https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

Machine learning for automated avalanche terrain exposure scale (ATES) classification

Kalin Markov[1,5], Andreas Huber[2,6], Momchil Panayotov[3,4,5], Christoph Hesselbach[2], Paula Spannring[2], Jan-Thomas Fischer[2], and Michaela Teich[2]

[1]National Institute of Geophysics, Geodesy and Geography, Bulgarian Academy of Sciences, Sofia, Bulgaria
[2]Austrian Research Centre for Forests (BFW), Department of Natural Hazards, Innsbruck, Tyrol, Austria
[3]University of Forestry, Sofia, Bulgaria
[4]Bulgarian Extreme and FreeSkiing Association (BEFSA), Sofia, Bulgaria
[5]Bulgarian Avalanche Association, Sofia, Bulgaria
[6]University of Innsbruck, Unit of Hydraulic Engineering, Innsbruck, Tyrol, Austria

Correspondence: Kalin Markov (kmarkov@geophys.bas.bg)

Abstract. Avalanche risk management is essential for backcountry safety. The Avalanche Terrain Exposure Scale (ATES) classifies mountain terrain based on its potential exposure to avalanche hazards and offers assistance to backcountry users in their terrain assessment. Initially, ATES maps were generated manually, a costly and time-consuming process. Automated ATES model chains (AutoATES) have been developed to address these limitations, but existing approaches require careful parametrisation when applied to novel areas.

This study applies machine learning methods, specifically Random Forests, for automated ATES classification by replacing expert-driven AutoATES classification trees with a data-driven approach. Using a labelled training dataset from the Pirin Mountains, Bulgaria, we trained and evaluated three Random Forest models to assess their potential in classifying avalanche terrain. We analysed the influence of various input features, including slope, potential release areas, and percent canopy cover, on classification performance. Our results indicate that Random Forests offer a robust and scalable method for ATES mapping and that incorporating additional input features can improve classification performance. The accuracies for our Random Forest models on a held-out test set were 79.31 %, 82.32 %, and 80.42 %, demonstrating their potential for automated avalanche terrain classification and supporting safer backcountry decision-making.

1 Introduction

Globally, around 250 fatalities are caused by snow avalanches every year (Acharya et al., 2023), with more than one third of annual fatalities occurring on average in the European Alps (Techel et al., 2016). The amount of non-fatal incidents significantly exceeds these numbers. A large proportion of fatal avalanche accidents involves people recreating in the mountains, particularly skiers and snowboarders, often triggering the avalanches themselves (Schweizer and Lüschg, 2001; Techel and Zweifel, 2013; Engeset et al., 2018). These numbers clearly highlight the importance of supporting recreationists' decision making in avalanche terrain by raising awareness, enhancing education, and providing information on avalanche terrain and current conditions (Toft et al., 2024).

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

In Bulgaria, backcountry skiing and snowboarding have seen a notable increase in popularity in recent years. Unfortunately, this increase has been accompanied by a growing number of avalanche-related incidents (Panayotov et al., 2021). Despite these developments, there is still limited information available to the public regarding current snowpack stability and avalanche danger or the spatial distribution of avalanche-prone terrain. Unlike in most alpine countries no government-supported avalanche forecast exists. The only available forecast is provided by a non-governmental organization (Bulgarian Avalanche Association), and is limited to the Bansko Ski Resort area (Panayotov et al., 2024b). As a result, backcountry recreationists must rely primarily on their own judgment when evaluating current avalanche danger and navigating potential avalanche terrain. Therefore, there is a pressing need for clear and accessible information to enhance public awareness of avalanche-prone terrain and current snowpack stability, especially aimed at backcountry users with limited avalanche education and experience in terrain assessment and management (Panayotov et al., 2021).

A simple and effective method for mapping avalanche-prone terrain and communicating it to backcountry users is the Avalanche Terrain Exposure Scale (ATES). Originally introduced by Statham et al. (2006) as a visitor management tool for national parks in Canada, ATES was developed to assess and convey the inherent avalanche risk of snow-covered mountainous terrain. Unlike avalanche forecasts, which provide dynamic assessments of current snowpack stability and avalanche danger, ATES offers a static classification that reflects the terrain's inherent avalanche hazard potential, independent of current snow or weather conditions. The initial version categorised terrain into the three classes, *simple*, *challenging*, and *complex* (Statham et al., 2006). A later revision expanded the system to five classes, adding an *extreme* category and a designation for *non-avalanche terrain* (Statham and Campbell, 2025).

The ATES framework has been applied to map linear terrain routes (Parks Canada Agency, 2017) as well as to create comprehensive spatial maps. The first of these were developed by expert-based manual delineation. Examples include the mapping of more than $8\,000\,\mathrm{km}^2$ of avalanche terrain in western Canada (Campbell and Gould, 2013) or the production of ATES maps for a part of the Central Pyrenees (Gavaldã et al., 2013). In Bulgaria, geospatial technologies were first employed for avalanche terrain hazard mapping by Markov and Ivanov (2021), while the country's first ATES maps were also produced through expert assessment and manual GIS-based classification of an area near Bansko Ski Resort (Panayotov et al., 2021).

In recent years, significant efforts have been dedicated to developing automated procedures for generating ATES maps (e.g. AutoATES, Larsen et al., 2020) or similar avalanche terrain classification approaches (Schmudlach and Köhler, 2016; Harvey et al., 2018, 2024). The first iteration of AutoATES was introduced in Norway by Larsen et al. (2020), outlining a structured methodology consisting of three key steps: (1) identifying potential avalanche release areas (PRAs), (2) modelling their potential runouts, and (3) integrating this information into a final ATES map using an expert-based classification scheme. A major limitation of the initial version was that it did not account for the influence of forest cover on potential avalanche hazard. Schumacher et al. (2022) demonstrated that incorporating forest data can substantially enhance the accuracy of ATES classification compared to approaches that omit this information.

