160 ms train sizes: 70 50	train time: 3106 ms test time: 1745 ms train sizes: 70 50
1 d	RMSE	65.496 90.971	37.047 118.178	36.168 87.924
	MAE	52.884 72.349	26.712 70.743	25.636 53.854
	R ²	0.895 0.797	0.966 0.657	0.968 0.810
5 d	RMSE	65.438 90.197	28.576 113.564	16.082 112.793
	MAE	52.376 71.350	22.306 70.166	11.163 83.407
	R ²	0.895 0.800	0.980 0.683	0.994 0.687
10 d	RMSE	65.300 89.897	37.441 173.928	47.703 132.402
	MAE	52.016 70.927	33.391 121.930	39.389 101.678
	R ²	0.895 0.801	0.966 0.256	0.944 0.569
15 d	RMSE	65.857 88.730	30.847 170.488	94.797 121.531
	MAE	51.459 69.427	27.584 120.744	90.365 103.779
	R ²	0.893 0.806	0.977 0.285	0.779 0.637
30 d	RMSE	62.334 99.616	60.608 214.624	63.358 146.955
	MAE	50.819 77.573	50.051 150.633	49.667 104.832
	R ²	0.904 0.755	0.909 0.139	0.901 0.466

The dimensions of the original dataset are 18×2709 . After feature selection, this is reduced to 4×2709 for Lasso and RF, and 5×2709 for VIF and XGBoost. The transformed feature space with all LDA components has a dimension of 4×2709 , and is reduced to

2×2709 for ‘selected’ features. The 70/30% train/testing split ratio is the most commonly used ratio in machine learning as it provides a significant amount of the data for training without compromising on sufficient data for testing. In our experiments, the performance with 60/40%, 70/30% and 80/20% split ratios are similar. In contrast, when reducing the training set and increasing the testing set through adoption of the 50/50% ratio, we demonstrate the robustness of the regression algorithms to reduced training sets as well as demonstrating prediction for the second unseen major failure of 2018 via prediction of explosive movements.

As expected, we observe that performance with 70/30% training/testing split ratio is better across all experiments for performance with 50/50% training/testing split ratio. While RF and XGBoost have similar performance (but better than Lasso) for the larger training set, we observe that RF is more robust to a relatively smaller training set (like Lasso) compared to XGBoost. Overall, RF has the best performance; therefore, we use RF to compare the effect of physical feature selection vs. feature extraction in the LDA transformed domain, as well as VIF-selected features which are independent of displacement. Note that in the case of RF-LDA ‘all’, all 4 LDA components were used as features vs. 2 for ‘selected’ as shown in Figure 3. Performance of RF with embedded feature selection vs. LDA feature extraction are similar for the 70/30% training/testing split ratio, but the former is more explainable since we know the physical features used. However, as observed by better performance of RF-LDA compared to RF for the 50/50% training/testing split ratio, we conclude that LDA feature extraction captures marginally better the displacement with a smaller training set than embedded feature selection. RF with VIF-selected features, being agnostic of displacement, has worse performance than with LDA or embedded RF feature selection.

