https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

Climate change increases landslide susceptibility in Aotearoa New Zealand: Development and application of a national-scale model using machine learning.

Oliver Wigmore[1,2]

[1]Department of Conservation, Wānaka, New Zealand
[2]Antarctic Research Centre, Victoria University of Wellington, Wellington, New Zealand

Correspondence to: Oliver Wigmore (oliver.wigmore@vuw.ac.nz)

ORCID: 0000-0002-6813-3884

Abstract.

Rainfall-induced landslides (RILs) pose a major hazard to infrastructure, ecosystems, and communities across Aotearoa New Zealand, with events such as Cyclone Gabrielle underscoring the potential scale of their impacts. In this study, we develop a relatively high-resolution national-scale RIL susceptibility model that includes both conditioning and triggering variables and use it to assess the impacts of climate change on RIL susceptibility. The model utilises machine learning (ML) (gradient boosted decision trees) to predict RIL susceptibility in response to extreme rainfall events under current and future climate scenarios at 25m spatial resolution. We use a training dataset of observed landslides triggered by Cyclone Gabrielle in the Hawke's Bay and Gisborne/Tairāwhiti regions. Predictor variables include topographic, geologic, and environmental factors, with rainfall intensity serving as the primary trigger. Model performance is evaluated using Shapley additive explanations (SHAP) analysis, alongside standard error metrics, achieving a receiver operating characteristic area under the curve (ROC-AUC) of 0.94. We then apply the model nationally to estimate RIL susceptibility under six current and 24 future storm scenarios based on NIWA's high-intensity rainfall design system (HIRDS) datasets and modelled temperature changes under different shared socioeconomic pathways (SSPs). Results show a substantial increase in RIL susceptibility under warmer climate futures, with susceptibility increasing disproportionately to rainfall increase. Forest cover is found to play an important role in mitigating susceptibility. This work presents a robust framework for national-scale RIL susceptibility assessment under specific storm scenarios and provides a national-scale dataset suitable to support climate-resilient land use planning and nature-based mitigation strategies.

1.1 Introduction

Rainfall-induced landslides (RILs) are a significant natural hazard in Aotearoa New Zealand. For example, Cyclone Gabrielle in February 2023 is estimated to have caused (NZD)$14 billion in damages, with around (NZD)$1.5 billion of that attributed to landslides (McMillan et al., 2023; Wilson et al., 2023). The cost of RIL damage is primarily due to lost assets and infrastructure, for example, productive farmland, crops, roads, and buildings. However, RILs can also impact downstream areas through the transport and subsequent deposition of large sediment loads, which increases turbidity and can bury critical river habitat, impacting native species. Following Cyclone Gabrielle (McMillan et al., 2023) estimated that around 300 million tonnes of sediment were mobilised by RILs, with much of that ending up in streams and rivers. For example, in the Esk Valley, they estimated 5.7 million tonnes of soil were eroded; around half of that was delivered to waterways, with 1.5 million tonnes of sediment deposited on the floodplain to an average thickness of 80cm (McMillan et al., 2023). RIL susceptibility is

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

influenced by a variety of topographic, geologic, and environmental factors, such as slope, rock type, soil depth and porosity, antecedent soil moisture, land cover (vegetation), and land use (Ehsan et al., 2025). Heavy rainfall then acts as a trigger, saturating the soil, which increases its mass and reduces shear strength, which can lead to slope failure and landslides. Thus, variability in the intensity and direction of rainfall (during a given storm), in concert with local topography (e.g. aspect), impacts where and when RILs occur (Ehsan et al., 2025). As a result of climate change, some regions are expected to experience increases in the frequency, quantity, and intensity of precipitation received during rainfall events. This effectively compresses the recurrence interval of extreme rainfall (storm) events; consequently, in these regions, the frequency and intensity of RILs is likely to increase in the future (Crozier, 2010a; Gariano & Guzzetti, 2016). Accurate susceptibility maps of RILs under current and future climate conditions are important tools for reducing RIL exposure to people and infrastructure, for both existing and future development. Additionally, these products can be used to better inform land use decision-making and facilitate the implementation of nature-based solutions, for example, retiring RIL-prone hill country pasture for more resilient forest cover (Griffiths et al., 2020; Phillips et al., 2021; Spiekermann et al., 2022). Thus, it is critical to improve our understanding at a national scale of what areas are most susceptible to RIL currently and how this may change under future storm and climate scenarios.

To date, most of New Zealand's RIL susceptibility mapping falls into two categories: broadscale erosion and/or landslide susceptibility mapping relying on a mixture of heuristic and parametric statistical models, and physically based RIL modelling. Broadscale maps delineating highly erodible land (HEL) (which includes landslide susceptibility) for New Zealand were first developed in 2006 (Dymond et al., 2006), with the most recent update in 2023 (Dymond & Shepherd, 2023). These HEL susceptibility maps were generated by applying thresholds to a series of data layers (slope, forest cover, geology) to identify areas of high landslide (and earthflow) susceptibility at the hillslope scale. Expert knowledge is used to determine these threshold values, and there is no trigger data (rainfall) included as part of the analysis. Critically, areas with forest cover are delineated as low susceptibility regardless of slope and geology, when in reality, forest cover mitigates but does not eliminate the likelihood of RIL, especially during large storm events (Fuller et al., 2016; Griffiths et al., 2020; Phillips et al., 2021). Consequently, these data are suitable for mapping areas of increased landslide susceptibility but are unable to generate predictions in response to specific storm or climate scenarios, and are of limited use in forested areas. Ex post facto analysis of these datasets in conjunction with observed landslides attributed to specific storm events has reported accuracy of 58% (Dymond et al., 2006) to 75% (Dymond & Shepherd, 2023) at the hillslope scale, but as low as 26% on a per-pixel basis (Dymond et al., 2006). However, commission errors have not been reported for these datasets. The application of statistical models has mostly focused on correlation and regression-type solutions at the catchment to the regional scale (De Rose, 2013; Massey et al., 2025; Smith et al., 2021; Spiekermann et al., 2022). Rosser et al. (2021) reported on the development of a national-scale logistic regression model for deployment as a forecasting tool, which combines conditioning factors and event rainfall. However, forecasts are produced in response to specific rainfall events and are not provided as a national dataset. Statistical models of earthquake-induced landslide susceptibility have also been produced (Bloom et al., 2023). Models based primarily on accounting for the water content of soils as a triggering factor have shown success over smaller study areas (Crozier, 1999; Hufschmidt & Crozier, 2008). Conversely, more complex physically based RIL modelling has been completed for some areas of New Zealand (Claessens et al., 2007; De Sy et al., 2013; Howard et al., 2025; Wilkinson et al., 2002). These models consider underlying conditioning and trigger factors (rainfall and earthquakes) and often integrate debris flow or runout models to predict both the initiation point and path of landslide debris (Howard et al., 2025). However, they are computationally and labour-intensive and typically rely on high-resolution digital elevation models (e.g., LiDAR), which are not yet available with complete national coverage. Because of these limitations, they have not been deployed at a national scale. Their use has

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

primarily focused on heavily populated urban regions, or areas with critical assets and infrastructure (e.g., road networks), and where there is sufficient data and investment to support their development.

Internationally, there have been recent advances in RIL susceptibility modelling that leverage nonparametric machine learning (ML) algorithms at a broad scale. (Chen et al., 2024; Ehsan et al., 2025; Mondini et al., 2023; Ngo et al., 2021; Semnani et al., 2025; Yin et al., 2023). ML algorithms can exploit large observational datasets of RIL events, alongside myriad underlying potential explanatory or conditioning variables (e.g., geology, topography, land cover), from which models of RIL susceptibility can be developed. In particular, tree-based ensemble methods (e.g. random forests and numerous boosted tree algorithms) have found widespread use and success (Ehsan et al., 2025). Tree-based ML models are particularly suited to RIL modelling because they are capable of capturing non-linear interactions among predictor variables, and have been shown to have higher predictive capability than logistic regression approaches, which assume linear relationships between predictors and log-odds (Al-Najjar et al., 2021; Chen et al., 2017; Couronné et al., 2018). Recently, ML models have been developed that move beyond static RIL susceptibility by combining underlying conditioning factors (e.g. topography, geology and land cover) and event-specific rainfall (Bordoni et al., 2021; Mondini et al., 2023; Segoni et al., 2018; Shu et al., 2024). They can therefore be used for dynamic RIL susceptibility mapping and potentially forecasting (Bordoni et al., 2021; Mondini et al., 2023), as well as assessing the impact of climate change on RIL susceptibility (Han & Semnani, 2025; Semnani et al., 2025; Yin et al., 2023). ML approaches are data-driven, with the most useful predictors identified through a combination of feature importance metrics and model cross-validation to facilitate feature selection (Ehsan et al., 2025; Han & Semnani, 2025; Semnani et al., 2025). Leveraging ML to map RIL susceptibility requires less expert knowledge of the underlying processes that drive RIL susceptibility, is cost-effective to scale nationally, and can provide high predictive accuracy that accounts for trigger events (i.e. storms) (Ehsan et al., 2025). However, large observational datasets are required for model training and testing. And, considerable thought must be put into variable and model selection, and subsampling strategies to deal with the often significant class imbalance between landslide and non-landslide observational instances (Achu et al., 2023; Gu et al., 2024; Han & Semnani, 2025). Post hoc analysis of model predictions can be performed using feature importance and tools such as local interpretable model-agnostic explanations (LIME) (Ribeiro et al., 2016) and Shapley additive explanations (SHAP) (Lundberg & Lee, 2017) to investigate how individual variables influence model predictions (Achu et al., 2023; Wen et al., 2025). These tools can provide confidence in model performance (i.e. how predictions are made) and consequently elucidate further insight into RIL processes and potentially causation (Heskes et al., 2020). Furthermore, once ML models are created, it is relatively straightforward to generate susceptibility maps to investigate the potential impact of changes in land cover and land use (Griffiths et al., 2020), and future storm/climate scenarios (Han & Semnani, 2025; Semnani et al., 2025). Despite their success internationally, ML approaches for RIL susceptibility mapping in New Zealand have, to our knowledge, been limited to smaller study areas. For example, Griffiths et al. (2020) used gradient boosted decision trees to model landslide susceptibility in the Tasman region, with the model output used to assess the importance of native forest cover in reducing landslide susceptibility. They reported excellent predictive capability when training and validating the model with landslides observed from Cyclone Gita; however, this model has not been expanded nationally. More recently, Smith et al. (2021) developed both logistic regression (parametric) and random forest (nonparametric) models to predict landslide susceptibility for a small number of study catchments in New Zealand. They found that random forest outperformed logistic regression when validated in the same study area, while logistic regression mostly outperformed random forest training and testing on different regions. These models have not been deployed nationally, and the output is not publicly available.

In this study, we first develop a ML model (Light-Gradient Boosting Machine (LightGBM)) capable of predicting RIL susceptibility consequent to rainfall events. To do this, we leverage a training and testing dataset of observed landslides over the Hawke's Bay and Gisborne/Tairāwhiti regions that were triggered by Cyclone Gabrielle in mid-February 2023. Using the

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

trained model, we then generate predictions at a national scale, for three different extreme rainfall event recurrence intervals (10, 50, and 100 years) and for two different future climate Shared Socioeconomic Pathways (SSPs) (2-4.5, 3-7.0) (O'Neill et al., 2017; Riahi et al., 2017), at two time periods (2041-2060 and 2080-2099). We then use these outputs to assess changes in RIL susceptibility due to different future climate pathways. The workflow developed here is designed to be transferable to other temperate and high-intensity rainfall regions where comparable datasets are available.

2 Methods

2.1 Study Domain

Cyclone Gabrielle is estimated to have been a 1 in 100-year average recurrence interval (ARI) event; however, rainfall delivery across the country was highly variable, with some areas exceeding a 250-year ARI (Massey et al., 2025). The Hawke's Bay and Gisborne/Tairāwhiti regions (Fig. 1) on the east coast of the North Island experienced the most intense rainfall and extensive RIL damage (Massey et al., 2025). However, even here, rainfall varied considerably depending on location. Consequently, the observed impacts from this event represent a range of different rainfall intensities across a large geographic area. Furthermore, within these regions exists a variety of topography, land cover, soils, and geology that are broadly reflective of the North Island. While the land cover and topography of the South Island are similar to those of the North Island, geology is markedly different. The mountainous regions of the South Island are dominated by harder metamorphic rock types (e.g. greywacke) compared to the erosion-prone soft sedimentary rocks (mudstone and sandstone) and volcanics (e.g. basalt) of the North Island. Consequently, their RIL susceptibility and the relative importance of the underlying conditioning factors may differ markedly. Despite this potential limitation, the impact of Cyclone Gabrielle on Hawke's Bay and Gisborne/Tairāwhiti provides an ideal scenario with which to develop a machine learning based model of RIL susceptibility that can be applied at a national scale.

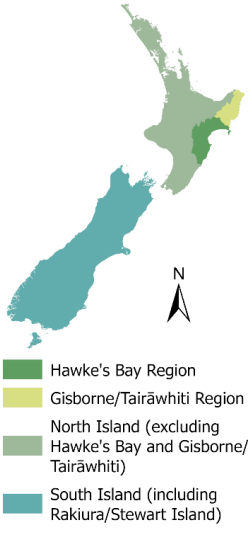

Figure 1: Study area domains.