To address this shortcoming and improve the general classification procedure, AutoATES 2.0 was developed by Toft et al. (2024), incorporating forest cover data to account for its protective effect against avalanches. AutoATES version 2.0 made use of ATES 2.0 (Statham and Campbell, 2025), which adds the *extreme* and *non-avalanche terrain* classes. Additionally,

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

this version leveraged the open-source software Flow-Py (D'Amboise et al., 2022) to enhance avalanche intensity and runout modelling. AutoATES 2.0 has since been applied to different mountainous regions worldwide. For instance, Avis et al. (2023) employed AutoATES to map avalanche terrain in the continental United States, while Huber et al. (2023) used a modified version to assess avalanche exposure in the Austrian Alps (AutoATES Austria).

The first use of automated ATES mapping in Bulgaria was carried out by Markov and Panayotov (2024), who based their work on the AutoATES 2.0 workflow. Subsequently, the methodology outlined in AutoATES Austria was applied to generate updated ATES maps for three test regions. While this approach proved effective in open alpine environments above the treeline, it faced challenges in accurately delineating avalanche terrain within forested areas, necessitating modifications of the expert-driven ATES classification tree and the creation of new algorithms to more precisely estimate forest percent canopy cover (Panayotov et al., 2024b). Furthermore, AutoATES relies on numerous thresholds and parameters, which must be carefully chosen and adjusted to local conditions. Identifying the optimal set of input features and thresholds for the classification is highly dependent on the specific characteristics of the terrain. As a result, expert knowledge is required to fine-tune all aspects of AutoATES, including PRA identification, runout approximation, and the final expert-driven ATES classification tree (Hesselbach, 2023). A method for optimizing parameters in the AutoATES classification step by reverse engineering human ATES maps was recently proposed by Sykes et al. (2024), highlighting the importance of refining threshold values for local conditions.

Applying the AutoATES procedure to novel mountain areas requires in-depth understanding of the AutoATES model chain itself as well as intimate knowledge of the mountainous terrain being analyzed. This process can be challenging and time-consuming, as it demands a deep understanding of avalanche dynamics and the intricate dependencies between input parameters. Given these challenges, an ideal approach would be to create an automated ATES mapping procedure that integrates expert knowledge on local terrain and mountain characteristics, without having to rely heavily on manual parameter tuning and threshold adjustments in the AutoATES procedure.

Machine learning (ML) presents a promising method to address these challenges. ML techniques have been widely employed across various domains to harness computational power and enhance human decision-making (Pugliese et al., 2021; Sarker, 2021). Applications range from pattern recognition, image and speech processing, and medical diagnostics to predicting the likelihood of natural disasters and other complex events (Pandey et al., 2025; Gowda V et al., 2024; Ahsan et al., 2024; Venkadesh et al., 2024).

A specialized subset of ML, known as supervised learning, uses a labelled training dataset – expert-annotated examples – from which algorithms learn the relationships and interactions among different input features (Nasteski, 2017). This approach has proven particularly effective in classification tasks, where the goal is to assign data points to predefined categories (Saraswat, 2022; Amin et al., 2024). Since ATES mapping inherently falls into this category, ML-based classifiers present a promising alternative to traditional rule-based approaches.

Several categories of classification algorithms exist that leverage labelled training data, including Naïve Bayes (NB), Random Forest (RF), and Neural Networks (NN) (Yolanda et al., 2023; Ul Hassan et al., 2018). Among these, RF methods have been successfully applied to a wide range of classification problems, including in the field of avalanche hazard assessment

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

(Cetinkaya and Kocaman, 2023; Viallon-Galinier et al., 2023), due to their ability to process diverse input features and provide feature importance rankings (Jaiswal and Samikannu, 2017). RF models can directly utilise raw, continuous input values, eliminating the need for expert-driven rule-based manual classification and enabling the model to capture more complex patterns and interactions within the data (Salman et al., 2024). These capabilities are particularly valuable for ATES mapping, as they allow researchers to determine which factors most strongly influence terrain classification.

To address the above mentioned drawbacks of previous AutoATES approaches, in this study we tested the application of the Random Forest algorithm for final ATES classification. Our research addressed three key questions:

1. Can the Random Forest algorithm effectively perform the final classification required for automated ATES mapping, generating maps with terrain categorised into ATES classes?
2. Can Random Forests serve as a tool for assessing the relevance and utility of different input features in automated avalanche terrain classification?
3. How can we systematically evaluate the performance of automated ATES classification using different Random Forest models?

These research questions served as the basis for our methodological design and shaped the structure of our study into three main parts.

In the first part of our study, we replaced the expert-driven ATES classification tree, which relies on predefined rules and thresholds, with a Random Forest-based classifier trained on a manually labelled dataset created by a local expert.

In the second part, we assessed the significance of different input features. By systematically introducing additional features, we analysed their impact on Random Forest model performance and evaluated their suitability for ATES classification.

Finally, we conducted a comprehensive evaluation of the Random Forest models. Standard performance metrics were computed on a held-out test set, ensuring robust validation. This evaluation provided an estimate of how the models would generalise to unseen terrain.

2 Data and methods

2.1 Study area

Our study was centered on the Pirin Mountain region in Bulgaria. Pirin, situated in southwestern Bulgaria, is the country's second-highest mountain range, with its highest peak, Vihren, rising to $2914\,\mathrm{m}$ a.s.l. The northern part of the range contains over 50 peaks exceeding $2500\,\mathrm{m}$ a.s.l. and is characterized by a rugged alpine landscape featuring steep, prominent peaks, sharp ridges, and deep glacially-carved valleys dotted with numerous lakes (PirinNationalPark, 2022).

The study area covered three adjacent valleys in northern Pirin – Bunderitsa, Demyanitsa, and Bezbog – spanning approximately $77\,\mathrm{km}^2$. These valleys are popular backcountry ski-touring destinations. Elevations within this area range from the

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

highest point at Vihren Peak (2914 m a.s.l.) to the lowest at 1462 m a.s.l. The valleys are characterized by considerable elevation changes, with some sections exhibiting vertical drops of up to 1000 m. While the lower portions of the study area are predominantly forested, elevations between 2200 and 2500 m a.s.l. are often densely covered by dwarf mountain pine (*Pinus mugo*). The highest elevations primarily consist of rocky terrain, characterised by extensive rock bands and cliff faces.

The northwestern section of the study area includes the region around Bansko Ski Resort and the Bunderitsa Valley. This is the area where all training data polygons were manually delineated (see Sect. 2.3). The trained models were then run on the entire study area, which includes the Demyanitsa and Bezbog valleys to the east.