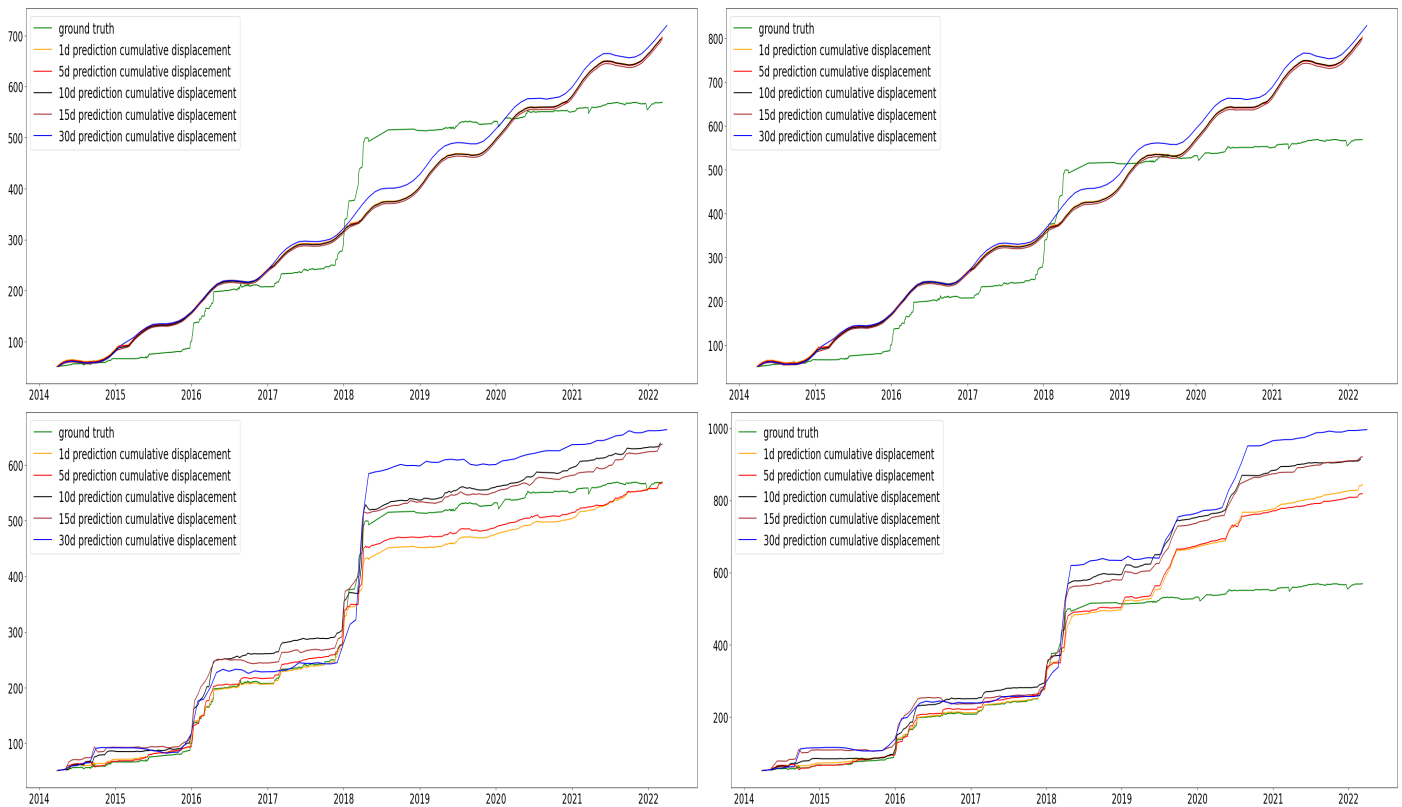


Figure 5. Cont.

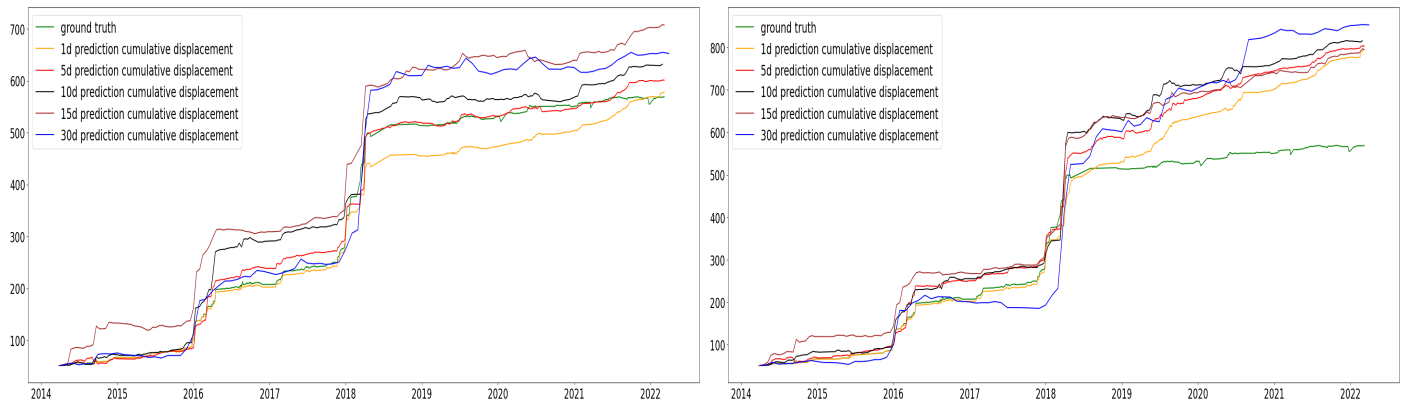


Figure 5. Horizontal axis on all the graphs shows time [days], while vertical shows cumulative displacement [mm]. Split ratio 70/30 (**left**); split ratio 50/50 (**right**). Lasso 1st row, RF 2nd row and XGBoost 3rd row are presented for recorded–predicted cumulative displacement on various time windows (1 d, 5 d, 10 d, 15 d, 30 d).

As shown in Figure 1, five distinct regions of displacement patterns can be observed according to the recorded gradient. That is, from 2014 to late 2015, the first intermittent failure can be observed, followed by the first major failure (explosive region) in early 2016, the second intermittent failure from 2016 to late 2017, the second major explosive failure in early 2018 and finally, from 2018 to 2022, the third intermittent region of displacement. The 70/30% train/test set split predicts only the last intermittent failure. Generalisation to different types of failures is shown by the 50/50% split which predicts the second explosive failure in addition to the last intermittent failure. Table 4 shows that in this case the performance drop in predicting the second explosive failure is negligible for all 3 methods. As above, results for all 3 methods are similar, with RF being negligibly better.

4.5. Prediction of Cumulative Displacement in Accumulation Time Windows of Various Sizes

Motivated by the good predictability of the models trained on relative displacement, we further attempt to indirectly predict the accumulated displacements on time windows of various sizes, such as $t = 1, 5, 10, 15$ and $t = 30$ days, by training on daily resolution relative displacement and then summing the predictions. The performance of prediction for all 4 methods can be seen in Table 5 for the training sizes of 70% and 50% of the total data. The time window is also used for the accumulation of the influencing factors that are selected by each of the 3 methods. Once the averaged accumulated relative displacement is predicted, then the cumulative accumulated displacement is calculated on the time window, according to the following equation:

$$disp[i] = t * disp_{mrel}[i] + disp[i - 1] \quad (6)$$

where $disp$ is the cumulative displacement array, $disp_{mrel}$ is the mean relative displacement array calculated on the examined time window, and t is the size of the accumulating time window. The results are shown in Table 5.

For the prediction of cumulative displacement in time windows, RF performed best for the 70/30% training/testing split ratio for all the provided time windows except the 5-day-window for which XGBoost outperformed other methods. However, Lasso is more robust to a smaller training set, consistently outperforming the other two regression methods for all time windows. Furthermore, Lasso has the shortest run time. Given the focus of our study on reducing computational effort, we show that relatively simpler models like Lasso can achieve comparable results to more complex ensemble models adopted in other studies. The models have been tested on a large multivariate dataset, their performance compared with different training and testing ratios to demonstrate generalisability for both gradual and explosive failure prediction.