2.2 RIL Mapping

To build a master dataset of RIL resulting from Cyclone Gabrielle in Hawke's Bay and Gisborne, we acquired Sentinel 2 imagery from just before (5, 7, 10 February 2023 pre-cyclone) and immediately after the cyclone (20, 22 February 2023 post-

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

cyclone) (scene IDs provided below), which impacted New Zealand from 11 February 2023 to 17 February 2023. Due to patchy cloud cover in the pre-cyclone and post-cyclone imagery, we built a cloud-free mosaic by selecting the pixel with the closest cloud-free date to the cyclone for each group. We then calculated the normalised difference vegetation index (NDVI) for the pre-cyclone and post-cyclone imagery. To identify RIL pixels, we calculated the change in NDVI (ΔNDVI) and empirically determined a suitable threshold value to identify transitions to bare ground between the two images. Areas with a ΔNDVI value exceeding the threshold were determined to be impacted land, while ΔNDVI values below this were not impacted. We then separated landslides from flood damage/silt deposition by applying a logic filter where only land with a slope >10° was considered a RIL, where slope was derived from the NZSOS DEM (Columbus et al., 2011), which was interpolated to the same 10m grid as the Sentinel 2 imagery. Technically, this method only identifies transitions to bare ground, and there are mechanisms other than RIL that may be responsible for this, e.g. forest clearance, crop harvesting, drought, mining, etc. However, due to the tight time bracketing between the pre-cyclone and post-cyclone imagery (~2weeks, with ~3-5 days of storm activity), it is unlikely that processes other than RIL are responsible for extensive areas of bare ground transition on land >10° during this short period. It is important to note that this method of landslide mapping identifies and includes both the initiation point and run-out zone (where the land exceeds a 10° slope).

To refine the RIL training data set, we downloaded a secondary Cyclone Gabrielle landslide inventory from Dragonfly Data Science (Dragonfly Data Science, 2023). They completed a landslide mapping exercise (post-Cyclone Gabrielle) using a similar methodology for the entire North Island. However, their work utilised a longer temporal window for both the pre-cyclone and post-cyclone imagery; consequently, it may include spurious results (bare ground transitions caused by non-RIL processes). We merged these data with our own, including only RIL pixels identified in both datasets as positive results. RIL pixels identified in only one of the datasets were classed as possible RIL. These areas were excluded from the training and testing datasets for both RIL and no-RIL classes. This process minimises potential errors in the identification of no-RIL:RIL in the training and test datasets. Based on this classification, we calculated that ~1% of the Hawke's Bay and Gisborne/Tairāwhiti regions experienced landslide damage, which aligns with post-cyclone assessments by Manaaki Whenua/Landcare Research (McMillan et al., 2023). Because we used 10m satellite imagery, it is probable that RILs that are significantly smaller than this were missed in the mapping exercise and consequently could be included in the no-RIL dataset. However, due to their low prevalence (relative to true no-RIL areas), the impacts of this are likely to be minimal.

2.3 Predictor Variables

Thirty potential predictor variables were collected and/or derived from various national and international datasets. The twelve predictor variables that were used in the final model are shown in Table 1, along with their source. Datasets were resampled to the same 25m pixel grid using bilinear interpolation. Vector datasets were rasterised to the 25m grid. Small data gaps were present in some of the datasets; these were filled using bilinear interpolation for floating-point variables or majority filtering for integer values. Not all datasets extended to the same coastal boundary; to remedy this, we interpolated (and filled holes) beyond the coastline before clipping all data to the same boundary.

Cyclone Gabrielle rainfall data were provided by the National Institute of Water and Atmospheric Research (NIWA) from the augmented virtual climate station network (VCSN) at ~500m. We used the 72-hr period spanning 12-14 February 2023 (starting 09:00 local time), corresponding to the period of active cyclone-related rainfall over New Zealand. Statistically resampling (i.e. ignoring local topography) to a 25m grid introduces uncertainty at local scales but still preserves the broad patterns of the dataset. A potentially more accurate physically based downscaling of the rainfall products is beyond the scope of this work. Storm rainfall as a proportion of annual rainfall was calculated against a 25m resampling of the mean annual rainfall from the New Zealand environmental data stack V1.1 (NZEnvDS), which is provided at 100m (McCarthy et al., 2021).

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

The topographic variables were all derived from the Forest and Buildings removed Copernicus 30m digital elevation model (FABDEM) version 1.2 (Hawker et al., 2022; Neal & Hawker, 2023). The FABDEM was resampled to the 25m grid using cubic convolution; child variables (e.g. slope, curvature, etc) were then calculated. Canopy height is a global 30m product for 2019 derived from a combination of multispectral Sentinel 2 data and spaceborne lidar from the Global Ecosystem Dynamics Investigation (GEDI) mission (Potapov et al., 2021), and has been shown to provide suitable accuracy for broad-scale canopy height assessment in New Zealand (Ng et al., 2025). Soil variables are taken from the New Zealand land resource inventory (NZLRI) soils (rooting depth, depth to zone of slow permeability) (Newsome et al., 2008), and the New Zealand fundamental soils layer (NZFSL) (particle size, drainage) (Barringer, 2018). It is important to note that the NZLRI layers do not provide data for urban areas; consequently, these areas were set as null values for model training and prediction. Depth to the hydrogeologic basement is a national dataset created by GNS Science (GNS Science, 2019; Westerhoff et al., 2019). Because not all national rock types are present within the training area (Hawke's Bay and Gisborne/Tairāwhiti), geology could not be used directly as a categorical predictor variable; instead, we use rock density as a proxy for rock strength. To do this, we created an updated version of Tenzer et al. (2011) national rock density map, based on the most recent NZGeol250 (Heron, 2023), which fills some data gaps present in the original dataset. We do not include aspect variables (e.g. northness) because their influence on RIL susceptibility is likely to depend on storm direction, which is unknown for future scenarios. Similarly, we do not include antecedent soil moisture, which reflects prior rainfall and therefore antecedent soil saturation, and has been shown to be an important control on RIL susceptibility (Crozier, 1999; Leonarduzzi et al., 2021). We omit this predictor because pre-storm rainfall and soil moisture cannot be specified for future storm scenarios.

Variable Name	Variable Description	Variable Source	Variable Group
Curvature	Combined profile and planimetric land curvature	Calculated from FABDEM (Neal & Hawker, 2023)	Topography
Cyc. Gab. Rain Prop. Ann.	Cyclone Gabrielle total rainfall (12-14 February) as a proportion of annual rainfall	Cyc. Gab. Total Rain (mm) divided by mean annual rainfall (McCarthy et al., 2021)	Water
Cyc. Gab. Total Rain (mm)	Cyclone Gabrielle total rainfall (12-14 February) (mm)	Cyclone Gabrielle rainfall data from NIWA VCSN	Water
Depth to Slow Perm. (m)	Depth to zone of slow permeability (m)	NZLRI (Newsome et al., 2008)	Sediments
Drainage (rank)	Soil drainage relative ranking	NZFSL (Barringer, 2018)	Sediments
Forest Height (m)	Height of forest canopy (m)	Potapov et al., (2021)	Vegetation
HGB Depth (m)	Depth to hydrogeological basement	GNS Science, (2019)	Geology
Particle Size (rank)	Soil particle size relative ranking	NZFSL (Barringer, 2018)	Sediments
Rock Density (Mg/m3)	Density of main rock type from NZGeol250	Main rock type from NZGeol250 (Heron, 2023) matched to rock density from Tenzer et al., (2011)	Geology
Rooting Depth (m)	Rooting depth (m)	NZLRI (Newsome et al., 2008)	Sediments
Slope (deg)	Slope angle in degrees	Calculated from FABDEM (Neal & Hawker, 2023)	Topography
Topo. Wetness Ind.	Topographic Wetness Index	Calculated from FABDEM (Neal & Hawker, 2023)	Water
Predictors tested but not used in the final model: Canopy Cover Percentage, Catchment Area, Catchment Slope, Drop (D8), Flow Accumulation, Forest/Not-forest, Gravel Content, Multiscale Terrain Position Index, Plan Curvature, Profile Curvature, Profile Curvature Variability (5x5 Standard Deviation), Ranked Rock Strength, Roughness, Slope Length, Slope Variability (5x5 Standard Deviation), Soil Induration, Terrain Ruggedness Index.			

Table 1: Final set of predictor variables used in the model and their source.

LightGBM is unaffected by multicollinearity; however, the inclusion of low-quality (redundant) variables can increase model complexity and the potential for overfitting. Thus, to help identify the most useful variables and reduce the total number of

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

variables, we calculated a correlation matrix for all potential predictors. This identified five main groups of data: geology, land cover, sediments, topography, and water. Variable selection aimed to retain at least one variable from each group in the final set of predictors and minimise the use of highly correlated variables. We used this correlation analysis in conjunction with SHAP analysis to identify the best predictor variables for use in the final model.

2.4 Model Development

2.4.1 Training, Validation, and Testing Data

We use the following definitions to refer to the data subsets described herein: 1) training data are a random subset of the initial data used for model training; 2) validation data are the out of fold data from a random K-fold (10-folds) cross validation of the training data, which is used to select predictor variables and tune hyperparameters; 3) testing data are a random subset of the initial data that are held out from training and validation and used solely to assess final model performance against unseen data.

To develop training and testing datasets from the landslide mapping dataset, we first polygonised the satellite-mapped landslides. This serves to aggregate individual pixels into connected landslide features. From this, we randomly selected a 20% subset of these landslide polygons for use as the holdout testing dataset. Polygonising the landslide dataset first is important, as it ensures that training and testing pixels are not potentially selected from the same landslide feature. This step minimises the likelihood (and hence impact) of spatial autocorrelation between the training and testing datasets, which can falsely inflate model accuracy metrics due to data leakage between the training and testing datasets (Kattenborn et al., 2022; Koldasbayeva et al., 2024).

The training dataset includes all pixels within the 80% landslide polygon subset, as well as a random selection of non-landslide pixels, which were selected as follows. After identifying the landslide training features, we then randomly selected pixels from the entire Hawke's Bay and Gisborne/Tairāwhiti regions where no landslides were identified. Pixels that were identified as possible landslides, i.e. mapped as a landslide by one but not both mapping products (2.2 above) were excluded. A 10m buffer was applied to the landslide and possible landslide polygons as an exclusion zone; this minimises the potential inclusion of mixed pixels along the edges of the mapped landslide features. Due to the severe class imbalance between no-RIL and RIL pixels (~99:1), no-RIL pixels were then undersampled to produce a 60:40, 55:45 and 50:50 no-RIL:RIL data ratio. Ultimately, a 50:50 split was selected as it provided the best model performance.

For the test dataset, we combined all pixels within the 20% hold-out subset of landslide polygons (above) with a random selection of no-RIL pixels that were not used in training, selected following the same criteria as above. Additionally, we applied a 72m (5x5 pixel neighbourhood) buffer to the no-RIL pixels used in training. This reduces potential spatial autocorrelation between the no-RIL training and testing instances. The test data set was compiled at the same ratio as observed in the event, i.e. 99:1 (no-RIL:RIL). The final training dataset includes 358,880 25m pixel observations (50:50 ratio), with the holdout test dataset comprising 3,898,278 25m pixel observations (99:1 ratio).

2.4.2 Model Architecture

The model was implemented in Python using LightGBM, a gradient-boosted decision tree algorithm. Hyperparameters were optimised using Optuna via the LightGBMTunerCV framework with ten-fold cross-validation on the training dataset. Optimisation minimised binary log loss. Up to 2000 boosting iterations were permitted, with early stopping after 50 rounds without improvement. The Optuna optimisation was allowed to run for up to 12 hours. Final tuned hyperparameters are reported in Appendix A. We calculated feature importance and SHAP values on the out-of-fold data and used this alongside

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

correlation analysis to identify the most useful predictor variables. The least useful variables were dropped, and the workflow was rerun to determine the final set of (twelve) variables (Table 1). Care was taken to minimise the use of highly correlated variables (to minimise model complexity), but retain variables from each correlation grouping (2.3 above). We then used the values from the Optuna hyperparameter tuning to retrain the model on the entire training dataset.

The model was trained on a 50:50 (majority undersampled) distribution of no-RIL:RIL, when in reality the positive case (RIL) has a much lower prevalence (approximately 1%, 99:1 noRIL:RIL). Thus, the probabilities output from the model may require calibration and/or scaling (e.g. prior probability adjustment) before they can be interpreted as real-world probabilities for quantitative exposure assessment (Pozzolo et al., 2015; Saerens et al., 2002). We assessed model calibration on a 50:50 distribution subsample of the test dataset and compared raw model output with both sigmoid and isotonic calibration to ensure stable prediction frequency across the probability range (Fig. B1). The raw model output was already well calibrated, and neither sigmoid nor isotonic calibration meaningfully improved model quality. To adjust the raw model probabilities (under the 50:50 training prevalence) to the observed event prevalence (99:1 no-RIL:RIL), we applied a prior-probability (scaling) adjustment following Saerens et al., (2002) using Eq. (1). Thus, the final probabilities reported here (and provided in associated datasets) represent event-conditional RIL probability for a specified storm scenario.

$$P_{corrected} = \frac{P_{model} \cdot \frac{P_{real}}{P_{train}}}{P_{model} \cdot \frac{P_{real}}{P_{train}} + (1 - P_{model}) \cdot \frac{1 - P_{real}}{1 - P_{train}}} \quad (1)$$

Where: P_{model}: raw model output probability from the classifier trained on a 50:50 dataset

P_{train} = 0.5: event prevalence in the training data

P_{real} = 0.01: observed event prevalence (Cyclone Gabrielle; 99:1 no-RIL:RIL)

$P_{corrected}$: prevalence-adjusted probability used for thresholding/exposure calculations

Finally, we reran the model on the holdout test data, calculated error metrics and completed SHAP analysis to assess real-world deployment performance. SHAP was used to interpret the fitted model's feature attributions (predictive explanations) rather than to infer causal effects. SHAP analysis, receiver operating characteristic (ROC) and precision-recall curves were calculated on the raw (unscaled) model output. The other error metrics (Brier score, F1, precision, recall, and accuracy) were calculated on the scaled probabilities. Test data were not seen in the model training phase, and through the polygonised randomisation process of RIL pixels, test data cannot be selected from the same landslide feature as the training data. Similarly, the buffering of non-RIL pixels minimises the effect of data leakage due to spatial autocorrelation. Finally, the test data represent the true class distribution ratio of non-RIL to RIL (99:1); thus, the accuracy metrics reported for this analysis are representative of absolute model accuracy for the Hawke's Bay and Gisborne/Tairāwhiti regions in relation to Cyclone Gabrielle.