Figure 1 provides an overview of the study area and its location on the Balkan peninsula and also indicates the location of the training area.

2.2 General input data

The essential input data for generating ATES maps include a digital terrain model (DTM) and at least one type of forest data, which can be chosen from several available options. In AutoATES 2.0, Toft et al. (2024) introduced three possible forest data types that can be used - percent canopy cover, stem density, and basal area.

For the majority of the study area, we used a 5 m resolution DTM provided by Pirin National Park for the purposes of this study (PirinNationalPark, 2022). Additionally, for a smaller section of the Bunderitsa Valley, a 2 m DTM provided by Geopolymorphic Cloud, a private geospatial data provider, was utilised. These two datasets were merged using open-source GIS software (QGIS) and resampled to a 10 m resolution. A 10 m DTM resolution has been utilised in a number of AutoATES case studies (Larsen et al., 2020; Huber et al., 2023) and coincides with the resolution of the available forest data in our study region. Additional processing of the DTM consisted of a script-based correction of elevations inside lake polygons to ensure even lake surfaces.

Forest structure data in Bulgaria is scarce, and the dataset available from Pirin National Park lacks sufficient detail for this study's purposes. Therefore, we utilised the satellite-based open-source Copernicus tree cover density product with 10 m resolution (Copernicus, 2020). This dataset assigns a percentage value to each pixel that corresponds to varying degrees of forest density. To maintain consistency with terminology used in AutoATES 2.0 (Toft et al., 2024), we referred to this layer as *percent canopy cover* (PCC). While this dataset has known limitations, such as an inability to detect small forest openings and occasional misclassification of dwarf mountain pine as forest (Panayotov et al., 2024b), its open accessibility makes it a useful candidate for evaluating the potential of low-cost, large-scale mapping approaches.

All datasets have been transformed to a common coordinate system (EPSG:32635, WGS 84 / UTM zone 35N).

2.3 Training data

Since RF is a supervised ML algorithm (Nasteski, 2017), it requires a training dataset to learn the complex patterns and dependencies between input features and the corresponding response variable (i.e. the ATES classes). The training dataset needs to contain manually labelled ATES regions, where the assigned ATES classes are treated as the ground truth from which the RF classifier learns. Ensuring a diverse training set is critical, as it must represent a variety of possible terrain features

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

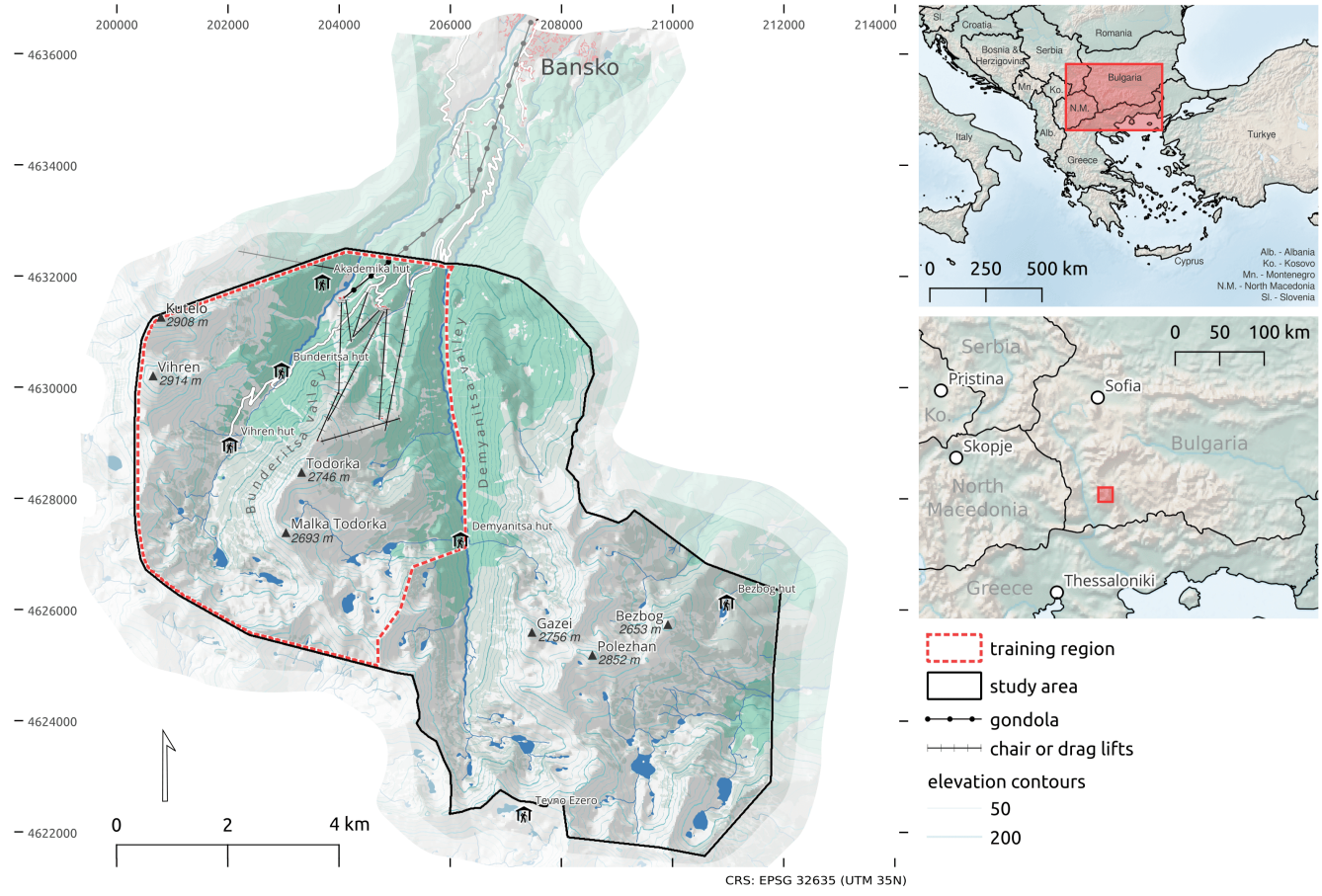

Figure 1. Study area located in the Pirin Mountains, Bulgaria south of Bankso. The western part of the study area covering the Bunderitsa valley, parts of the Demyanitsa valley and parts of Bansko Ski Resort has been used for model training. The study area also comprises parts of the rest of the Demyanitsa Valley and the Bezbog area, on which the models have not been trained. The map utilises the following sources. **(a)** ©OpenStreetMap contributors 2025. Distributed under the Open Data Commons Open Database license (ODbL) v1.0. **(b)** Natural Earth. Free vector and raster map data at https://www.naturalearthdata.com/ (last access: 26 March 2025) **(c)** FAB-DEM (https://doi.org/10.5523/bris.s5hqmjcdj8yo2ibzi9b4ew3sn). Distributed under a Non-Commercial Government Licence for public sector information.