Figure 5 shows the same set of results in terms of cumulative displacement vs. time, that is, the prediction of cumulative displacement over 1-, 5-, 10-, 15- and 30-day periods to verify how well the 3 regression models capture the unseen slope movements visually. As observed in Figure 5, for the 70/30% training/testing split ratio, XGBoost does indeed closely follow the ground-truth (in green) for the 5 day window (in red) and RF for the larger prediction windows. This is in line with the results in Table 5. Indeed with XGBoost and RF, the different stages of failure can be predicted more accurately than with Lasso. However, for the 50/50% training/testing split ratio, RF for all time windows generated 2 major false gradients after the 2018 failure that do not correspond to recorded explosive movements. XGBoost, on the other hand, performed better than RF without false gradients for the 50/50% training/testing split ratio, which is in line with the results of Table 5, accurately capturing failure. Overall, for both split ratios, both RF and XGBoost accurately capture the magnitude of the total displacement increment that occurred during the unseen major event for all time windows except for the 30 d period (the 2nd vertical section in the beginning of 2018, in the graphs of the 2nd and 3rd rows shown in Figure 5). Whilst Lasso, in terms of quantitative metrics (Table 5), performed better than other methods for larger time windows (10 d, 15 d, 30 d) for the 50/50% training/testing split ratio, it captured the trend well, but did not succeed in predicting any failure patterns of explosive failures and intermittent landslide movements, as seen in all cases in Figure 5. This is due to Lasso's tendency to smooth predictions. Visual explanation of failure prediction results, shows that XGBoost, with a relatively smaller run time than RF and comparable to Lasso, is the model with highest accuracy in capturing unseen failure. Therefore, performance metrics are not always a good indicator of particular events since they average performance, and visual reconstruction is also needed.

5. Methodology for Unsupervised Detection of the Stages of Landslide Displacement

In the previous section, we proposed several feature selection methods and discussed how effective these methods are for prediction of relative landslide displacement. Next, we will assess the suitability of the selected features for the task of clustering the data points in time, for the purpose of grouping the samples to identify different stages of landslide displacement in an unsupervised manner. To perform clustering of the HHLO recordings, we use dendrograms and agglomerative or bottom-up hierarchical clustering, as a popular approach that does not require the number of clusters to be pre-specified.

Hollin Hill Observatory is a landslide zone where failure has been monitored through the years with heavy instrumentation and occasional visual confirmation. In the landslide zone no man-made events triggering or leading to failure have purposefully taken place over the eight years of recordings, which could have influenced failures. Additionally, the site is remote, far from residential areas and roads and human activity in general. The adopted data-driven approach aims to provide a framework for failure prediction through continuous site monitoring not focusing on the material investigation but on the relationship between environmental recordings and previously recorded slope movement patterns. So, conditions were considered only through physical explanation of the inter relationships between the ground parameters and not directly as predictors which serves as the scope of this study.

5.1. Clustering Performance Evaluation Metrics

To assess the performance of clustering methods, Minkowski distance is often used. This distance determines the similarity of distances between two or more vectors in space, as is given by

$$D_M(X, Y) = (\sum_{j=1}^N |x_j - y_j|^p)^{\frac{1}{p}} \quad (7)$$

where x_j, y_j are the j -th elements of N -dimensional data vectors X and Y , respectively, and $D_M(X, Y)$ is the distance between them. Minkowski distance is often used with $p = 1$ or $p = 2$, which correspond to the Manhattan distance and the Euclidean distance, respectively.

The agglomerative clustering method predicts subgroups of data within the data. This is achieved through calculating the distances between each data point (or a cluster of points) and its (their) nearest neighbors and by linking the closest neighbors. We consider the three

most commonly used distance metrics, namely Euclidean, Manhattan and Cosine distance, and three ways to merge the closest neighbors, namely, Ward, Average and Complete linkages.

In order to identify the unique subgroups (clusters), we use dendrogram visualisation, and prune the tree based on a threshold that is set using Silhouette analysis and Calinski–Harabasz Index. The calculation of silhouette coefficient combines inter- and intra-cluster distance into a single score. Specifically, for a given observation o , the score $S(o)$ is calculated as

$$S(o) = \frac{b(o) - a(o)}{\max[a(o), b(o)]} \tag{8}$$

where $a(o)$ is the average distance between observation o and all the other observations in the cluster that o belongs to, and $b(o)$ is the minimum distance from observation o to all clusters to which o does not belong to. The Calinski–Harabasz (CH) index [32] evaluates the cluster validity based on the ratio of the within-cluster variance to the between-cluster variance, where higher values indicate compact and well-separated clusters, and is given by [33]

$$CH = \frac{\text{trace}(S_B)}{\text{trace}(S_W)} * \frac{N - K}{K - 1}, \tag{9}$$

where N is the total number of data points, K is the number of clusters, $\text{trace}(S_B)$ is the trace of the between-cluster scatter matrix that should be maximised and is calculated by (10), $\text{trace}(S_W)$ is the trace of the internal scatter matrix that should be minimised and is computed by (11):

$$\text{trace}(S_B) = \sum_{k=1, K} (n_k \times \|C_k - C\|^2), \tag{10}$$

$$\text{trace}(S_W) = \sum_{k=1, K} \sum_{X_{ik} \in C_k} \|X_{ik} - C_k\|^2 \tag{11}$$

where n_k is the number of observation in cluster k , C_k is the centroid of cluster k , C is the centroid of the dataset and X_{ik} is the i th observation of cluster k .

5.2. Clustering Parameter Selection

As the previous two sections have shown, using all 18 features implies multicollinearity, which adds unnecessary complexity and negatively affects performance. Therefore, we leverage the best features selected from Sections 3 and 4; namely, we perform predictive agglomerative clustering for fitting to relative displacement. These are the five features selected by VIF (PRECIP, RN, G1, G2, and WS) and four by RF (PRECIP, PA, TDT1VWC, TDT2VWC).

Table 6 shows the results in terms of S and CH for three different distance metrics and three different linkage methods, and 2, 3 and 4 clusters. The results indicate that the optimal number of clusters is $n = 2$ for both VIF-based and RF-based feature selection methods. Euclidean distance and Ward linkage leads to the most accurate results.