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

2.5 Predicting RIL susceptibility due to future storms

To predict RIL susceptibility at the national scale, we used data from the NIWA high-intensity rainfall design system (HIRDS) (Carey-Smith et al., 2018). HIRDS is a national scale model output of rainfall depths (mm) for a range of different storm durations (10 minutes to 120 hours) and recurrence intervals (1.58-250 years), with 144 different scenarios output at 2000m spatial resolution. For this study, we used a 24-hour and 72-hour storm duration, with a recurrence interval of 10, 50, and 100 years (3 scenarios). We statistically resampled the HIRDS data to the 25m study grid using bilinear interpolation and calculated storm rainfall proportional to annual rainfall calculated from the NZEnvDSV1.1, which was also statistically resampled from 100m to 25m (McCarthy et al., 2021). We then reran the model, replacing the Cyclone Gabrielle precipitation variables (storm total precipitation and storm total precipitation as a proportion of annual precipitation) with each of the HIRDS-derived datasets. Model outputs were generated at the national scale, i.e. extending outside of the initial training and testing geographic domain (Hawke's Bay and Gisborne/Tairāwhiti). To provide confidence in this approach, we compared the predictor data distributions for Hawke's Bay and Gisborne/Tairāwhiti with the rest of the North Island and the South Island (including Rakiura/Stewart Island) (Appendix C).

HIRDS data can also be used to estimate rainfall under future climate scenarios by applying a percentage change factor per degree of warming (Table 2) (Carey-Smith et al., 2018). Where temperature change is calculated as a national scale mean for different climate scenarios (Table 3). In this case, we used two SSP (2-4.5, 3-7.0) scenarios for the time periods 2041-2060 and 2080-2099. Temperature change data comes from the most recent downscaled climate projections for New Zealand produced by NIWA (NIWA, 2025) from a subset (six) of global models (Bodeker et al., 2022; Ministry for the Environment, 2018). The same three 24-hour and 72-hour storm durations and recurrence intervals (10, 50, 100 years) were run for each SSP. Thus, a total of 6 current climate and 24 future climate RIL prediction layers were created, relating to each storm, climate, and recurrence scenario. These output datasets comprise the main data asset contribution of this work.

A statistical analysis of the potential increased RIL susceptibility due to a warming climate was then completed. First, we calculated the total land area in each New Zealand region with prevalence-adjusted RIL susceptibility exceeding 5% threshold ($\tau = 0.05$) for each storm and climate scenario, denoted A_{pred} (true positive (TP) + false positive (FP)). To estimate the total area likely to actually experience a RIL event for a given scenario, we applied a first-order correction based on the confusion matrix at the same threshold τ, converting the thresholded susceptibility area (TP +FP) into an expected impacted area (TP + false negative (FN)) using precision and recall estimated from the holdout test region. Because these error rates were derived for the training/test domain and may vary across regions and scenarios, we interpret the corrected values only as screening-level estimates suitable for national comparison.

$$A_{imp} \approx \frac{precision * A_{pred}}{recall} \quad (2)$$

Where: A_{pred} = *area above threshold (TP + FP)*

A_{imp} = estimated impacted area (TP + FN)

Precision and recall are evaluated on the holdout test set at threshold τ

We then calculated the total number of buildings and total length and area of the road network exposed to RIL susceptibility exceeding a 5% threshold, using the Land Information New Zealand (LINZ) NZ Building Outlines product (LINZ, 2025a),

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

and NZ Road Centrelines (LINZ, 2025b), respectively. The building outlines product includes all structures that are mapped through feature extraction of national aerial imagery, and thus includes higher value structures, e.g. houses, hospitals, schools, high rises, and low-cost structures, e.g. small sheds, garages, and greenhouses. These are mostly undifferentiated; therefore, building counts cannot be directly equated with financial exposure. To identify buildings that are potentially exposed to RIL, we simply counted all building footprints that fell within or intersected the RIL susceptibility areas exceeding 5% probability. To assess the potential impact of modelled RIL scenarios on the national road network, we first buffered the road centrelines product by the typical New Zealand road lane width (3.5m), and the number of lanes, i.e. road lane width (3.5m) multiplied by half the number of road lanes. We then intersected this layer with the areas of RIL susceptibility exceeding a 5% probability to identify the impacted road area; finally we intersected this layer with the road centrelines to calculate the impacted road length.

	ARI 10yr	ARI 50yr	ARI 100yr
24hr	**8.10**	**8.40**	**8.60**
24hr 2-4.5 2041-2060	*10.47*	*10.86*	*11.12*
24hr 2-4.5 2080-2099	*17.34*	*17.98*	*18.41*
24hr 3-7.0 2041-2060	*12.58*	*13.04*	*13.35*
24hr 3-7.0 2080-2099	*25.69*	*26.65*	*27.28*
72hr	**6.50**	**6.80**	**6.90**
72hr 2-4.5 2041-2060	*8.40*	*8.79*	*8.92*
72hr 2-4.5 2080-2099	*13.91*	*14.56*	*14.77*
72hr 3-7.0 2041-2060	*10.09*	*10.56*	*10.71*
72hr 3-7.0 2080-2099	*20.62*	*21.57*	*21.89*

Table 2: HIRDS change factor percentages applied for each storm length (hrs) and recurrence interval (yr) (in bold). Values extracted from Carey-Smith et al., (2018). HIRDS change factor applied for each scenario (italics) based on multiplication with the projected change in mean national air temperature (Table 3) (NIWA, 2025).

Climate Scenario	Mean Tair change (°C)
2-4.5 2041-2060	1.29
2-4.5 2080-2099	2.14
3-7.0 2041-2060	1.55
3-7.0 2080-2099	3.17

Table 3: Mean temperature change for each SSP scenario, national mean calculated from NIWA projections (NIWA, 2025).

3 Results

3.1 Model Accuracy and Predictor Variable Importance

Table 4 shows several different error metrics calculated for the model and for different probability thresholds. The model produces well-calibrated probabilities, with Brier scores of 0.018 (training) and 0.010 (testing) and log loss of 0.088 (training)

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

and 0.040 (testing), reflecting improved calibration on the real-world, imbalanced (99:1) testing dataset. Precision, recall and accuracy metrics are included for three threshold values (1%, 5%, 10%), with balanced accuracy (which accounts for class imbalance) ranging from 75-86% depending on the threshold applied.

	Training	Testing	10% Probability Threshold	5% Probability Threshold	1% Probability Threshold
ROC AUC (RIL)	NA	0.940	NA	NA	NA
ROC AUC (No-RIL)	NA	0.940	NA	NA	NA
PR AUC (RIL)	NA	0.276	NA	NA	NA
PR AUC (No-RIL)	NA	0.999	NA	NA	NA
Binary Log Loss	0.088	0.040	NA	NA	NA
Brier Score	0.018	0.010	NA	NA	NA
Precision	NA	NA	0.202	0.145	0.072
Recall	NA	NA	0.531	0.650	0.849
F1 Score	NA	NA	0.293	0.237	0.132
RIL Accuracy	NA	NA	53.1%	65.0%	84.9%
No-RIL Accuracy	NA	NA	97.6%	95.7%	87.5%
Accuracy	NA	NA	97.1%	95.3%	87.5%
Balanced Accuracy	NA	NA	75.4%	80.3%	86.2%

Table 4: Selected error metrics. ROC AUC = Receiver operating characteristic area under curve; PR AUC = Precision recall area under curve.

ROC curves for the holdout test dataset are presented in Fig. 2a. An area under curve (AUC) value of 0.94 is reported for both the RIL and no-RIL classes (where a value of 1 is perfect prediction and 0.5 is no better than random), which indicates that the model has high predictive capability; however, for highly imbalanced classes, ROC-AUC alone can be misleading. Figure 2b shows the precision-recall (PR) curve for both classes, along with the baseline precision for a model that predicted all instances as true for each class (i.e. a majority and a minority baseline classifier is included). No-RIL PR-AUC is very high at 0.999 and exceeds the baseline AUC of 0.989. However, for the (positive) RIL class, PR-AUC is only 0.276. This lower value is because the RIL class (positive case) is a minority in the testing dataset (99:1); thus, for a specific probability threshold, a relatively small percentage error for the no-RIL class equates to a large total number of pixels being misclassified. For example, at a 5% probability threshold (Fig. 3), false positives are just 4.3%; however, this results in 166,808 pixels being misclassified, considerably exceeding the true positives at 28,353 pixels. This is expected when there is extreme class imbalance (99:1). Selecting a suitable threshold for binary classification thus depends on the intended application (e.g. screening versus intervention) and the cost and benefits associated with false positives versus false negatives. Despite this low absolute precision-recall AUC value, the model significantly exceeds the baseline AUC of 0.011 (i.e., it performs ~26 times better than the baseline), confirming that it has strong predictive power for both classes. In real-world applications, this means that the model tends to be over-predictive, whereby large areas are identified as susceptible to RIL, but they may not actually experience a RIL consequent to a specific precipitation event. Together, these results indicate that the final model has strong predictive power in identifying both high and low-susceptibility RIL areas on unseen real-world data over the training area (Hawke's Bay and Gisborne/Tairāwhiti).

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

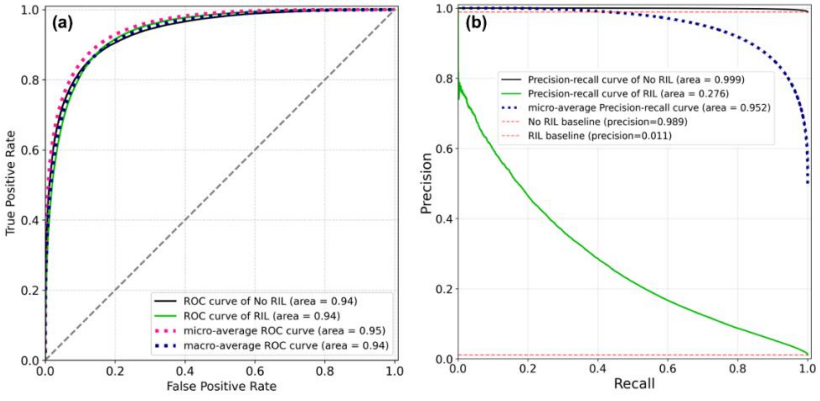

Figure 2: Receiver operating characteristic (ROC) curves (a) and precision-recall curves (b) for the test dataset.

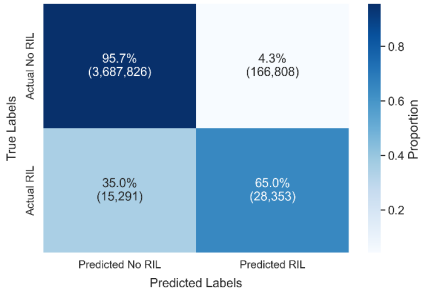

Figure 3: Confusion matrix for the test dataset at 5% probability threshold.

A bee swarm plot of SHAP feature importance for the test data is presented in Fig. 4, with the three most important variables being identified as: storm precipitation as a proportion of annual precipitation, slope, and forest height. SHAP values are used here to interpret the fitted model by quantifying each predictor's contribution to the predicted landslide probability. These attributions describe associations within the model and dataset, not causal effects, especially where predictors are correlated. On the x-axis, a more negative SHAP value indicates a lower RIL probability, while a more positive SHAP value indicates increased RIL probability. The bee swarm plots are coloured by the variable magnitude (red is higher, blue is lower). Figure 4 shows that high precipitation as a proportion of annual precipitation, steep slopes, and low forest height increase the likelihood of the model predicting that a pixel is susceptible to RIL.

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

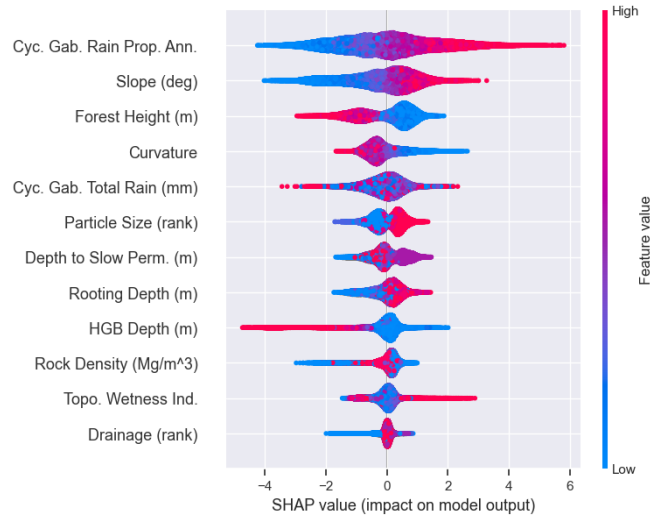

Figure 4: Bee swarm plot of SHAP values for the test dataset. Abbreviations: Cyc. Gab. Rain Prop. Ann. = Cyclone Gabrielle rain proportional to annual rainfall; Cyc. Gab. Total Rain (mm) = Cyclone Gabrielle total rainfall (mm); Depth to Slow Perm. = Depth to zone of slow permeability; HGB Depth (m) = Depth to hydrogeologic basement (m); Topo. Wetness Ind. = Topographic wetness index.