associated with each ATES class. A well-balanced and comprehensive dataset enhances the algorithm's ability to learn the defining characteristics of the different classes (Gong et al., 2023). Additionally, the accuracy of the spatial boundaries in the training dataset is crucial. Poorly defined polygon boundaries, such as one polygon extending into a region belonging to a different ATES class, can introduce ambiguity into the model, reducing its ability to generalise to unseen data. Another key factor is class balance within the training dataset. While RF is robust to class imbalance, maintaining a relatively balanced number of samples per class generally improves model performance (Japkowicz, 2001).

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

Table 1. Distribution of ATES classes in the training set and test set. Numbers correspond to the number of 10 m ×10 m pixels, multiplication with 10^2 yields area in m^2. The relative share of each class in % is provided in brackets.

ATES class	training set	test set	total
Simple	4441 (34.1)	4503 (34.6)	8944 (34.3)
Challenging	3633 (27.9)	3573 (27.4)	7206 (27.7)
Complex	3809 (29.2)	3841 (29.5)	7650 (29.4)
Extreme	1142 (8.8)	1108 (8.5)	2250 (8.6)
Total	13025 (100)	13025 (100)	26050 (100)

In this study, the training data was delineated in cooperation with a local expert for an area covering the Bunderitsa Valley and a small portion of the Demyanitsa Valley near Bansko Ski Resort (see Fig. 1) and was based on ATES class definitions following Statham and Campbell (2025) and Toft et al. (2024). The remainder of the Demyanitsa valley and the Bezbog region were entirely excluded during the training set development. The decision to use only a subset of the study area for training was motivated by the objective of evaluating the models' abilities to generalise beyond the region they were trained on.

The delineation of training data polygons was based on local expertise and records of well-known and observed avalanche release areas and runouts in the region. The mapping process was additionally informed by slope maps, results of dendrogeomorphologic studies (Tsvetanov, 2024), and local-scale avalanche simulations conducted with RAMMS (Panayotov et al., 2024a). The training set was constructed to ensure a diverse representation of different terrain features associated with each ATES class. Consequently, it consisted of small, precisely delineated polygons, each assigned to one of the four ATES classes. Using small polygons for training facilitated labelling regions with high confidence, ensuring accurate classification. In contrast, we found that manually drawing a continuous ATES map across the entire study area is significantly more challenging and may introduce generalisation errors, making it harder for the classifier to learn effectively. Efforts were also made to achieve class balance as much as possible across the *simple*, *challenging*, and *complex* categories. However, the *extreme* class was underrepresented due to the more limited availability of extreme terrain in the study area. Figure 2 illustrates the training data polygons used in this study, as well as the used PCC data for the same region.

The training polygons were subsequently converted to a 10 m raster dataset and aligned with the other raster input-features described in sections 2.2 and 2.4. The dataset was then randomly split into a 50/50 train–test partition, with one half used for model training and the other reserved as a hold-out test set to evaluate performance and generalisation (see Sect. 2.5). Table 1 presents the total number of pixels for each ATES class and the approximate 50/50 split between training and test sets. Figure A3 displays the distributions of some of the most important input features across each of the four ATES classes in the training dataset.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

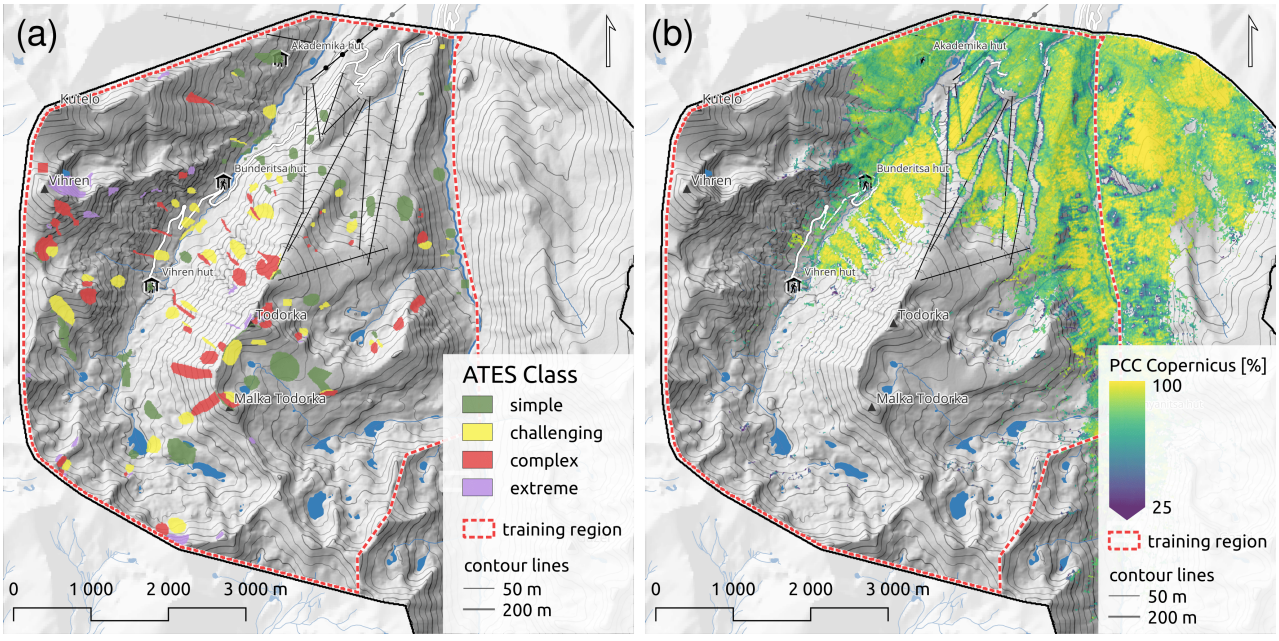

Figure 2. (a) Training data based on local expert assessment of ATES classes for the training area, covering the Bunderitsa Valley and parts of the Demyanitsa Valley. **(b)** Copernicus percent canopy cover (PCC) layer for the training region.