Table 6. Agglomerative clustering results for features selected via VIF and RF. S scores and CH indices are shown for several combinations of distance metrics, linkage methods and the number of clusters $n = 2, 3, 4$. Values in bold represent best performance.

Distance Metrics	Linkage Distances	VIF, $n = 2$		RF, $n = 2$		VIF, $n = 3$		RF, $n = 3$		VIF, $n = 4$		RF, $n = 4$	
		S	CH	S	CH	S	CH	S	CH	S	CH	S	CH
Euclidean	Ward	0.86	2596.35	0.87	2971.54	0.80	2060.55	0.82	2615.26	0.44	1889.32	0.76	2431.59
Euclidean	Average	0.95	310.54	0.95	337.59	0.90	1219.56	0.92	1416.38	0.87	897.82	0.91	986.19
Euclidean	Complete	0.87	2594.02	0.94	726.28	0.86	1710.72	0.81	1835.78	0.70	1507.52	0.77	2191.65
Manhattan	Average	0.95	310.54	0.95	337.60	0.90	1260.08	0.92	1416.38	0.87	927.53	0.88	1051.01
Manhattan	Complete	0.90	352.97	0.90	2928.77	0.88	1429.78	0.89	1869.05	0.86	1052.41	0.80	1651.85
Cosine	Average	0.40	1.29	0.79	1985.32	0.17	131.91	0.65	1273.78	0.21	352.21	0.54	938.35
Cosine	Complete	0.20	352.97	0.55	857.54	0.14	237.26	0.56	593.37	0.16	282.88	0.48	1831.04

5.3. Results and Discussion

Hence, we set $n = 2$, which defines dendrogram thresholds to be equal to 12.5 and 8, for VIF-based and RF-based metrics, respectively, and in the following use Euclidean distance with Ward linkage.

Figure 6 shows the resulting dendrograms using VIF- and RF-selected features. As discussed, based on Table 6, we prune the tree to obtain $n = 2$ clusters. We can see that, in both cases, pruning leads to a very compact cluster (orange) with very low inter-cluster distance, and another more dispersed and much larger cluster (green).

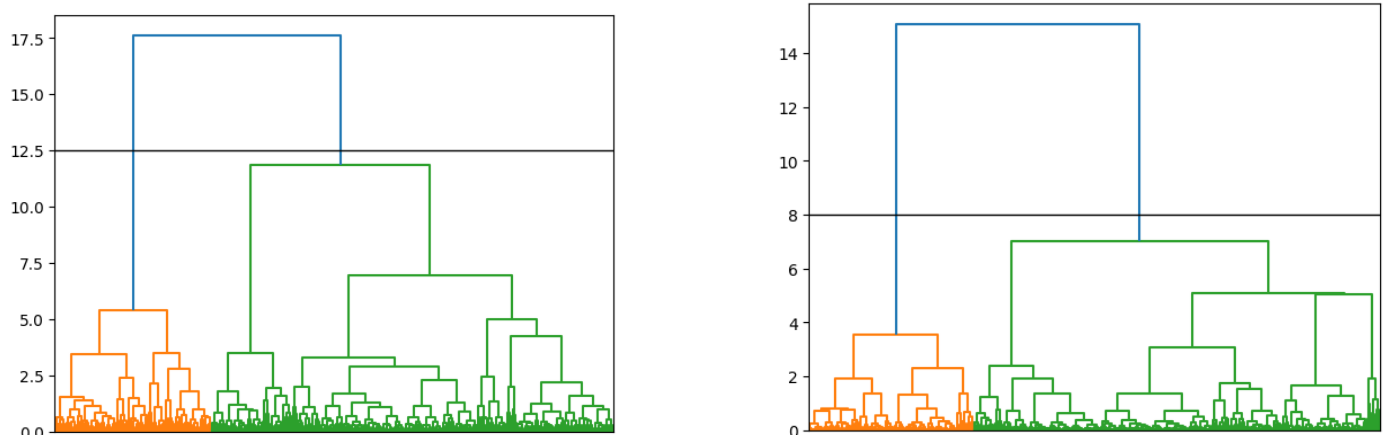


Figure 6. Clustering of data points to explain daily differential displacement, according to features selected by VIF (left) and RF (right): dendrogram with thresholds at 12.5 and at 8, as derived in Section 5.

Figure 7 shows the clustering results with the two methods in the daily differential displacement (in mmday^{-1}) vs. time (in days) plot. It can be seen that both methods led to similar clustering results, successfully isolating major explosive failures (red triangles corresponding to the orange cluster in the dendrogram plot) that took place in 2016 and 2018. This suggests that the identified features indeed capture changes in the relative displacement well, with only few outliers that are similarly positioned in both graphs: around late 2014, mid 2020 and, with the RF method, early 2021.