The SHAP dependence plots (Fig. 5) elucidate further insight into the relationships between model decisions of RIL probability in relation to the predictor variables. However, they do not assess causality, especially when predictors may be correlated (e.g. slope and curvature). The SHAP value is plotted on the y-axis and the predictor value on the x-axis; points are coloured by the strongest interacting feature for the x-axis predictor, as identified using SHAP interaction ranking. These plots show that the relationship between modelled RIL susceptibility and precipitation as a proportion of annual rainfall is linear, and that steep slopes with high rainfall have increased RIL susceptibility (variable interaction). Slope exhibits a non-linear (logarithmic) relationship to landslide susceptibility. Increasing rapidly from 0-15°, with a more gradual increase in RIL susceptibility for slopes greater than 15°. Note, the reduced landslide susceptibility for slopes less than 10° may also be a consequence of the RIL mapping exercise, where slopes under 10° that transitioned to bare ground were not classified as RIL. Forest height is the third most important predictor variable, where areas with trees over ~5 m tall have lower RIL susceptibility, and the highest susceptibility areas have steep slopes and short/no trees. Interestingly, the lowest susceptibility areas are areas with tall trees and steep slopes, possibly because these are the areas with the most intact native forest cover. Negative slope curvature values are more likely to experience RIL, where negative values identify zones of convergence, i.e. plan and/or profile concavity, while positive curvature areas (zones of divergence) have lower susceptibility. Zones of convergence with high proportional precipitation have the highest susceptibility, while zones of divergence with high proportional precipitation have the lowest RIL susceptibility. Increasing depth to the hydrogeologic basement decreases RIL susceptibility because deeper sediments are found in the valley floors. Conversely, increasing rooting depth and the depth to the zone of permeability (which essentially describes the soil thickness) increases RIL susceptibility, presumably because more surface sediment is susceptible to sliding. However, where depth to the zone of slow permeability is at its maximum (>2.25 m), RIL susceptibility is low; these areas are valley floors. Rooting depths' highest interacting feature is forest height, where tall trees and deep soils have elevated RIL susceptibility, and tall trees in shallow soils have reduced RIL susceptibility. Both small and large particle sizes have an increased probability of RIL; these correspond to clays (class 1) and loose gravels (class 5), with the greatest risk for larger

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

particles. For areas with rock density exceeding ~2.1 Mg/m^3, RIL susceptibility reduces with increasing rock density; below ~2.1 Mg/m^3, RIL susceptibility is noticeably lower (step change). This counterintuitive relationship is likely due to the lowest-density geologies (e.g. unconsolidated sand) being only present in the valley floors, where RIL susceptibility is low. Areas with higher topographic wetness index values and steeper slopes generally have an elevated RIL susceptibility. Drainage is the least important predictor variable; areas with the lowest RIL susceptibility are 'very poorly drained' (class 1), which is counterintuitive. However, again, this is likely because clay-type substrates with poor drainage are predominantly located in valley floors with low slopes.

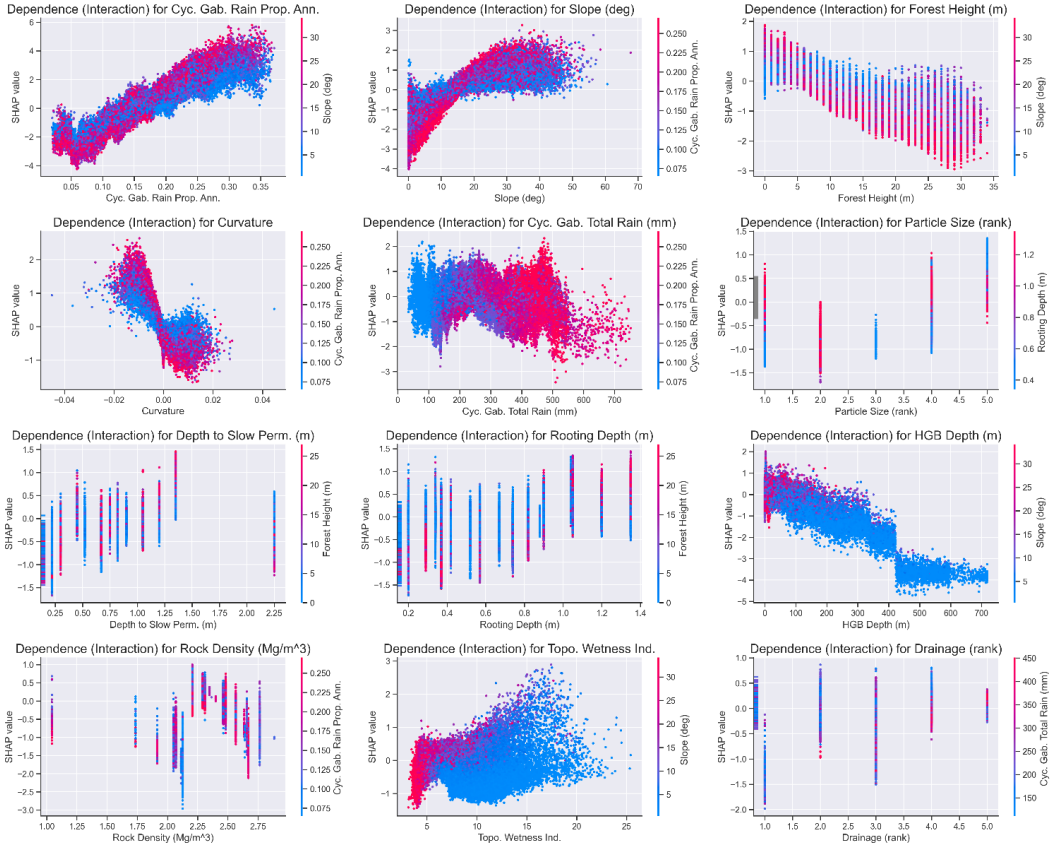

Figure 5: SHAP dependence plots for each predictor variable on the test set, coloured by the highest interacting variable. Abbreviations: Cyc. Gab. Rain Prop. Ann. = Cyclone Gabrielle rain proportional to annual rainfall; Cyc. Gab. Total Rain (mm) = Cyclone Gabrielle total rainfall (mm); Depth to Slow Perm. = Depth to zone of slow permeability; HGB Depth (m) = Depth to hydrogeologic basement (m); Topo. Wetness Ind. = Topographic wetness index.

3.2 Model Predictions and Climate Projections

Model outputs for New Zealand (Wigmore, 2026) under the two storm durations, three recurrence intervals, and four SSP scenarios/periods are shown in Fig. 6. The figures show that as storm precipitation increases, so too does RIL susceptibility, which is expected due to the two rainfall-related variables being the first and fifth most important predictor variables in the SHAP analysis Fig. 4. For the different SSP scenarios, the results indicate that a warmer climate increases RIL susceptibility,

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

with the greatest susceptibility posed under scenario 3-7.0 at 2080-2099 (the warmest temperature scenario modelled here). Figure 7 presents a close-up view of the model output for Banks Peninsula for a 24-hour storm under several of the different modelled scenarios. These results show the increasing RIL susceptibility for a 24-hour storm from a 10-100-year recurrence interval; of particular interest is the much larger area of susceptible land under the highest temperature future modelled here (SSP 3-7.0, 2080-2099). Figure 8 shows a close-up of Hawke's Bay for a 72-hour storm. Again, we can clearly see the considerable increase in the extent of susceptible land under the warmest future climate scenario, where storm events are projected to deliver the greatest increase in precipitation. Of note is the much lower increase in susceptible areas where there is a well-established native forest, e.g. Te Urewera.

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

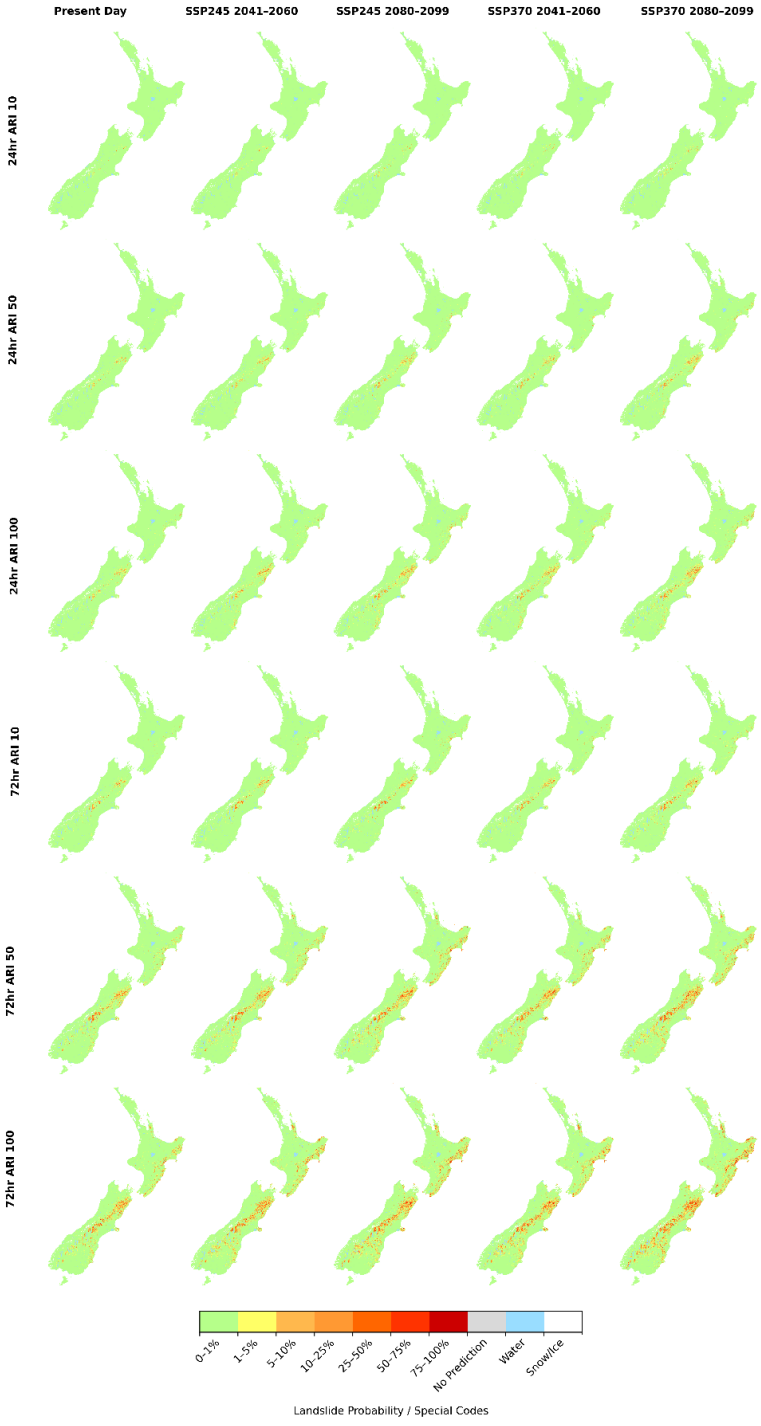

Figure 6: Model output of all storm lengths, recurrence intervals, and SSPs. Rows denote the different storm lengths and ARI, and columns show the different SSPs for each row (Wigmore, 2026).

440

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

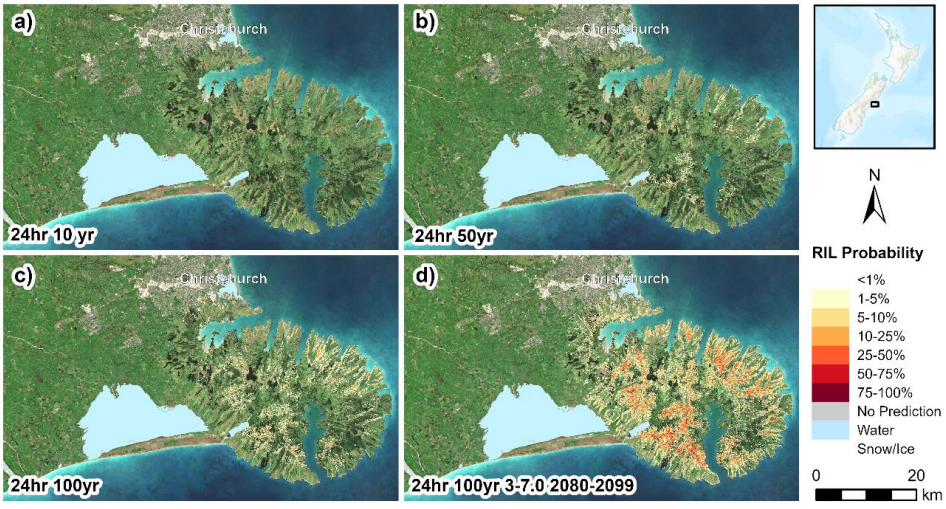

Figure 7: Close-up of selected model outputs for Banks Peninsula. Basemap imagery is a Sentinel 2 national mosaic 2023-2024 compiled by Land Information New Zealand and distributed under a CC-BY 4.0 license.

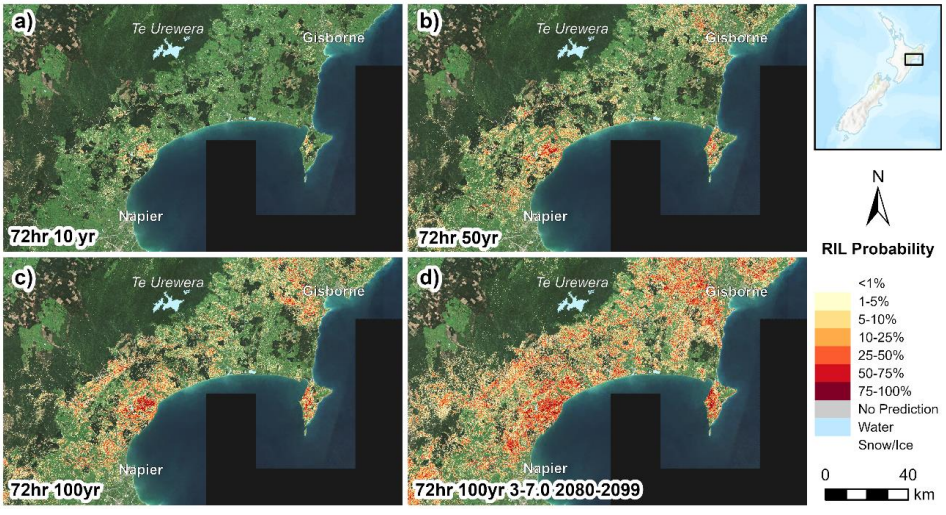

Figure 8: Close-up of selected model outputs for Hawke's Bay. Basemap imagery is a Sentinel 2 national mosaic 2023-2024 compiled by Land Information New Zealand and distributed under a CC-BY 4.0 license.