2.4 AutoATES model chain

The basic AutoATES model chain (e.g. Larsen et al., 2020; Huber et al., 2023; Panayotov et al., 2024b; Toft et al., 2024; Sykes et al., 2024) is based on open-source components developed by different research groups. It has been designed to work with limited input data requirements (DTM, forest layer), and involves three main steps. These steps comprise (1) the identification of potential avalanche release areas (PRAs), (2) the delineation of potential avalanche runouts originating from the PRAs, and (3) a final classification and mapping step, which has been referred to as the "AutoATES classifier" in previous studies (Toft et al., 2024; Sykes et al., 2024). This study focused on the application of supervised ML methods in the final AutoATES classification step in place of the previously proposed deterministic, expert-based set of thresholds and rules (Fig. 3). The next sections provide an overview of the setup and parameterisation that have been used in the PRA and avalanche runout modelling steps in this study, and briefly discuss the ”classic” decision tree-based AutoATES classifier, before Sect. 2.5 introduces the modifications to the classifier investigated in this study.

2.4.1 Modelling of potential avalanche release areas (PRAs)

The PRA model used in this study followed the approach outlined by Toft et al. (2024) and Sykes et al. (2024). The original PRA model proposed by Veitinger et al. (2016a) has been modified by Sharp (2018), Schumacher et al. (2022) and Toft et al. (2024) to include forest information and operate on input data with coarser resolution ($\approx$ 10 m as opposed to 2 m used by

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

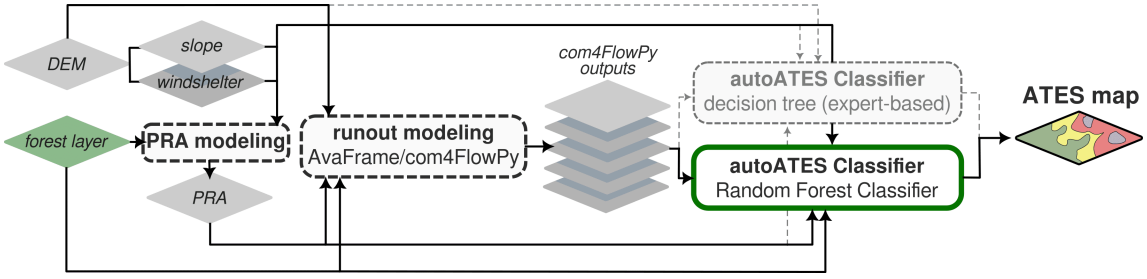

Figure 3. Main components of the AutoATES model chain (represented by rounded rectangles) and model's input, intermediate, and output layers (depicted as diamond shapes). Parts of the model chain that are shared with the approach of Toft et al. (2024) are highlighted with dashed outlines. The introduction of machine learning (ML) techniques in the AutoATES classifier, as investigated in this study, is indicated by a green outline.

Veitinger et al. (2016a)). The model maps a set of input layers (slope, wind shelter, percent canopy cover) to fuzzy membership values μ using three-parameter bell-shaped membership functions with varying parametrisations (Eq. (1)).

$$\mu(x) = \frac{1}{1 + (\frac{x-c}{a})^{2b}} \tag{1}$$

The fuzzy membership values of the three layers are then combined with a "fuzzy AND" operator (Werners, 1988) to a continuous PRA membership value $PRA_{cont} \in [0, \ldots, 1]$, which indicates PRA likelihood (Eq. (2)).

$$PRA_{cont} = \gamma \times min_{1<i\leq n}(\mu_i(x)) + \frac{(1-\gamma)\sum_1^n \mu_i(x)}{n}$$
$$\text{with :} \tag{2}$$
$$\gamma = 1 - min_{1<i\leq n}(\mu_i(x)), \;\; x \in X, \, \gamma \in [0, 1]$$

We used a DTM and the Copernicus PCC layer, both at $10\,\mathrm{m}$ resolution, as inputs for the PRA model. To parametrise the bell-shaped membership functions, we adopted values for slope, wind shelter, and percent canopy cover from Toft et al. (2024), with a slight modification to the Cauchy membership function for slope. We assigned membership values of 0 to areas with slopes $< 28°$ or $> 60°$, in accordance with the original PRA model (Veitinger et al., 2016b). To convert the continuous PRA membership values $PRA_{cont} \in [0, \ldots, 1]$ into a binary PRA layer $PRA_{bin} \in [0|1]$, we applied a threshold value of 0.3. This binarisation is required for the subsequent runout modelling step. The threshold was selected based on experience from previous studies (Hesselbach, 2023; Huber et al., 2023; Panayotov et al., 2024b), site inspections, and local expert knowledge. Figures A1 (a) and (b) present the results of the PRA modelling step (PRA_{cont}, PRA_{bin}) for the terrain surrounding Todorka Peak, which is easily accessible from the top chairlift of the Bansko Ski Resort.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