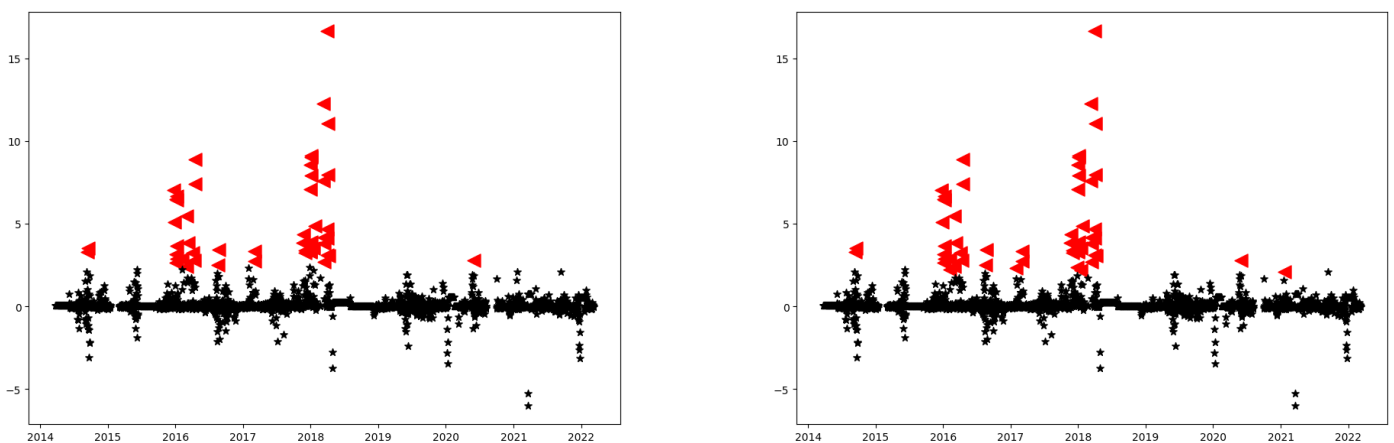


Figure 7. Clustering of data points according to features selected by VIF (left) and RF (right): Daily differential displacement [mmday^{-1}] vs. time [days]. Red indicates clustered data points related to major explosive movements while black indicates points related to intermittent movements.

Figure 8 shows the clustering results presented as each of the selected feature vs. time. It can be seen that the areas of explosive failure happened during high peaks in PRECIP. However, there are a number of outliers, which means that PRECIP alone cannot be used as

a feature for distinguishing the two types of displacement. Note that parameter RN, i.e., net solar radiation, expresses the total amount of solar energy that comes into the soil, i.e., and is generally low between autumn and spring. One can see from Figure 8 that major failures are focused on relatively low values of PA and solar radiation RN, which are associated with cloudiness and rainy days. G, soil heat flux, expresses the amount of thermal energy that moves through an area of soil in a unit of time [34]; daytime peak hourly values of G for a bare dry soil in midsummer could be in excess of 300 Wm^{-2} and much lower, in the range of $-20, 20 \text{ Wm}^{-2}$, for moistured soils [35,36]. Low values of G during the two failures indicate moistured soil, as also evidenced by the peaks of soil moisture features. Overall, one can see the value of using at least two of these features to accurately identify the two distinct types of displacement, where precipitation, net radiation and soil moisture have clearer clusters.

Similar observations can be taken from Figure 9, which shows the clustering results as each of the selected features vs. daily differential displacement. The major movement occurred mainly, but not necessarily, during high PRECIP (first subfigure, both rows). The failures correspond to extreme values of soil moisture TDT1VWC, TDT2VWC (third, fourth figure, bottom), low positive and negative values of RN (second, top, sub-figure) and low values of G (third, fourth, top subfigures).

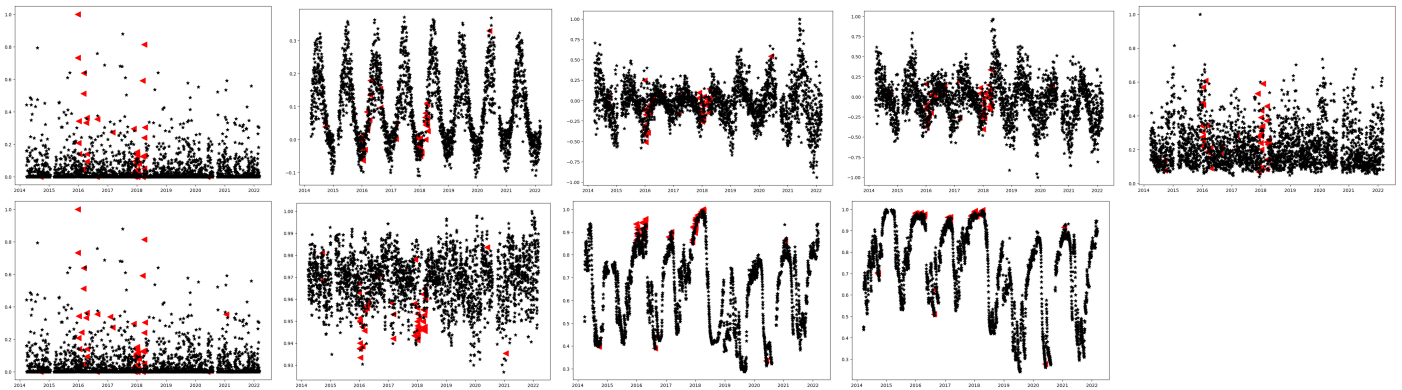


Figure 8. Clustering of data points to explain daily differential displacement, according to features (unitless, normalised per absolute maximum) selected by VIF (**top**) and RF (**bottom**): selected features vs. time, from left to right PRECIP, RN, G1, G2 and WS (**top**); PRECIP, PA, TDT1VWC, and TDT2VWC (**bottom**). Red indicates clustered data points related to major explosive movements while black indicates points related to intermittent movements.