Figure 9 shows the total area at or exceeding a 5% RIL probability under each of the modelled storm and climate scenarios, broken down by the sixteen administrative regions of New Zealand. The white dots show the estimated total land area that is likely to actually experience a RIL event under each scenario when accounting for model true positives and false negatives.

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

Canterbury, Marlborough, Otago, Hawke's Bay, and Gisborne/Tairāwhiti have the highest RIL exposure under each of the scenarios. The area of susceptible land increases under each of the modelled climate futures, with the largest increase under SSP 3-7.0 at 2080-2099, the warmest scenario explored here. This is expected because the HIRDS rainfall projections increase in response to temperature. Figure 10 shows the percentage change in total corrected RIL area (accounting for true positives and false negatives) relative to the baseline exposure of current conditions aggregated at the national scale. The largest increase is for a 24-hour 50-year storm under SSP370 2080-2099, where the RIL impacted area is estimated to increase by 272%. For this scenario, the minimum increase is 93% for a 72-hour 100-year storm. In all cases, the shorter-duration 24-hour storms experience a larger increase in RIL-impacted area compared to 72-hour storms. There is at least a 34% (and up to 84%) increase in RIL-impacted area by 2041-2060 under all the scenarios modelled here. However, an important finding is that for all scenarios, the percentage increase in RIL impacted area in 2080-2099 under a lower emission scenario (SSP2-4.5) is almost half what it is under the higher emission scenario (SSP3-7.0). Figure 11 shows the total length of road (centreline) and the number of buildings that are modelled as impacted by the different modelled scenarios. For all scenarios, the impacted road length and number of buildings are approximately double what they are currently under SSP3-7.0 2080-2099. It is notable that the percentage increase in land area and assets exposed is disproportionate to the modelled increase in rainfall consequent to warmer temperatures. For example, for a 72-hour 100-year storm scenario, the susceptible area (Fig. 10), road length, and number of buildings (Fig. 11) roughly double under SSP 3-7.0 2080-2099; however, the percentage increase in storm precipitation is only 21.89% (HIRDS change factor Table 2)

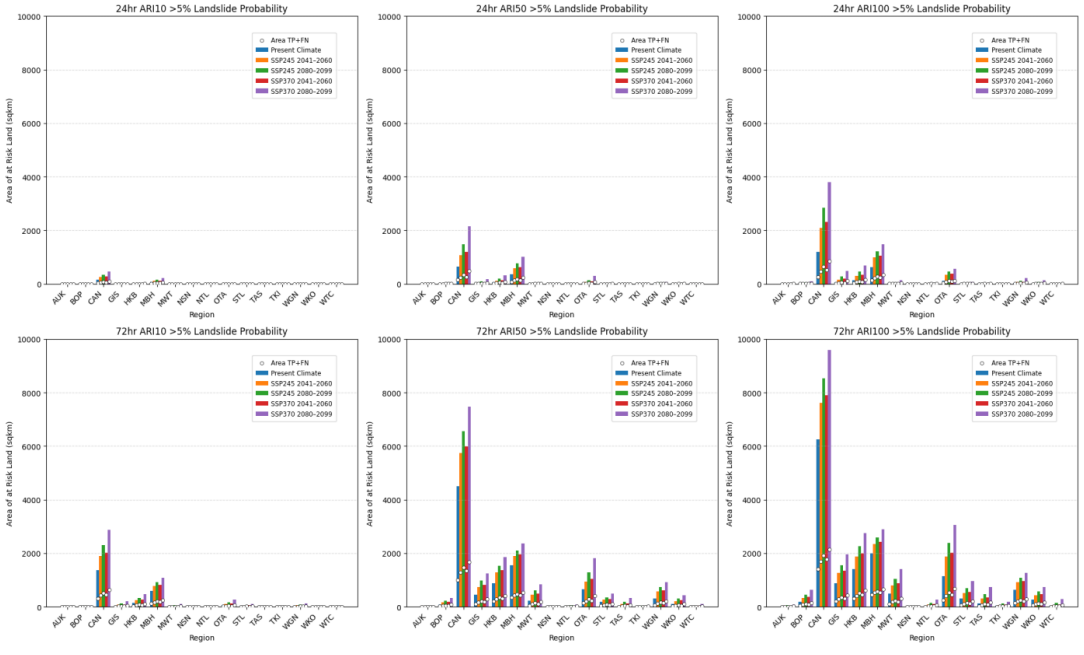

Figure 9: Bar plot of total land area at or exceeding 5% RIL probability for each storm scenario, broken down by region. Individual coloured bars correspond to the different climate SSPs. White dots (Area TP+FN) show the corrected total area likely to experience a RIL using confusion-matrix error propagation to account for both TP and false negatives FN at the 5% probability threshold. Regional abbreviations are as follows: AKL: Auckland, BOP: Bay of Plenty, CAN: Canterbury, GIS: Gisborne/Tairāwhiti, HKB: Hawke's Bay, MBH: Marlborough, MWT: Manawatu, NSN: Nelson, NTL: Northland, OTA: Otago, STL: Southland, TAS: Tasman, TKI: Taranaki, WGN: Wellington, WKO: Waikato, WTC: West Coast.

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

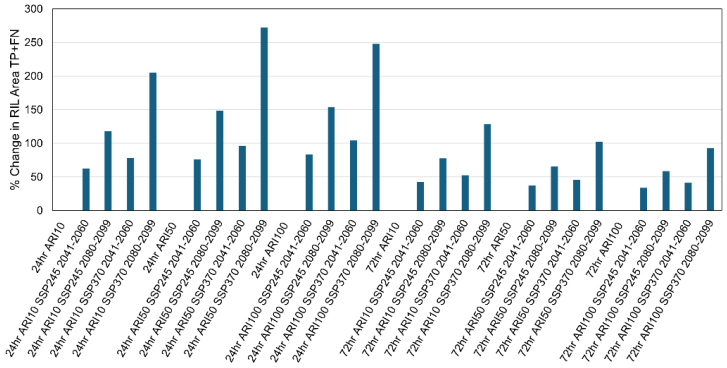

Figure 10: Percentage change in RIL area (national scale) accounting for true positives and false negatives, relative to baseline of exposure under current climate conditions.

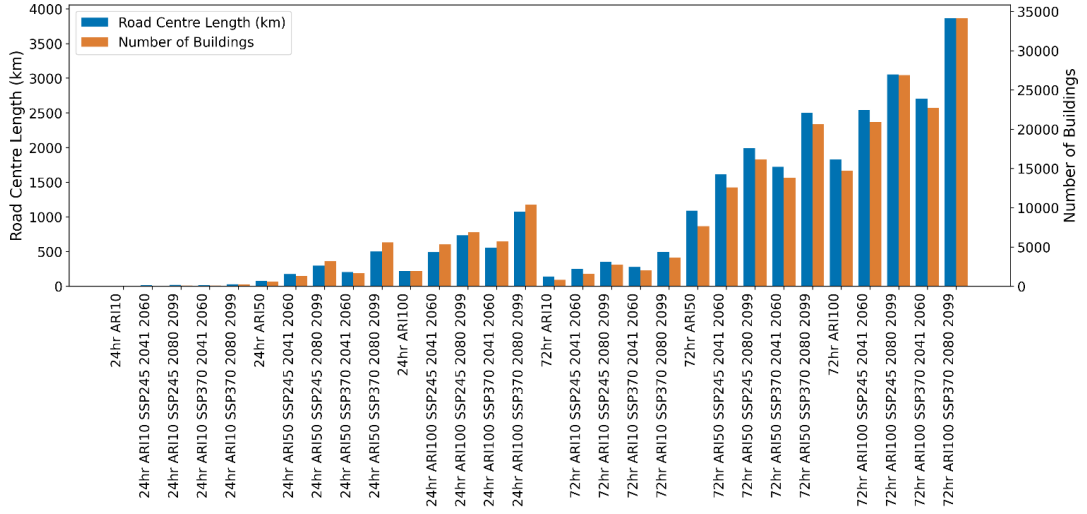

Figure 11: Total length of roads and number of buildings exposed to greater than 5% RIL probability under each storm and climate scenario.

4 Discussion

4.1 Model Insights

The model was able to predict RIL susceptibility with a ROC AUC of 0.94. This is similar to the ROC AUC value of 0.93 reported by Griffiths et al. (2020) for their model in the Tasman region, trained on Cyclone Gita observations. At a 5% probability threshold, the model has a 65.0% accuracy of predicting RILs and 95.7% accuracy of no-RIL prediction (class balanced accuracy 80.3%); at a 1% probability threshold, accuracy is 84.9% for RIL and 87.5% for no-RIL (class balanced accuracy 86.2%). Given the significant class imbalance (99:1 no-RIL:RIL), there are large areas that are predicted as susceptible, which may not actually experience a RIL under a given storm scenario. However, these results are a considerable

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

improvement on the existing national erosion susceptibility maps (which includes landslides), which have been shown to have an accuracy of only 26% at the pixel level (Dymond et al., 2006) and 58-75% at the hillslope scale (Dymond et al., 2006; Dymond & Shepherd, 2023) (false positive rates are not reported). Importantly, the model accounts for rainfall intensity as a triggering factor and provides not just a simple binary classification but RIL probability for each specified storm. It is necessary to note that the training dataset, model, and resultant susceptibility maps do not separate the RIL initiation point from the debris run-out zone; consequently, neither is explicitly modelled. Because some of the predictor variables describe pixel neighbourhoods (e.g. TWI and curvature), in many cases, both the initiation point and run-out zone are included in the mapped RIL probability area. However, it is possible that the impacted downslope area extends beyond the defined RIL susceptible area.

Under the highest emissions scenario modelled in this study (SSP 3-7.0 at 2080-2099), New Zealand's mean annual air temperature is projected to increase by 3.17°C compared to the 1986-2005 mean (NIWA, 2025). This results in a 27.3% (24-hour storm) and 21.9% (72-hour storm) increase in the precipitation received in a 1 in 100-year storm. Because of this increased precipitation, we model a disproportionate increase in the land area likely to be impacted by RIL (accounting for true positives and false negatives), in this case, a 248% (24-hour storm) and 94% (72-hour storm) increase, respectively. This percentage increase is reduced for the lower emission SSP 2-4.5 scenario, which highlights the importance of reducing emissions to mitigate amplification of RIL susceptibility, and minimise potential asset exposure and loss of life. Studies in Southwest China (Yin et al., 2023) and California (Semnani et al., 2025) have identified a similar relationship between higher emissions scenarios (warmer temperatures) and increased RIL susceptibility. In New Zealand, Neverman et al., (2023) have shown that warmer temperatures are likely to increase river sediment loads, which is driven in part by increased RIL activity. Similarly, Crozier (2010) showed that observationally, RIL frequency increased non-linearly with mean rainfall. And posited that climate change was likely to increase RIL frequency because of increased frequency and intensity of rainfall and significant storm events.

SHAP analysis allows us to better interpret the black box of ML and gain insight into how the different variables influence model predictions. We showed that many of these relationships are non-linear and have important and informative interactions. This illustrates the suitability of gradient boosted decision trees for RIL susceptibility modelling, due to their robustness to both non-linearity and multicollinearity. Leveraging SHAP analysis, we can hypothesise potential mechanisms for these relationships. For example, previous research has documented the importance of tree cover in reducing landslide susceptibility, and is driving calls for nature-based solutions in the form of reforestation to mitigate this risk (Griffiths et al., 2020; Phillips et al., 2021; Spiekermann et al., 2022). The model indicates that tree height is the third most important predictor variable (behind rainfall and slope). We show consistently lower RIL susceptibility over areas with extensive and established native forest cover (where tree height exceeds 5m), despite other risk factors (steep slopes, high rainfall, and weak geology), e.g. Te Uruwera (Fig. 8), even under the most extreme storm and climate scenarios modelled here. This is consistent with earlier work by (Griffiths et al., 2020), who found that landcover was the most important determinant of landslide susceptibility in the Tasman region of New Zealand, where limiting clear-fell forestry and increasing permanent forest cover could substantially reduce landslide susceptibility. Crozier (2010), argued for the importance of considering human factors, notably deforestation, in driving landslide risk, with this being of potentially greater significance than climate change. To further explore the impact of forest cover and the potential impact of afforestation to reduce RIL susceptibility, the model could be rerun with synthetic (adjusted) tree height data; however, this is outside of the scope of this paper. The relatively high resolution and national scale of the work presented herein may be useful in targeting catchment-scale reforestation interventions where they are likely to have the most benefit.

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

Importantly, despite the reduced susceptibility of RIL in forested areas, the probability is not zero, because RIL susceptibility depends on both the canopy height and the intensity of precipitation received. The only currently available national landslide susceptibility assessment (Dymond & Shepherd, 2023) does not include either of these variables, and simply classifies forested areas as 'no risk'. Hence, the modelling efforts presented herein are a significant advance in our understanding of the nuanced and variable reduction in RIL susceptibility that is provided by forest cover. This improved understanding of potential RIL susceptibility in forested areas is critical for managing biodiversity in these regions, understanding potential exposure of river ecosystems to landslide-derived sediment deposition, and exposure of assets and infrastructure within these areas, e.g. roads, buildings, huts, campsites, track networks, etc. Furthermore, the ability to quantitatively assess the likely recurrence frequency of landslide damage can facilitate better-informed planning and decision-making with respect to asset management.