2.4.2 Modelling of avalanche runout and intensities with com4FlowPy

AutoATES utilises the empirical, raster-based model Flow-Py (D'Amboise et al., 2022) for calculations of avalanche runouts and intensities from the binary PRAs identified in the previous step (Toft et al., 2024; Sykes et al., 2024). Flow-Py has been integrated into the Open Avalanche Framework AvaFrame (Oesterle et al., 2025) in 2024 and is now available as AvaFrame module `com4FlowPy` (Huber et al., 2024). In this study we utilised `com4FlowPy` with its *forest-friction* option enabled (D'Amboise et al., 2021; Huber et al., 2024). The *forest-friction* functionality is used to account for increased energy dissipation in forest areas by increasing the basal friction on forested cells in the model domain. The basal friction increase on forested cells is modulated by a lumped forest structure index (FSI) and modelled avalanche intensities z^{δ} (D'Amboise et al., 2021; Huber et al., 2024). The FSI layer used in this study was obtained by linearly scaling the Copernicus PCC layer (Fig. 2 (b)) to the interval $[0, \ldots, 1]$, where 0 represents no forest and 1 represents dense forest canopy cover. The model parameters used for `com4FlowPy` are listed in Table 2. The choice of α angle, z^{δ}_{limit}, exp and R_{stop} has been based on previous studies (Hesselbach, 2023; Huber et al., 2023; McClung and Gauer, 2018) with a focus on capturing typical runouts of large avalanches (size 3 according to EAWS[1], Huber et al. (2023)), while the parametrisation of the *forest-friction* function was informed by a limited parameter sensitivity study (Huber et al., 2024). Although a rigorous evaluation of the modelled avalanche runouts was outside the scope of this study, good agreement was found between the model results and local records of historical avalanches, as well as dendrochronological investigations in the Bunderitsa area (Tsvetanov, 2024), using the chosen model parameter set.

`Com4FlowPy` computes several model outputs in the form of raster layers (D'Amboise et al., 2022), which can be used to characterise avalanche intensities, proximity to release areas, and overhead exposure (Sykes et al., 2024). Recent studies have used different combinations of output layers from `com4FlowPy` in applications of AutoATES (Schumacher et al., 2022; Toft et al., 2024; Sykes et al., 2024). In this study, we focused on the maximum modelled energy line height z^{δ}_{max}, the maximum modelled local travel angle (γ), *cell counts*, and the sum of modelled fluxes ($\sum routFlux$) (see Fig. A1 panels c-f). The z^{δ}_{max} output layer can be interpreted in terms of potential avalanche intensities and more intuitively be converted to maximum potential avalanche velocities by $v_{max} = (2\,g\,z^{\delta}_{max})^{\frac{1}{2}}$ (Körner, 1980). The maximum local travel angle (γ) provides a measure of the proximity of a potential runout location to the nearest PRA, taking into account both the distance and the average gradient of the avalanche path connecting the two locations (D'Amboise et al., 2022). The *cell counts* layer describes the number of PRA cells potentially affecting a location, while the $\sum routFlux$ layer sums the modelled flux that is routed through a cell (D'Amboise et al., 2022). Both layers provide an estimate of potential overhead hazard, with high values typically characterising gullies and/or overlapping avalanche paths (fig. A1 e, f). While the *cell counts* output focuses on the sizes of overhead PRAs, $\sum routFlux$ provides a more nuanced output, highlighting terrain traps such as gullies and deposit zones.

2.4.3 AutoATES classifier

The final step in the AutoATES model chain – the AutoATES classifier – combines layers derived from preceding stages of the model chain (PRAs, runout and intensity outputs, slope, and forest information) into ordinal ATES ratings. Although

[1] https://www.avalanches.org/standards/avalanche-size/

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

Table 2. Parameters used for modelling of avalanche runouts and intensities with `com4FlowPy`.

Parameters:	α	z_{lim}^{δ}	exp	R_{stop}	$\Delta_{max}^{\alpha,for}$	$\Delta_{min}^{\alpha,for}$	v_{for}^{lim}
used value:	26°	185 m ($\approx$ 60 m s^{-1})	8	3×10^{-4}	21°	0°	45 m s^{-1}

the specific configuration of input layers employed by the AutoATES classifier varies slightly across different versions of AutoATES (Larsen et al., 2020; Schumacher et al., 2022; Toft et al., 2024) and its local applications (von Avis et al., 2023; Huber et al., 2023; Panayotov et al., 2024b), the underlying classification methodology has remained largely consistent since the release of AutoATES v1.0 (Larsen et al., 2020).

In summary, existing AutoATES classifiers are based on a deterministic framework that applies expert-defined thresholds and rules to derive ATES ratings from a set of input layers. This rule-based approach offers the advantage of being transparent and interpretable from a human perspective, and also mirrors the general structure of both the original and updated ATES technical models suggested for manual ATES mapping (Statham et al., 2006; Statham and Campbell, 2025).

However, to eliminate the need for parameter optimisation (Sykes et al., 2024) or rule modification (Panayotov et al., 2024b) when applying AutoATES to new regions, this study tested an alternative classification approach based on supervised machine learning.

2.5 Random Forests

2.5.1 Basic principles

Random Forests (RF) is a supervised machine learning algorithm based on ensemble learning, designed for both classification and regression tasks (Breiman, 2001). In the context of ATES class prediction, the model operates on a per-pixel basis, assigning each pixel one of four possible terrain classes: *simple*, *challenging*, *complex*, or *extreme*. The algorithm constructs multiple decision trees during training, linking input features that describe the characteristics of each pixel to its manually labelled ATES class. Each individual tree generates a class prediction, and for a new, unseen data point, the Random Forest aggregates predictions from all trees, assigning the most frequent class as the final output. This ensemble approach reduces overfitting compared to a single decision tree, enhancing model robustness and generalisation performance (Breiman, 2001).

Each tree in the RF is trained using either a bootstrap sample (random sampling from the training dataset with replacement) or the entire dataset, as controlled by the *bootstrap* parameter. During tree construction, the algorithm determines the optimal split at each node by selecting a feature and a threshold that minimises impurity, commonly measured using Gini impurity and information gain (Prasetiyowati et al., 2021). Impurity measures quantify the degree of class heterogeneity within a dataset. A node containing only samples from a single ATES class has zero impurity, whereas a node with an evenly distributed mix of classes exhibits higher impurity. The goal of decision tree construction within the RF is to iteratively partition the data so that the resulting subsets (leaf nodes) are as pure as possible. This process is guided by information gain, which quantifies the reduction in impurity achieved by a particular split. At each decision node, the algorithm evaluates a subset of input features

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

(determined by the *max_features* parameter) and identifies the feature and threshold that maximises information gain. The search space for threshold values is usually restricted to midpoints between unique observed values within the selected feature, ensuring computational feasibility while optimising for split quality (Salman et al., 2024).