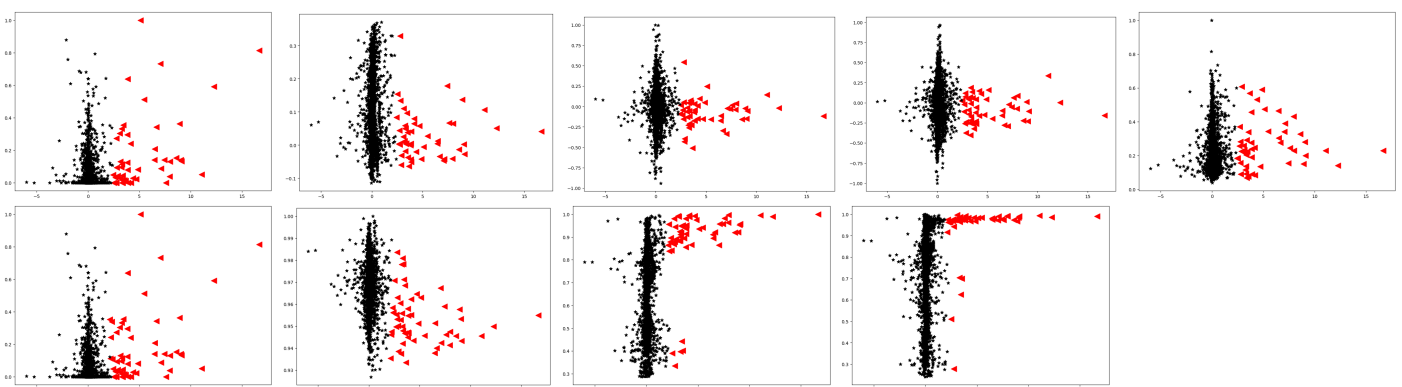


Figure 9. Clustering of data points to explain daily differential displacement, according to features (unitless, normalised per absolute maximum) selected by VIF (**top**) and RF (**bottom**): selected features vs. relative displacement, from left to right. Top: PRECIP, RN, G1, G2 and WS; Bottom: PRECIP, PA, TDT1VWC, and TDT2VWC. Red indicates clustered data points related to major explosive movements while black indicates points related to intermittent movements.

While the above conclusions are expected, it can be seen from Figures 8 and 9 that none of these features alone can be used as a good indicator of a failure. Indeed, while PRECIP is generally high during failure, extremely low values of PRECIP are also linked to the explosive failure, and high PRECIP often did not lead to a failure. Similarly, very low values of RN, G1 and G2, or high moisture TDT1VWC and TDT2VWC did not necessarily occur only during the failures. This leads to the conclusion that joint consideration of the selected features is needed to provide good landslide prediction.

6. Discussion of Key Findings

This study bridges the gap that exists in the current literature, between physical finite analysis models that consider many influencing factors for predicting landslide displacement and machine learning models that consider a small subset of influencing factors. Through correlation analysis and embedded feature selection, our study shows that, among 18 sensor recordings of a range of meteorological and ground parameters, the following sensor recordings, as summarised in Table 3, have the strongest influence on prediction of relative displacement: precipitation, soil heat flux, atmospheric pressure, and soil moisture. Note that the literature mostly tends to consider precipitation measurements and ground water level [18–20].

Furthermore, for completeness, we also consider feature extraction for dimensionality reduction, although the features extracted are in the transform domain and not physically interpretable. In order to predict displacement, we leverage ensemble regression methods, RF and XGBoost, which have been discussed in Section 1.1, as robust for limited training feature data (5–8 years) as well as Lasso regression, which is a relatively less complex model with fewer parameters and shorter execution time (as shown in first rows of Table 5).

As shown in Table 4, RF and XGBoost have similar relative prediction performance in general, but RF, as Lasso, is more robust for the smaller training set compared to XGBoost. Overall, RF has the best average performance and therefore we use it to compare the effect of physical feature selection vs. feature extraction in the LDA transformed domain, as well as with VIF-selected features that are independent of displacement. Key findings are that LDA feature extraction captures marginally better relative displacement with a smaller training set than embedded feature selection. As shown in Figure 5, cumulative prediction over 1, 5, 10, 15 and 30 days for the last intermittent failure and an explosive failure show that, whilst performance metrics in Table 5 indicate otherwise due to averaging over the five displacement regions, the reconstruction plots of predicted displacement are most accurate with XGBoost regression with inputs comprising precipitation, atmospheric pressure, wind direction and soil moisture. Generalisation to different types of failure is shown by the 50/50% train/test set split that predicts the second unseen major failure of 2018 in addition to the last intermittent failure. Our study provided quantitative prediction results (RMSE = 16.082 mm, MAE = 11.163 mm and $R^2 = 0.994$ for 5 days accumulation time window) comparable to other studies, using less computationally expensive models compared to deep learning models, as reviewed in Section 1.1, and small predictor sets (5 features for XGBoost) for up to 2.2 unseen years of movement. The final prediction was able to accurately capture the time at which the major event occurred and the magnitude of the total displacement increment that occurred in the duration of the particular major event.

Whilst the above methodology introduced a rigorous approach for embedded feature selection with supervised machine learning for predicting relative and cumulative displacement, to solve the problem of grouping the selected features into different stages of failure, an unsupervised hierarchical clustering approach is proposed in Section 5, where it is concluded that joint consideration of four to five selected features (PRECIP, RN, G1, G2, and WS with VIF) and (PRECIP, PA, TDT1VWC, TDT2VWC with RF) led to a better understanding of the underlying mechanisms related to the investigated instability.

The Hollin Hill failure is a moisture-induced and generally slow-moving landslide with intermediate periods of fast movements. The daily rate of normal movements in

this site is within the range [6.45×10^{-6} mmday⁻¹, 3.55 mmday⁻¹], while in periods of major events, movements can accumulate up to 250 mm per event, reaching rates up to 16.68 mmday⁻¹. The approach adopted in this study forms a framework according to which, relative and cumulative movements are predicted through the utilisation of a multiparameter dataset of long-term recordings related to distinct movement patterns. Since it is focused on daily relative displacement, this methodology is more applicable to cases of landslides where failure follows a behaviour dominated by periods of fast and slow movements where those stages can be distinguished. All steps followed across the process, such as dimensionality reduction, feature selection and the identification of subgroups within the recordings in an unsupervised manner but also the prediction of cumulative movements via regression, are generic and suitable for any dataset and type of sensors used, and so can be utilised in landslide early warning systems. It is worth mentioning here that the relative importance of features will be dependent on the specific type of landslide. For example, in the case of slopes with significant vegetation height (high-rise trees), the feature “wind speed” could play a more decisive role compared to our case, while gravitational forces that come from dense vegetation (densely located trees for example) can play a destabilising role in the failure process. In other cases, thawing permafrost triggers the landslide and so features related to temperature should play the most decisive role since the warming effect associated with climate change leads to melting the weakened and highly saturated frozen soil, thus leading to generalised instabilities.