4.2 Applications and Limitations

The output model predictions provide a first-of-their-kind national dataset of RIL susceptibility for New Zealand, which considers precipitation (event trigger) under different storm scenarios and climate futures (SSPs), and provides prevalence-scaled event-conditional RIL probability for each specified storm. It is at a suitable scale (25m) for national planning and to guide the targeted acquisition of more intensive studies (e.g. geotechnical assessments, and fine-scale physically based modelling) of landslide susceptibility as required. However, there are important caveats to consider. The most important being the limited geographical scale of the training and testing dataset, which was only taken from Hawke's Bay and Gisborne/Tairāwhiti. At a national scale, Hawke's Bay and Gisborne/Tairāwhiti regions can be considered as likely having elevated RIL susceptibility due to the presence of steep unforested slopes, deep soils, and relatively weak geology comprised primarily of sandstone and mudstone (Heron, 2023). In the model, we account for a full range of precipitation intensities, forest cover, topography and soil depth variables; however, geology is only constrained through the rock density variable (Fig. C1-2). Higher rock densities are correlated with rock types that are less susceptible to landslides, so this is an efficient manner of numerically quantifying this variable; however, it does not directly address the relationship between geology and RIL susceptibility. Another approach could be to either use simplified geology as a categorical variable or to score geology types with respect to their RIL susceptibility. The former was not possible because not all geology classes are present within the training region, while the latter is somewhat subjective and depends on not just the rock type but also on the determination of slope failure angles. Finally, while the highest rock densities are present in the training and testing region (i.e. the model is not predicting out of range), they are less prevalent there than in the South Island, and particularly the Southern Alps. The Southern Alps are dominated by hard (and dense) metamorphic rocks; whether the rock density RIL susceptibility relationship modelled here remains the same for these different lithologies is unproven. Furthermore, mass wasting in much of the Southern Alps tends to be dominated by active scree/talus slopes, rock avalanches, and large deep-seated landslides (often earthquake-induced) compared with the frequent RIL, shallow slips and slumps of the North Island (particularly the East Coast and Northland) (Allen et al., 2011; Crozier, 2010b; Hales & Roering, 2005; Korup, 2006). Consequently, the interpretation of RIL susceptibility maps for the Southern Alps should be undertaken cautiously. In the future, the model could be tested against observational data collected after a strong, widespread storm event over the South Island and potentially retrained with the inclusion of this data.

A secondary caveat is that the training precipitation data comes from a single storm, which may not be reflective of the national predictions of future storm events (HIRDS data), i.e. because Cyclone Gabrielle came from the northeast. We minimised the influence of this by not including directionally dependent predictor variables (e.g. aspect), which have been shown to be strong predictor variables in other similar studies (Chen et al., 2024). Fortunately, the rainfall distribution during Cyclone Gabrielle was highly heterogeneous. Therefore, the training dataset has a broad range of values that cover the data range of all future

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

storm and climate scenarios modelled herein, thus we are not predicting out of range. Antecedent rainfall and consequent antecedent soil moisture are also not included as predictor variables, because it is not possible to predict these as an input for future climate/storm scenarios. The Hawke's Bay and Gisborne/Tairāwhiti training regions had experienced significant rainfall in the preceding weeks, and soil moisture was uniformly high (Massey et al., 2025). Previous research has documented the importance of high antecedent soil moisture as a contributor to RIL occurrence (Crozier, 1999). Consequently, it is possible that the model developed herein over-predicts RIL susceptibility. However, Massey et al. (2025) found that in the case of Cyclone Gabrielle, where rainfall intensities were sufficiently high (e.g. much of Hawke's Bay and Gisborne/Tairāwhiti), antecedent soil moisture had minimal influence on landslide occurrence. Regardless, the potential overprediction of landslide susceptibility should be considered when interpreting these results or using the output datasets of this work.

Finally, it is important to consider the spatial scale, temporal context, and accuracy of the predictor variables used in this study. We use a 25m satellite-derived bare earth DEM (resampled from 30m) (Hawker et al., 2022; Neal & Hawker, 2023) to derive topographic metrics (slope, TWI, curvature). However, RIL triggering processes are likely to be driven by finer resolution topography than this. Superior results could be achieved through the utilisation of a higher accuracy and/or higher resolution DEM, e.g. national LiDAR when it is eventually completed. However, working at a higher resolution increases computational requirements. Additionally, the DEM, while generally of high quality, does include both relative and absolute elevation errors. Consequently, there are likely to be on-the-ground errors in RIL susceptibility driven by DEM errors; i.e. the model is predicting correctly, but the underlying DEM variables (slope, TWI, curvature) are incorrect, producing poor on-the-ground results. Rainfall variables (both those for training and prediction) were statistically downscaled (bilinear interpolation) from coarser 500m (training and testing) and 2000m grids (HIRDS prediction) to 25m pixels. This does not account for local influences (e.g. topography) on rainfall distribution. Improvements could potentially be made by considering more elegant methods for rainfall downscaling. Forest height is a satellite-derived product derived in 2019 (Potapov et al., 2021), which precedes Cyclone Gabrielle by approximately four years. During this time, changes in forest height are likely to have occurred, particularly in commercial forestry land, where large-scale planting and clear felling may have occurred, the timing of which can significantly impact RIL susceptibility (Phillips et al., 2024). This may confound the model through the inclusion of noise in the predictor variable signal. Because tree height was the third most important predictor variable in the model, it is important to consider potential changes in tree height between the training data (2019), the present day, and future land cover. For example, an area with tall trees in the training data that has since been clear-felled has likely experienced an increase in RIL susceptibility compared to the model output. Conversely, a bare ground area in the training data with high RIL susceptibility that has been reforested has likely experienced a reduction in RIL susceptibility.

Internationally, RIL susceptibility mapping has increasingly aimed to integrate static conditioning factors with event-specific rainfall scenarios, with the aim of improving understanding of recurrence frequency, storm-specific forecasting (Bordoni et al., 2021; Fumagalli et al., 2025; Mondini et al., 2023; Rosser et al., 2021; Segoni et al., 2018), and the impacts of climate change (Han & Semnani, 2025; Semnani et al., 2025; Yin et al., 2023). This can support planning applications and potentially the development of early warning systems, potentially reducing RIL hazard exposure. Our workflow could be applied to other temperate or high-intensity rainfall locations, provided that a suitable RIL inventory is available for training and that suitable conditioning variable layers and rainfall products exist. The primary contribution of this study is a national-scale implementation for New Zealand that combines ML RIL susceptibility modelling with scenario-specific rainfall events. Quantifies the impact of climate change on RIL susceptibility and assesses potential infrastructure exposure across these scenarios.

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

5 Conclusion

This study presents a national-scale model for assessing RIL susceptibility across Aotearoa New Zealand that includes both conditioning and triggering factors. By leveraging a robust training dataset derived from Cyclone Gabrielle's impacts in the Hawke's Bay and Gisborne/Tairāwhiti regions, the model demonstrates high predictive capability (ROC AUC = 0.94). Key predictors such as storm precipitation intensity, slope, and forest height were identified as critical determinants of RIL susceptibility through comprehensive SHAP analysis. The model's application to a suite of current and future storm scenarios, informed by NIWA's HIRDS dataset and SSP scenarios, reveals a consistent and significant increase in RIL susceptibility under warmer climate futures. We find that even for the shortest timelines (2041-2060), there is at least a 34% (and up to 84%) increase in RIL-impacted area for all the climate and storm scenarios modelled here. Furthermore, we estimate a 93-272% increase in the RIL-impacted land under the highest temperature future climate scenario modelled here (SSP3-7.0 2080-2099). The lower emission SSP scenario (SSP2-4.5 2080-2099) roughly halves this increase, emphasising the importance of reducing greenhouse gas emissions. This underscores the urgent need for climate-adaptive land use planning and the implementation of nature-based solutions, particularly reforestation in erosion-prone areas. The model highlights the protective role of established native forest cover, which consistently reduces RIL susceptibility even under extreme precipitation scenarios. Despite its strengths, the model has limitations, including the geographic concentration of training and testing data, limited geological classification, and potentially the exclusion of antecedent soil moisture. Furthermore, some of the predictor datasets may be out of date (e.g. forest height) or inaccurate. Future research should aim to expand the training dataset to include diverse geological settings, incorporate more accurate and temporally coincident forest height data, and utilise accurate high-resolution topographic data, i.e. the national lidar DEM, which is nearing completion. Overlaying the model's output with socio-economic exposure data can enhance its utility for disaster risk reduction and infrastructure planning. Overall, this work provides a scalable and interpretable framework for national RIL susceptibility assessment and offers valuable insights for policymakers, conservation planners, and emergency management agencies in building climate resilience.

6 Appendix A

Model hyperparameters were tuned with Optuna using 10-fold cross-validation on the training dataset with early stopping after 50 rounds. The final tuned model parameters were:

- Objective: binary
- Metric: binary log loss
- Boosting type: gbdt
- Number of leaves (num_leaves): 249
- Feature fraction: 1.0
- Bagging fraction: 0.952
- Bagging frequency (bagging_freq): 1
- L1 regularisation (lambda_l1): 0.499

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

- L2 regularisation (lambda_l2): 1.07×10^{-5}
- Minimum child samples (min_child_samples): 5
- Early stopping rounds: 50
- Maximum boosting rounds: 2000

All other parameters were left at LightGBM defaults.

7 Appendix B

Model calibration was assessed on a random subset of the holdout test dataset, sampled at 50:50 distribution (to match the training data distribution). Sigmoid and isotonic calibration were compared against the raw model probabilities. The raw model probabilities showed stable prediction frequency across the probability range, with no meaningful improvements using either sigmoid or isotonic calibration. Raw model probabilities were therefore used to calculate the prevalence-adjusted probabilities.

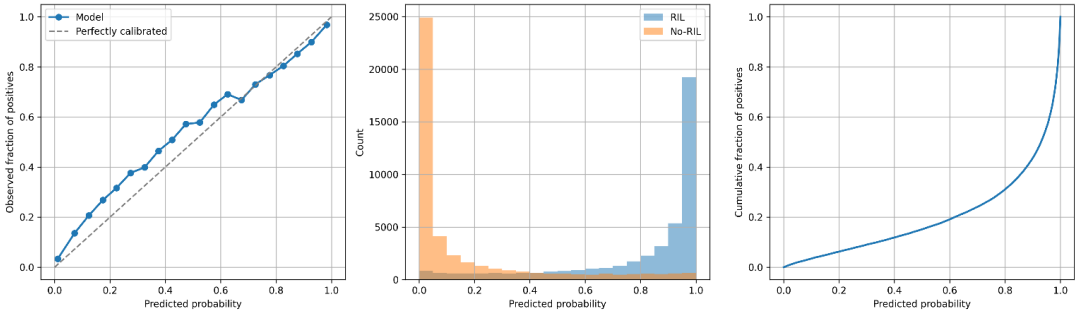

Figure B1: Calibration plots for the trained model, on a 50:50 distribution subset of the holdout test dataset. Calibration reliability curve (left); Probability histogram (centre); Cumulative fraction of RIL (right).

8 Appendix C

The model was trained and tested on RIL observation from across the Hawke's Bay and Gisborne/Tairāwhiti regions of New Zealand's North Island and then run across the entire country. Figures C1 and C2 compare the data distribution for the ten non-rainfall predictor variables for the training/testing domain (Hawke's Bay and Gisborne/Tairāwhiti), the rest of the North Island, and the South Island (including Rakiura/Stewart Island). The plots were generated from a 1% random sample of the data. For all variables, the training/testing data ranges are the same as those on which predictions are made, i.e. the model is not being run outside its training range. Data distributions for most variables are relatively similar; however, the South Island has generally higher rock densities and shallower rooting depths. These plots provide greater confidence in the findings when the model is run outside its training and testing domain.

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

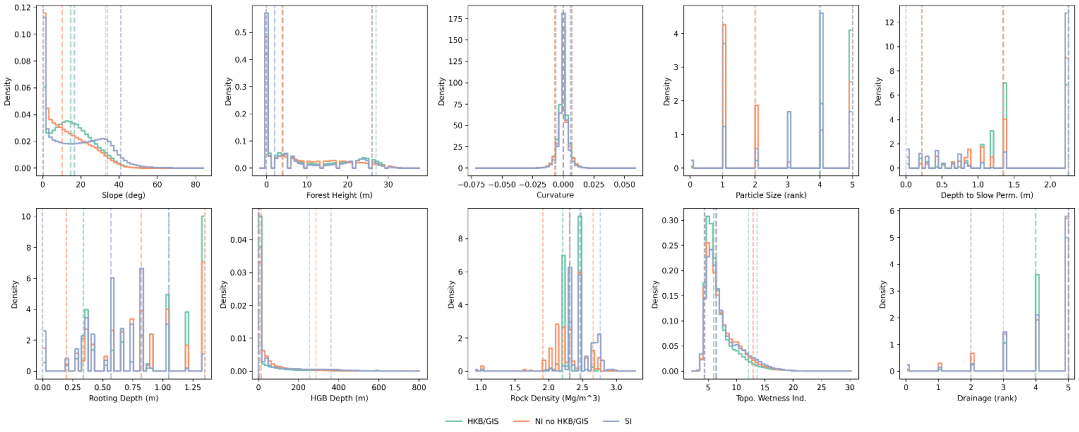

Figure C1: Step density histograms for the ten non-rainfall predictor variables. HKB/GIS = Hawke's Bay and Gisborne/Tairāwhiti, i.e. the model training and testing domain. NI no HKB/GIS = all regions in the North Island of New Zealand excluding Hawke's Bay and Gisborne/Tairāwhiti, i.e. North Island areas outside the training and testing domain. SI = South Island (including Rakiura/Stewart Island), which are also outside the training and testing domain. Vertical dashed lines mark the 5th, 50th (median), and 95th percentiles per region.