The RF algorithm has several hyperparameters that control the training process. While in-depth optimisation of hyperparameters is beyond of the scope of this study, a basic evaluation on how tuning a few key parameters affects model performance was conducted (see Sect. 2.5.2).

2.5.2 Application of Random Forest for ATES mapping

Following the procedure outlined in Sect. 2.4, a series of raster layers were generated in the PRA and avalanche modelling steps. Together with the general input data layers (Sect. 2.2), these layers served as input features for three different RF models investigated in this study – RF1, RF2, and RF3.

For all models, both the training labels and the input features were stored in georeferenced TIFF files with identical dimensions and pixel resolutions. Each input feature was represented as a separate raster band within the input feature files. The implementation was realised in Python 3.9.21 (Python Software Foundation), leveraging libraries such as GDAL and NumPy for data processing, while scikit-learn (Pedregosa et al., 2011) was used to define and train the RF classifiers.

The scikit-learn library provides extensive control over the model's hyperparameters, enabling their optimisation to enhance generalisation and predictive performance. Hyperparameter tuning is a widely researched topic, with numerous studies dedicated to optimising model configurations (Rimal et al., 2024; Bischl et al., 2023; Ramadhan et al., 2017; G and B.Sumathi, 2020; Probst et al., 2019; Contreras et al., 2021; Wang et al., 2021). However, previous research has also demonstrated that the default hyperparameter settings of the RF algorithm generally yield strong predictive performance (Probst et al., 2019). In this study, we adopted the approach of Contreras et al. (2021), using their predefined parameter grid with minor modifications. We also added the *bootstrap* parameter to the hyperparameter search together with *max_samples*. If *bootstrap* is set to *false*, the entire training dataset is used to build each tree. If it is set to *true*, then *max_samples* (as a fraction from all samples) training points are taken at random, with replacement, from the full training set. Finally, we also introduced the *class_weight* parameter, which can help training when classes are imbalanced, as was the case with the *extreme* class in our training set. The parameters *max_depth*, *max_leaf_nodes*, *min_samples_leaf*, and *min_samples_split* together control how much the trees grow, preventing overfitting to the training dataset and ensuring the model generalises better to new data.

The complete hyperparameter grid, along with a brief description of each parameter, is presented in Table 3. All unspecified parameters remain at their default values. Given computational constraints, we opted for a random search strategy rather than an exhaustive grid search, which would otherwise result in an infeasible number of combinations. Specifically, we sampled $10\,000$ hyperparameter combinations at random from the defined grid.

To evaluate model performance, the training dataset was randomly partitioned into training ($50\,\%$) and test ($50\,\%$) sets, ensuring that the test set remains entirely separate from the training process. Additionally, we employed k-fold cross-validation with $k = 3$ (Varoquaux and Colliot, 2023), wherein the training set was further subdivided into three equal partitions (folds). Each fold alternates between training and validation, with two folds used for training while the third serves as the validation

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

Table 3. Hyperparameter setup for training the Random Forest models (RF1-3)

Hyperparameter	Values tested	Description	Best values per model		
	[val 1, val 2, ...] and/or [min - max, interval]		RF1	RF2	RF3
max_depth	[5 - 70, 5]	max depth of each tree	10	45	70
max_features	[None, sqrt, log2], [0.1 - 1, 0.1]	features considered per split	0.8	0.8	log2
max_leaf_nodes	[5 - 50, 5]	max terminal nodes per tree	50	50	50
max_samples	[0.1 - 1, 0.1]	fraction of samples to use per tree	0.7	0.9	1.0
min_samples_leaf	[1, 5, 10, 15, 20]	min samples required per leaf node	1	1	5
min_samples_split	[2], [5 - 40, 5]	min samples needed to further split node	5	2	20
n_estimators	[10 - 100, 10], [200 - 1000, 100]	number of trees in the forest	100	60	800
bootstrap	[True, False]	whether to sample with replacement	True	True	True
class_weight	[None, balanced_subsample]	adjusts class imbalance	None	None	None

Decimal values or functions should be interpreted as that fraction of the total number. For example, for the `max_features` *hyperparameter,* `sqrt` *means try the square root of the total number of features as the value, while* 0.1 *means try* 10 % *of the total number of features.* `None` *means that there is no limit or a parameter is not used.*

set. This process is repeated across all fold combinations, allowing the model to select the hyperparameter configuration that maximizes accuracy on the validation folds. The best combination of hyperparameter values found per model is also displayed in Table 3.

Feature selection plays a crucial role in model development, as the accuracy of a RF classifier relies heavily on the relevance of input features (Rogers and Gunn, 2006). Initially, we adopted a feature set as proposed by Huber et al. (2023) that was also utilised in a previous AutoATES application in Bulgaria (Panayotov et al., 2024b). The feature set consisted of slope, percent canopy cover (PCC), a binary-thresholded PRA (PRA_{bin}) layer, and the maximum modelled local travel angle (γ). The goal at this stage was to address the first research question – whether a Random Forest model can effectively replace manually designed classification trees. This first model trained on these four features is referred to as RF Model 1 (RF1).

Next, we investigated whether RF can provide additional insights into feature importance and whether incorporating more input features enhances performance. Since the algorithm records the features used at each decision node, feature importance can be calculated to reveal which features most influence ATES classification. The first modification involved replacing the binary PRA layer with its continuous counterpart (PRA_{cont}) to leverage the full range of information it provides, rather than reducing it to categories prematurely. We then investigated whether additional `com4FlowPy` outputs could enhance classification. A limitation of RF1, which mirrors AutoATES Austria's inputs, is its lack of consideration of overhead exposure or avalanche intensity (Huber et al., 2023), which were utilised in other studies by incorporating additional `com4FlowPy` output layers in the final AutoATES classification step (Toft et al., 2024; Sykes et al., 2024).

In our approach, after replacing PRA_{bin} with its continuous counterpart, we incrementally introduced *cell counts*, z^{δ}_{max}, and the sum of modelled route fluxes ($\sum routFlux$), training a new model for each using the same parameter grid as shown in

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

Table 3. This incremental addition of features aimed to assess the impact of each individual feature on the overall classification performance. After incorporating all additional features, we constructed a seven-feature model, referred to as RF Model 2 (RF2). Based on its feature importance rankings, we identified the four most influential features – $\sum routFlux$, slope, local travel angle (γ), and PRA_{cont} – and used them to train a final model, RF Model 3 (RF3). A summary of the features used and their descriptions is presented in Table 4.