7. Conclusions

Recent years have seen a growth in machine learning approaches to predict landslides or displacement in general. These require an appropriate choice of features that capture the influencing factors that have the most importance for learning displacement.

We propose a three-fold methodology whereby a statistical approach based on Variational Inflation Factor (VIF) is first used to remove multicollinearity among 18 possible influencing factors that are being monitored on the Hollin Hill Landslide Observatory over a period of 8 years. However VIF does not consider importance of the selected features in relation to displacement. Thus, the second proposed approach is to use supervised feature extraction, with two-component LDA and embedded feature selection tied to three regression approaches, namely Lasso, Random Forest (RF) and XGBoost. RF feature selection that identified precipitation, atmospheric pressure and soil moisture as the most important features, has best overall daily differential displacement prediction performance even with a smaller training set. However, XGBoost feature selection, which selected precipitation, atmospheric pressure, wind direction and soil moisture, has the best overall performance for cumulative displacement prediction.

We also show that standard performance metrics such as RMSE do not always capture the ability of a regressor to accurately reconstruct the explosive and intermittent stages of failure, unlike the actual plot with point to point reconstruction. Finally, in order to identify, in an unsupervised manner, what the key distinguishable stages of displacement are in relation to daily differential displacement, we propose agglomerative clustering with dendrogram visualisation. These confirm, through clusters of selected features from VIF and RF against time and daily differential displacement, that no one feature is sufficient, but rather joint consideration of selected features is needed to provide good landslide prediction.

Author Contributions: Conceptualisation, A.P., L.S. and V.S.; methodology, A.P., V.S. and L.S.; software, A.P.; validation, A.P.; formal analysis, A.P., V.S. and L.S.; writing—original draft preparation, A.P.; writing—review and editing, A.P., L.S. and V.S.; visualisation, A.P.; supervision, L.S. and V.S.; funding, L.S. and V.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by EPSRC New Horizons research programme EP/X01777X/1.

Data Availability Statement: The datasets used by this paper are not publicly available and belong to British Geological Survey (BGS) and to the COSMOS-UK project team.

Acknowledgments: We would like to thank BGS for providing the Hollin Hill Observatory data measurements as part of EP/X01777X/1.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CH	Calinski–Harabasz
GWL	Ground Water Level
HHLO	Hollin Hill Landslide Observatory
Lasso	Least Absolute Shrinkage and Selection Operator
LDA	Linear Discriminant analysis
LULC	Land Use Land Cover
MAE	Mean Absolute Error
NDVI	Normalised Difference Vegetation Index
PCA	Principal Component Analysis
RF	Random Forest
RMSE	Root Mean Squared Error
SVM	Support Vector Machine
TRI	Terrain Ruggedness Index
TWI	Topographic Wetness Index
VIF	Variance Inflation Factor
XGBoost	Extreme Gradient Boosting

References

- Whiteley, J.S.; Chambers, J.E.; Uhlemann, S.; Boyd, J.; Cimpoiasu, M.O.; Holmes, J.L.; Inauen, C.M.; Watlet, A.; Hawley-Sibbett, L.R.; Sujitapan, C.; et al. Landslide monitoring using seismic refraction tomography—The importance of incorporating topographic variations. *Eng. Geol.* **2020**, *268*, 105525. [[CrossRef](#)]
- Whiteley, J.S.; Chambers, J.E.; Wilkinson, P.B.; Kendall, J.M. Geophysical monitoring of moisture-induced landslides: A review. *Rev. Geoph.* **2019**, *57*, 106–145. [[CrossRef](#)]
- Sengani, F.; Mulenga, F. Application of Limit Equilibrium Analysis and Numerical Modeling in a Case of Slope Instability. *Sustainability* **2020**, *12*, 8870. [[CrossRef](#)]
- Griffiths, D.V.; Marquez, R.M. Three-dimensional slope stability analysis by elasto-plastic finite elements. *Geotechnique* **2007**, *57*, 537–546. [[CrossRef](#)]
- Sengani, F.; Mulenga, F. A review on the application of particle finite element methods (PFEM) to cases of landslides. *Int. J. Geotech. Eng.* **2022**, *16*, 367–381. [[CrossRef](#)]
- Huber, M.; Scholtes, L.; Lave, J. Stability and failure modes of slopes with anisotropic strength: Insights from discrete element models. *Geomorphology* **2024**, *444*, 108946. [[CrossRef](#)]
- Karthik, A.V.R.; Manideep, R.; Chavda, J.T. Sensitivity analysis of slope stability using finite element method. *Innov. Infrastruct. Solut.* **2022**, *7*, 184. [[CrossRef](#)]
- Guerrero-Rodriguez, B.; Salvador-Meneses, J.; Garcia-Rodriguez, J.; Mejia-Escobar, C. Improving Landslides Prediction: Meteorological Data Preprocessing Based on Supervised and Unsupervised Learning. *Cybern. Syst.* **2024**, *55*, 1332–1356. [[CrossRef](#)]
- Moraes, M.V.D.; Pampuch, L.A.; Bortolozzo, C.A.; Mendes, T.S.G.; Andrade, M.R.M.D.; Metodiev, D.; Pryer, T. Thresholds of Instability: Precipitation, Landslides, and Early Warning Systems in Brazil. *Int. J. Geosci.* **2023**, *14*, 895–912 [[CrossRef](#)]
- Le Breton, M.; Bontemps, N.; Guillemot, A.; Baillet, L.; Larose, E. Landslide monitoring using seismic ambient noise correlation: Challenges and applications. *Earth-Sci. Rev.* **2021**, *216*, 103518. [[CrossRef](#)]
- Abdo, H.G.; Richi, S.M. Application of machine learning in the assessment of landslide susceptibility: A case study of mountainous eastern Mediterranean region, Syria. *J. King Saud Univ.-Sci.* **2024**, *36*, 103174. [[CrossRef](#)]
- Stanley, T.A.; Kirschbaum, D.B.; Sobieszczyk, S.; Jasinski, M.F.; Borak, J.S.; Slaughter, S.L. Building a landslide hazard indicator with machine learning and land surface models. *Environ. Model. Softw.* **2020**, *129*, 104692. [[CrossRef](#)]
- Tien Bui, D.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2016**, *13*, 361–378. [[CrossRef](#)]
- Martínez-Álvarez, F.; Reyes, J.; Morales-Esteban, A.; Rubio-Escudero, C. Determining the best set of seismicity indicators to predict earthquakes. Two case studies: Chile and the Iberian Peninsula. *Knowl.-Based Syst.* **2013**, *50*, 198–210. [[CrossRef](#)]
- Krkač, M.; Špoljarić, D.; Bernat, S.; Arbanas, S.M. Method for prediction of landslide movements based on random forests. *Landslides* **2017**, *14*, 947–960. [[CrossRef](#)]