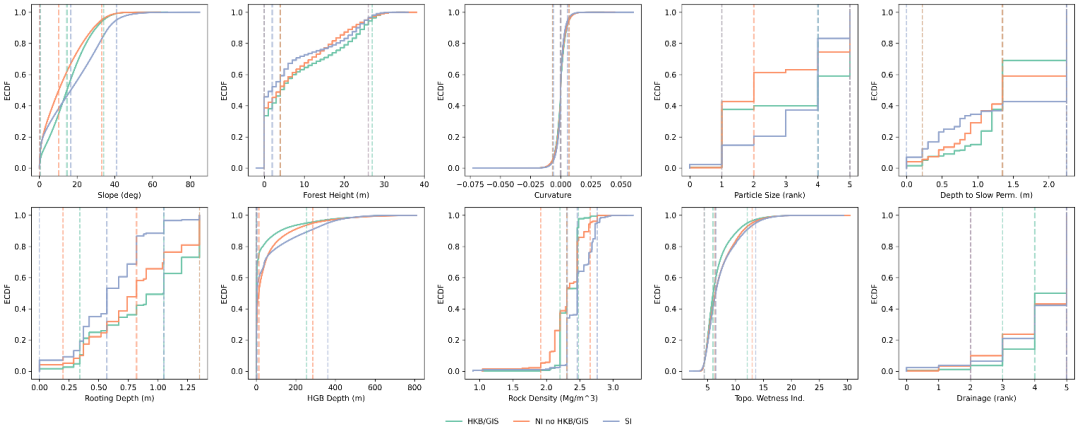

Figure C2: Empirical cumulative distribution function (ECDF) for the ten non-rainfall predictor variables, separated by region. HKB/GIS = Hawke's Bay and Gisborne/Tairāwhiti, i.e. the model training and testing domain. NI no HKB/GIS = all regions in the North Island of New Zealand excluding Hawke's Bay and Gisborne/Tairāwhiti, i.e. North Island areas outside the training and testing domain. SI = South Island (including Rakiura/Stewart Island), which is also outside the training and testing domain. Vertical dashed lines mark the 5th, 50th (median), and 95th percentiles per region.

Data Availability

RIL predictions for all SSP scenarios are available here:

Wigmore, O. (2026). *National Rainfall-Induced Landslide Susceptibility Maps (25 m) for Extreme Rainfall Scenarios under Present-Day and Future Shared Socioeconomic Pathways (SSPs), New Zealand* (Version 1.0.0) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.17850323

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

Note: data access is embargoed until completion of the peer review process. Reviewers have been provided with anonymous access to support the peer review process.

Sentinel 2 Data Strip IDs used for Cyclone Gabrielle RIL mapping.

5/02/2023 (NZDT date) – pre-cyclone
S2B_MSIL2A_20230204T221559_N0510_R129_T60HWD_20240731T061234.SAFE
S2B_MSIL2A_20230204T221559_N0510_R129_T60HXD_20240731T061234.SAFE
S2B_MSIL2A_20230204T221559_N0510_R129_T60HWC_20240731T061234.SAFE
S2B_MSIL2A_20230204T221559_N0510_R129_T60HXC_20240731T061234.SAFE
S2B_MSIL2A_20230204T221559_N0510_R129_T60HVB_20240731T061234.SAFE
S2B_MSIL2A_20230204T221559_N0510_R129_T60HWB_20240731T061234.SAFE
S2B_MSIL2A_20230204T221559_N0510_R129_T60GVA_20240731T061234.SAFE

7/02/2023 (NZDT date) – pre-cyclone
S2A_MSIL2A_20230206T220621_N0511_R086_T60HWD_20250806T055140.SAFE
S2A_MSIL2A_20230206T220621_N0511_R086_T60HXD_20250806T055140.SAFE
S2A_MSIL2A_20230206T220621_N0511_R086_T60HWC_20250806T055140.SAFE
S2A_MSIL2A_20230206T220621_N0511_R086_T60HXC_20250806T055140.SAFE
S2A_MSIL2A_20230206T220621_N0511_R086_T60HVB_20250806T055140.SAFE
S2A_MSIL2A_20230206T220621_N0511_R086_T60HWB_20250806T055140.SAFE
S2A_MSIL2A_20230206T220621_N0511_R086_T60GVA_20250806T055140.SAFE

10/02/2023 (NZDT date) – pre-cyclone
S2A_MSIL2A_20230209T221601_N0510_R129_T60HXD_20240801T113458.SAFE
S2A_MSIL2A_20230209T221601_N0510_R129_T60HWD_20240801T113458.SAFE
S2A_MSIL2A_20230209T221601_N0510_R129_T60HXC_20240801T113458.SAFE
S2A_MSIL2A_20230209T221601_N0510_R129_T60HWC_20240801T113458.SAFE
S2A_MSIL1C_20230209T221601_N0510_R129_T60HVC_20240801T090929.SAFE
S2A_MSIL2A_20230209T221601_N0510_R129_T60HWB_20240801T113458.SAFE
S2A_MSIL1C_20230209T221601_N0510_R129_T60HVB_20240801T090929.SAFE

20/02/2023 (NZDT date) – post-cyclone
S2A_MSIL2A_20230219T221601_N0510_R129_T60HXD_20240729T200132.SAFE
S2A_MSIL2A_20230219T221601_N0510_R129_T60HWD_20240729T200132.SAFE
S2A_MSIL2A_20230219T221601_N0510_R129_T60HXC_20240729T200132.SAFE
S2A_MSIL2A_20230219T221601_N0510_R129_T60HWC_20240729T200132.SAFE
S2A_MSIL2A_20230219T221601_N0510_R129_T60HVC_20240729T200132.SAFE
S2A_MSIL2A_20230219T221601_N0510_R129_T60HWB_20240729T200132.SAFE
S2A_MSIL2A_20230219T221601_N0510_R129_T60HVB_20240729T200132.SAFE
S2A_MSIL2A_20230219T221601_N0510_R129_T60GVA_20240729T200132.SAFE

22/02/2023 (NZDT date) – post-cyclone
S2B_MSIL2A_20230221T220619_N0510_R086_T60HWD_20240730T062532.SAFE
S2B_MSIL2A_20230221T220619_N0510_R086_T60HXD_20240730T062532.SAFE
S2B_MSIL2A_20230221T220619_N0510_R086_T60HWC_20240730T062532.SAFE
S2B_MSIL2A_20230221T220619_N0510_R086_T60HXC_20240730T062532.SAFE
S2B_MSIL2A_20230221T220619_N0510_R086_T60HVB_20240730T062532.SAFE
S2B_MSIL2A_20230221T220619_N0510_R086_T60HWB_20240730T062532.SAFE
S2B_MSIL2A_20230221T220619_N0510_R086_T60GVA_20240730T062532.SAFE

690

Author contributions

OW completed all aspects of the work contained herein.

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

Competing Interests

The author declares no competing interests.

Acknowledgements

OW would like to acknowledge Prof. Bryan G. Mark and Prof. Robert McKay for providing feedback on the manuscript prior to submission. OW would also like to acknowledge Nicki Atkinson and Michelle Crowell (Department of Conservation) for supporting this work.

AI tools were used for code debugging and language editing. ChatGPT-4 (OpenAI) and Microsoft Copilot (GPT-5) were used to generate initial code snippets and troubleshoot/debug; Grammarly (Grammarly Inc.) and Microsoft Copilot (GPT-5) were used for grammar and style improvements. All outputs were reviewed and edited by the author, who takes full responsibility for the final manuscript.

References

Achu, A. L., Aju, C. D., Di Napoli, M., Prakash, P., Gopinath, G., Shaji, E., & Chandra, V. (2023). Machine-learning based landslide susceptibility modelling with emphasis on uncertainty analysis. *Geoscience Frontiers*, *14*(6), 101657. https://doi.org/10.1016/j.gsf.2023.101657

Allen, S. K., Cox, S. C., & Owens, I. F. (2011). Rock avalanches and other landslides in the central Southern Alps of New Zealand: A regional study considering possible climate change impacts. *Landslides*, *8*(1), 33–48. https://doi.org/10.1007/s10346-010-0222-z

Al-Najjar, H. A. H., Pradhan, B., Kalantar, B., Sameen, M. I., Santosh, M., & Alamri, A. (2021). Landslide Susceptibility Modeling: An Integrated Novel Method Based on Machine Learning Feature Transformation. *Remote Sensing*, *13*(16), 3281. https://doi.org/10.3390/rs13163281

Barringer, J. (2018). *Fundamental Soil Layer—NZ Soil Classification* [Dataset]. Manaaki Whenua Landcare Research. https://doi.org/10.7931/L10T0

Bloom, C. K., Singeisen, C., Stahl, T., Howell, A., Massey, C., & Mason, D. (2023). Coastal earthquake-induced landslide susceptibility during the 2016 M_w 7.8 Kaikōura earthquake, New Zealand. *Natural Hazards and Earth System Sciences*, *23*(9), 2987–3013. https://doi.org/10.5194/nhess-23-2987-2023

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

Bodeker, G., Cullen, N., Katurgi, M., McDonald, A., Morgenstern, O., Noone, D., Renwick, J., Revell, L., & Tait, A. (2022). *Aotearoa New Zealand climate change projections guidance: Interpreting the latest IPCC WG1 report findings* (Report number CR 501; p. 51). Prepared for the Ministry for the Environment.

Bordoni, M., Vivaldi, V., Lucchelli, L., Ciabatta, L., Brocca, L., Galve, J. P., & Meisina, C. (2021). Development of a data-driven model for spatial and temporal shallow landslide probability of occurrence at catchment scale. *Landslides*, *18*(4), 1209–1229. https://doi.org/10.1007/s10346-020-01592-3

Carey-Smith, T., Henderson, R., & Singh, S. (2018). *High Intensity Rainfall Design System Version 4* (No. 2018022CH; p. 73). https://niwa.co.nz/sites/default/files/2018022CH_HIRDSv4_Final.pdf

Chen, W., Guo, C., Lin, F., Zhao, R., Li, T., Tsangaratos, P., & Ilia, I. (2024). Exploring advanced machine learning techniques for landslide susceptibility mapping in Yanchuan County, China. *Earth Science Informatics*, *17*(6), 5385–5402. https://doi.org/10.1007/s12145-024-01455-8

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., Duan, Z., & Ma, J. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *CATENA*, *151*, 147–160. https://doi.org/10.1016/j.catena.2016.11.032

Claessens, L., Schoorl, J. M., & Veldkamp, A. (2007). Modelling the location of shallow landslides and their effects on landscape dynamics in large watersheds: An application for Northern New Zealand. *Geomorphology*, *87*(1–2), 16–27. https://doi.org/10.1016/j.geomorph.2006.06.039

Columbus, J., Sirguey, P., & Tenzer, R. (2011). A free fully assessed 15 metre digital elevation model for New Zealand. *Survey Quarterly*, *66*, 16–19.

Couronné, R., Probst, P., & Boulesteix, A.-L. (2018). Random forest versus logistic regression: A large-scale benchmark experiment. *BMC Bioinformatics*, *19*(1), 270. https://doi.org/10.1186/s12859-018-2264-5

Crozier, M. J. (1999). Prediction of rainfall-triggered landslides: A test of the Antecedent Water Status Model. *Earth Surface Processes and Landforms*, *24*(9), 825–833. https://doi.org/10.1002/(SICI)1096-9837(199908)24:9%253C825::AID-ESP14%253E3.0.CO;2-M

Crozier, M. J. (2010a). Deciphering the effect of climate change on landslide activity: A review. *Geomorphology*, *124*(3–4), 260–267. https://doi.org/10.1016/j.geomorph.2010.04.009

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

Crozier, M. J. (2010b). Landslide geomorphology: An argument for recognition, with examples from New Zealand. *Geomorphology*, *120*(1–2), 3–15. https://doi.org/10.1016/j.geomorph.2009.09.010

De Rose, R. C. (2013). Slope control on the frequency distribution of shallow landslides and associated soil properties, North Island, New Zealand. *Earth Surface Processes and Landforms*, *38*(4), 356–371. https://doi.org/10.1002/esp.3283

De Sy, V., Schoorl, J. M., Keesstra, S. D., Jones, K. E., & Claessens, L. (2013). Landslide model performance in a high resolution small-scale landscape. *Geomorphology*, *190*, 73–81. https://doi.org/10.1016/j.geomorph.2013.02.012

Dragonfly Data Science. (2023). *Cyclone Gabrielle Impact Map* [Dataset]. https://swa-impactmap.dragonfly.co.nz/

Dymond, J. R., Ausseil, A.-G., Shepherd, J. D., & Buettner, L. (2006). Validation of a region-wide model of landslide susceptibility in the Manawatu–Wanganui region of New Zealand. *Geomorphology*, *74*(1–4), 70–79. https://doi.org/10.1016/j.geomorph.2005.08.005

Dymond, J., & Shepherd, J. (2023). *Update to Highly Erodible Land and Estimated Long-term Soil Erosion data sets for Environmental Reporting* (Contract Report: LC4365). Manaaki Whenua – Landcare Research. https://environment.govt.nz/publications/update-to-highly-erodible-land-and-estimated-long-term-soil-erosion-data-sets-for-environmental-reporting