16. Chen, C.; Fan, L. Selection of contributing factors for predicting landslide susceptibility using machine learning and deep learning models. *Stoch. Environ. Res. Risk Assess.* **2023**, 1–26. [[CrossRef](#)]
17. Bravo-López, E.; Fernández Del Castillo, T.; Sellers, C.; Delgado-García, J. Analysis of conditioning factors in cuenca, ecuador, for landslide susceptibility maps generation employing machine learning methods. *Land* **2023**, *12*, 1135. [[CrossRef](#)]
18. Nava, L.; Carraro, E.; Reyes-Carmona, C.; Puliero, S.; Bhuyan, K.; Rosi, A.; Monserrat, O.; Floris, M.; Meena, S.R.; Galve, J.P.; et al. Landslide displacement forecasting using deep learning and monitoring data across selected sites. *Landslides* **2023**, *20*, 2111–2129. [[CrossRef](#)]
19. Yu, C.; Huo, J.; Li, C.; Zhang, Y. Landslide Displacement Prediction Based on a Two-Stage Combined Deep Learning Model under Small Sample Condition. *Remote Sens.* **2022**, *14*, 3732. [[CrossRef](#)]
20. Wang, Y.; Tang, H.; Huang, J.; Wen, T.; Ma, J.; Zhang, J. A comparative study of different machine learning methods for reservoir landslide displacement prediction. *Eng. Geol.* **2022**, *298*, 106544. [[CrossRef](#)]
21. Xu, J.; Jiang, Y.; Yang, C. Landslide displacement prediction during the sliding process using XGBoost, SVR and RNNs. *Appl. Sci.* **2022**, *12*, 6056. [[CrossRef](#)]
22. Boyd, J. Hydrogeophysical Characterisation for Improved Early Warning of Landslides. Doctoral Thesis, Lancaster University, Lancaster, UK, 2024.
23. Whiteley, J. Geophysical Indicators of Slope Stability: Towards Improved Early Warning of Moisture-Induced Landslide Hazards. Doctoral Thesis, University of Bristol, Bristol, UK, 2022.
24. COSMOS-UK Project Team. *Cosmos-UK User Guide*; Version 3.08; UK Centre for Ecology & Hydrology: Wallingford, UK, 2024.
25. Sujitapan, C. Insights into Moisture-Driven Landslides Using Electrical and Seismic Methods: Case Studies from Hollin Hill, UK and Thungsong, Thailand. Doctoral Dissertation, University of Bristol, Bristol, UK, 2021.
26. Liao, D.; Valliant, R. Variance inflation factors in the analysis of complex survey data. *Surv. Methodol.* **2012**, *38*, 53–62.
27. Xanthopoulos, P.; Pardalos, P.M.; Trafalis, T.B. *Robust Data Mining*; Springer Science & Business Media: Berlin, Germany, 2012.
28. Parasyris, A.; Stankovic, L.; Stankovic, V. Dimensionality reduction for visualisation of hydrogeophysical and meteorological recordings on a landslide zone. In Proceedings of the IGARSS Conference, Athens, Greece, 7–12 July 2024.
29. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. Stat. Methodol.* **1996**, *58*, 267–288. [[CrossRef](#)]
30. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
31. James, G.; Witten, D.; Hastie, T.; Tibshirani, B.; Taylor, G. *An Introduction to Statistical Learning with Applications in Python*; Springer: Berlin/Heidelberg, Germany, 2023.
32. Liu, Y.; Li, Z.; Xiong, H.; Gao, X.; Wu, J. Understanding of Internal Clustering Validation Measures. In Proceedings of the 2010 IEEE International Conference on Data Mining, Sydney, Australia, 13–17 December 2010.
33. Mishra, S.; Saha, S.; Mondal, S. A multiobjective optimization based entity matching technique for bibliographic databases. *Expert Syst. Appl.* **2016**, *65*, 100–115. [[CrossRef](#)]
34. Sauer, T.J.; Peng, X. Soil Temperature and Heat Flux. *Agroclimatol. Link. Agric. Clim.* **2020**, *60*, 73–93.
35. Sauer, T.J.; Peng, X. Soil heat flux. *Micrometeorol. Agric. Syst.* **2005**, *47*, 131–154.
36. Fuchs, M.; Hadas, A. The heat flux density in a non-homogeneous bare loessial soil. *Bound.-Layer Meteorol.* **1972**, *3*, 191–200. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.