Ehsan, M., Anees, M. T., Bakar, A. F. B. A., & Ahmed, A. (2025). A review of geological and triggering factors influencing landslide susceptibility: Artificial intelligence-based trends in mapping and prediction. *International Journal of Environmental Science and Technology*. https://doi.org/10.1007/s13762-025-06741-6

Fuller, I. C., Riedler, R. A., Bell, R., Marden, M., & Glade, T. (2016). Landslide-driven erosion and slope–channel coupling in steep, forested terrain, Ruahine Ranges, New Zealand, 1946–2011. *CATENA*, *142*, 252–268. https://doi.org/10.1016/j.catena.2016.03.019

Fumagalli, M., Previati, A., Frattini, P., & Crosta, G. B. (2025). Shaping shallow landslide susceptibility as a function of rainfall events. *Natural Hazards and Earth System Sciences*, *25*(11), 4405–4422. https://doi.org/10.5194/nhess-25-4405-2025

Gariano, S. L., & Guzzetti, F. (2016). Landslides in a changing climate. *Earth-Science Reviews*, *162*, 227–252. https://doi.org/10.1016/j.earscirev.2016.08.011

GNS Science. (2019). *Depth to hydrogeological basement* [Dataset]. GNS Science, Ministry for the Environment. https://doi.org/10.21420/FQXD-VY44

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

Griffiths, J. W., Lukens, C. E., & May, R. (2020). Increased forest cover and limits on clear-felling could substantially reduce landslide occurrence in Tasman, New Zealand. *New Zealand Journal of Forestry Science*, *50*. https://doi.org/10.33494/nzjfs502020x94x

Gu, T., Duan, P., Wang, M., Li, J., & Zhang, Y. (2024). Effects of non-landslide sampling strategies on machine learning models in landslide susceptibility mapping. *Scientific Reports*, *14*(1), 7201. https://doi.org/10.1038/s41598-024-57964-5

Hales, T. C., & Roering, J. J. (2005). Climate-controlled variations in scree production, Southern Alps, New Zealand. *Geology*, *33*(9), 701. https://doi.org/10.1130/G21528.1

Han, Y., & Semnani, S. J. (2025). Important considerations in machine learning-based landslide susceptibility assessment under future climate conditions. *Acta Geotechnica*, *20*(1), 475–500. https://doi.org/10.1007/s11440-024-02363-3

Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., & Neal, J. (2022). A 30 m global map of elevation with forests and buildings removed. *Environmental Research Letters*, *17*(2), 024016. https://doi.org/10.1088/1748-9326/ac4d4f

Heron. (2023). *Geological map of New Zealand 1:250,000. 4th ed* [Dataset]. GNS Science. https://doi.org/10.21420/5XTJ-5718

Heskes, T., Sijben, E., Bucur, I. G., & Claassen, T. (2020). *Causal Shapley Values: Exploiting Causal Knowledge to Explain Individual Predictions of Complex Models* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2011.01625

Howard, M. E., Hunter, A. J., Roberts, R. C., & Brook, M. S. (2025). Landslide hazard and loss-of-life risk assessment for Muriwai, New Zealand following Cyclone Gabrielle in February 2023. *New Zealand Journal of Geology and Geophysics*, 1–18. https://doi.org/10.1080/00288306.2025.2470433

Hufschmidt, G., & Crozier, M. J. (2008). Evolution of natural risk: Analysing changing landslide hazard in Wellington, Aotearoa/New Zealand. *Natural Hazards*, *45*(2), 255–276. https://doi.org/10.1007/s11069-007-9158-6

Kattenborn, T., Schiefer, F., Frey, J., Feilhauer, H., Mahecha, M. D., & Dormann, C. F. (2022). Spatially autocorrelated training and validation samples inflate performance assessment of convolutional neural networks. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, *5*, 100018. https://doi.org/10.1016/j.ophoto.2022.100018

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

Koldasbayeva, D., Tregubova, P., Gasanov, M., Zaytsev, A., Petrovskaia, A., & Burnaev, E. (2024). Challenges in data-driven geospatial modeling for environmental research and practice. *Nature Communications*, *15*(1). https://doi.org/10.1038/s41467-024-55240-8

Korup, O. (2006). Effects of large deep-seated landslides on hillslope morphology, western Southern Alps, New Zealand. *Journal of Geophysical Research: Earth Surface*, *111*(F1), 2004JF000242. https://doi.org/10.1029/2004JF000242

Leonarduzzi, E., McArdell, B. W., & Molnar, P. (2021). Rainfall-induced shallow landslides and soil wetness: Comparison of physically based and probabilistic predictions. *Hydrology and Earth System Sciences*, *25*(11), 5937–5950. https://doi.org/10.5194/hess-25-5937-2021

LINZ. (2025a). *NZ Building Outlines* [Dataset]. https://data.linz.govt.nz/layer/101290-nz-building-outlines/

LINZ. (2025b). *NZ Road Centrelines (Topo, 1:50k)* [Dataset]. https://data.linz.govt.nz/layer/50329-nz-road-centrelines-topo-150k/

Lundberg, S., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing System*, 4768–4777. https://doi.org/10.48550/ARXIV.1705.07874

Massey, C., Leith, K., Robinson, T. R., Lukovic, B., McColl, S., Carey-Smith, T., Rosser, B., Wotherspoon, L., Smith, H., Betts, H., Buxton, R., & Bidmead, J. (2025). What controlled the occurrence of more than 116,000 human-mapped landslides triggered by Cyclone Gabrielle, New Zealand? *Landslides*. https://doi.org/10.1007/s10346-025-02591-y

McCarthy, J., Leathwick, J., Roudier, P., Barringer, J., Etherington, T., Morgan, F., Odgers, N., Price, R., Wiser, S., & Richardson, S. (2021). New Zealand Environmental Data Stack (NZEnvDS): A standardised collection of spatial layers for environmental modelling and site characterisation. *New Zealand Journal of Ecology*. https://doi.org/10.20417/nzjecol.45.31

McMillan, A., Dymond, J., Jolly, B., Shepherd, J., & Sutherland, A. (2023). *Rapid Assessment of Land Damage – Cyclone Gabrielle* (Contract Report: LC4292). Manaaki Whenua – Landcare Research. https://environment.govt.nz/assets/Rapid-assessment-of-land-damage-Cyclone-Gabrielle-Manaaki-Whenua-Landcare-Research-report.pdf

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

Ministry for the Environment. (2018). *Climate Change Projections for New Zealand: Atmospheric Projections Based on Simulations Undertaken from the IPCC 5th Assessment* (2nd edition). Ministry for the Environment Manatū Mō Te Taiao.

Mondini, A. C., Guzzetti, F., & Melillo, M. (2023). Deep learning forecast of rainfall-induced shallow landslides. *Nature Communications*, *14*(1), 2466. https://doi.org/10.1038/s41467-023-38135-y

Neal, J., & Hawker, L. (2023). *FABDEM V1-2* [Application/zip,text/plain; charset=windows-1252,application/pdf,text/plain; charset=UTF-8,text/plain; charset=ISO-8859-1]. University of Bristol. https://doi.org/10.5523/BRIS.S5HQMJCDJ8YO2IBZI9B4EW3SN

Neverman, A. J., Donovan, M., Smith, H. G., Ausseil, A.-G., & Zammit, C. (2023). Climate change impacts on erosion and suspended sediment loads in New Zealand. *Geomorphology*, *427*, 108607. https://doi.org/10.1016/j.geomorph.2023.108607

Newsome, P. F. J., Wilde, R. H., & Willoughby, E. J. (2008). *Land Resource Information System spatial data layers: Data dictionary*. https://doi.org/10.7931/DL1V88

Ng, S. K., Sirguey, P., & Redpath, T. (2025). Performance of global canopy height models across varied New Zealand vegetation types. *New Zealand Journal of Forestry Science*, *55*. https://doi.org/10.33494/nzjfs552025x446x

Ngo, T. Q., Dam, N. D., Al-Ansari, N., Amiri, M., Phong, T. V., Prakash, I., Le, H. V., Nguyen, H. B. T., & Pham, B. T. (2021). Landslide susceptibility mapping using single machine learning models: A case study from Pithoragarh District, India. *Advances in Civil Engineering*, *2021*(1), 9934732. https://doi.org/10.1155/2021/9934732

NIWA. (2025). *Updated National Climate Projections for New Zealand*. NIWA. https://niwa.co.nz/climate-and-weather/updated-national-climate-projections-new-zealand

O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., van Ruijven, B. J., van Vuuren, D. P., Birkmann, J., Kok, K., Levy, M., & Solecki, W. (2017). The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change*, *42*, 169–180. https://doi.org/10.1016/j.gloenvcha.2015.01.004

Phillips, C., Betts, H., Smith, H. G., & Tsyplenkov, A. (2024). Exploring the post-harvest 'window of vulnerability' to landslides in New Zealand steepland plantation forests. *Ecological Engineering*, *206*, 107300. https://doi.org/10.1016/j.ecoleng.2024.107300

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

Phillips, C., Hales, T., Smith, H., & Basher, L. (2021). Shallow landslides and vegetation at the catchment scale: A perspective. *Ecological Engineering*, *173*, 106436. https://doi.org/10.1016/j.ecoleng.2021.106436

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C. E., Armston, J., Dubayah, R., Blair, J. B., & Hofton, M. (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, *253*, 112165. https://doi.org/10.1016/j.rse.2020.112165

Pozzolo, A. D., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). Calibrating Probability with Undersampling for Unbalanced Classification. *2015 IEEE Symposium Series on Computational Intelligence*, 159–166. https://doi.org/10.1109/SSCI.2015.33

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., Kc, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., … Tavoni, M. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, *42*, 153–168. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*. arXiv. https://doi.org/10.48550/ARXIV.1602.04938

Rosser, B., Massey, C., Lukovic, B., Dellow, S., & Hill, M. (2021). Development of a Rainfall-Induced Landslide Forecast Tool for New Zealand. In N. Casagli, V. Tofani, K. Sassa, P. T. Bobrowsky, & K. Takara (Eds.), *Understanding and Reducing Landslide Disaster Risk* (pp. 273–277). Springer International Publishing. https://doi.org/10.1007/978-3-030-60311-3_32

Saerens, M., Latinne, P., & Decaestecker, C. (2002). Adjusting the outputs of a classifier to new a *Priori* probabilities: A simple procedure. *Neural Computation*, *14*(1), 21–41. https://doi.org/10.1162/089976602753284446

Segoni, S., Tofani, V., Rosi, A., Catani, F., & Casagli, N. (2018). Combination of Rainfall Thresholds and Susceptibility Maps for Dynamic Landslide Hazard Assessment at Regional Scale. *Frontiers in Earth Science*, *6*, 85. https://doi.org/10.3389/feart.2018.00085

Semnani, S. J., Han, Y., Bonfils, C. J., & White, J. A. (2025). Assessing the impact of climate change on rainfall-triggered landslides: A case study in California. *Landslides*. https://doi.org/10.1007/s10346-024-02428-0

https://doi.org/10.5194/egusphere-2026-768
Preprint. Discussion started: 17 February 2026

Shu, H., He, J., Zhang, F., Zhang, M., Ma, J., Chen, Y., & Yang, S. (2024). Construction of landslide warning by combining rainfall threshold and landslide susceptibility in the gully region of the Loess Plateau: A case of Lanzhou City, China. *Journal of Hydrology*, *645*, 132148. https://doi.org/10.1016/j.jhydrol.2024.132148

Smith, H. G., Spiekermann, R., Betts, H., & Neverman, A. J. (2021). Comparing methods of landslide data acquisition and susceptibility modelling: Examples from New Zealand. *Geomorphology*, *381*, 107660. https://doi.org/10.1016/j.geomorph.2021.107660

Spiekermann, R. I., Smith, H. G., McColl, S., Burkitt, L., & Fuller, I. C. (2022). Quantifying effectiveness of trees for landslide erosion control. *Geomorphology*, *396*, 107993. https://doi.org/10.1016/j.geomorph.2021.107993

Tenzer, R., Sirguey, P., Rattenbury, M., & Nicolson, J. (2011). A digital rock density map of New Zealand. *Computers & Geosciences*, *37*(8), 1181–1191. https://doi.org/10.1016/j.cageo.2010.07.010

Wen, H., Yan, F., Huang, J., & Li, Y. (2025). Interpretable machine learning models and decision-making mechanisms for landslide hazard assessment under different rainfall conditions. *Expert Systems with Applications*, *270*, 126582. https://doi.org/10.1016/j.eswa.2025.126582

Westerhoff, R. S., Tschritter, C., & Rawlinson, Z. J. (2019). *New Zealand Groundwater Atlas: Depth to Hydrogeological Basement* (Consultancy Report No. 2019/140; p. 19). GNS Science.

Wigmore, O. (2026). *National Rainfall-Induced Landslide Susceptibility Maps (25 m) for Extreme Rainfall Scenarios under Present-Day and Future Shared Socioeconomic Pathways (SSPs), New Zealand* (Version 1.0.0) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.17850323

Wilkinson, P. L., Anderson, M. G., & Lloyd, D. M. (2002). An integrated hydrological model for rain-induced landslide prediction. *Earth Surface Processes and Landforms*, *27*(12), 1285–1297. https://doi.org/10.1002/esp.409

Wilson, N., Broadbent, A., & Kerr, J. (2023). *Cyclone Gabrielle by the numbers – A review at six months* (The Briefing). Public Health Communication Centre Aotearoa. https://www.phcc.org.nz/briefing/cyclone-gabrielle-numbers-review-six-months

Yin, H., Zhang, J., Mondal, S. K., Wang, B., Zhou, L., Wang, L., & Lin, Q. (2023). Projected rainfall triggered landslide susceptibility changes in the Hengduan mountain region, Southwest China under 1.5–4.0 °C warming scenarios based on CMIP6 models. *Atmosphere*, *14*(2), 214. https://doi.org/10.3390/atmos14020214