After training the three models, we applied them to classify the pixels in the test set and evaluated their performance using a range of metrics, as detailed in Sect. 2.6. We also applied the models to the entire study area, including the Bunderitsa, Demyanitsa, and Bezbog valleys, to qualitatively assess their generalisation and performance across different regions.

The RF algorithm also produces an additional output that reflects the confidence of its predictions. For each data point, the model generates four probability values, each indicating the likelihood of belonging to a specific ATES class. The class with the highest probability is selected as the final prediction. Higher confidence is achieved when a larger proportion of decision trees in the forest agree on the same class (Bhattacharyya, 2011). We generated rasters that visualise the model's confidence in the predicted ATES classes across the entire study area. Additionally, we analysed the distribution of predictive confidence for each model and examined how it varies across different ATES classes.

It is essential to emphasise that all random processes involved in model training, such as the train-test split, model initialisation, and random selection from the hyperparameter grid, were consistently performed using a fixed seed. Using the same seed ensures that, despite the randomness of these steps, they produce identical results each time they are executed (Dutta et al., 2022). As a result, the train-test split remained unchanged across all three models (RF1, RF2, RF3), and the same set of hyperparameter combinations was explored. Keeping these factors constant enabled a more reliable comparison between models, ensuring that any observed differences were attributable to variations in the input features rather than differences in training data or hyperparameter selection.

2.6 Methods used for performance evaluation

To assess model performance, we computed several evaluation metrics on the test set, including accuracy and the per-class precision, recall, and F1 score (Rainio et al., 2024). Estimation based on a held-out test set has been shown to offer reliable estimates of model generalisation performance in previous studies (Varoquaux and Colliot, 2023).

The most fundamental metric is accuracy. It is defined as the proportion of correctly classified predictions out of all pixels in the test set (Eq. (3)):

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \tag{3}$$

While accuracy is useful, it does not provide insight into how well the model performs for each individual class, especially in imbalanced datasets, as was the case with the *extreme* terrain class. Therefore, we also computed precision, recall, and F1-score, using binary confusion matrices for each class containing the numbers of *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), and *False Negative* (FN) predictions.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

Table 4. Description of input features and the corresponding Random Forest models (RF1-3) in which each feature was used. More detailed information on each feature is provided in Sect. 2.2 and 2.4.

Feature	Description	Used in models		
slope	local slope gradient computed from the DTM after Horn (1981), using the `gdal` library	RF1	RF2	RF3
PCC	percent canopy cover obtained from the Copernicus dataset	RF1	RF2	
$\mathrm{PRA}_{\mathrm{cont}}$	potential release areas (PRA) without thresholding, $\mathrm{PRA} \in [0,1]$, retaining continuous values		RF2	RF3
$\mathrm{PRA}_{\mathrm{bin}}$	PRA thresholded at 0.3 and converted to a binary classification (0 or 1)	RF1		
γ	local travel angle, used to describe avalanche runout behaviour, computed using `com4FlowPy`	RF1	RF2	RF3
cell counts	number of release cells affecting each raster cell in the model domain, computed using `com4FlowPy`		RF2	
z^{δ}_{max}	maximum modeled avalanche intensity per cell, derived from `com4FlowPy` simulations		RF2	
$\sum \mathrm{routFlux}$	sum of all routed flux values, representing the cumulative effect of potential avalanche paths, computed using `com4FlowPy`		RF2	RF3

Precision is defined as the proportion of instances predicted to belong to a class that are actually members of that class – in other words predicted correctly. It is calculated as:

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)} \tag{4}$$

Recall is defined as the proportion of instances that actually belong to a certain class and are correctly predicted as belonging to that class. It is calculated by:

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)} \tag{5}$$

The F1 score is the harmonic mean of precision and recall, providing a balanced measure of model performance. It is calculated by:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

In addition to the per-class skill-scores, we also computed the weighted mean of each metric by weighting the per-class metrics according to the proportion of each class in the test set. The raw numbers used for calculating the evaluation metrics were also visualised in a multi-class confusion matrix. The full confusion matrix allows for a more detailed interpretation of model predictions compared to the test set. For example, distinctions between instances where model predictions differ from the true labels of the test data by only one ordinal class and more severe instances of misclassification can be made (Rainio et al., 2024).

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

Another common method to evaluate model performance for classification tasks are plots of the precision-recall curve and evaluation of the associated areas under the curve (AUC). The precision-recall curve visually represents the trade-off between precision and recall as the decision threshold is varied. The decision threshold defines the probability above which an instance is classified as belonging to a particular class. For example, if the threshold is set to 0.5 and the maximum predicted probability for a class is 0.4, the instance will not be assigned that class, as its probability does not exceed the threshold. Precision and recall are both influenced by the threshold. As the threshold decreases, recall tends to increase while precision may decrease, and vice versa. The curve is useful for evaluating model performance, especially in the case of imbalanced classes, with a larger area under the curve indicating better overall accuracy (Rainio et al., 2024). In this study, we produced class-wise precision-recall curves for all three trained models and calculated the corresponding AUCs. The precision-recall curve was chosen over the commonly used receiver-operator curve (ROC), which plots the true positive against the false positive rate, due to its more robust performance in the presence of class imbalances.

In addition to the quantitative methods used to evaluate the models' performances on the test set, we also conducted a qualitative assessment of model predictions across the entire study area. This involved a careful visual inspection of the model predictions, including in regions outside the training area, thereby providing initial insights into the generalisation capabilities of the models.

3 Results

3.1 Quantitative evaluation against test set

Table 5 presents the full multi-class confusion matrices for each model, comparing the model predictions against the held-out test set. The area overlap between the predicted and the actual ATES classes for each class and each model is provided. Since each pixel represents an area of $100\ \mathrm{m}^2$, the number of overlapping pixels can be calculated by dividing the area value by 100. The diagonal of the matrix, highlighted by bold numbers, represents correct classifications, while misclassifications below the diagonal correspond to underclassification errors (lower ATES class predicted than actual one), and those above it indicate overclassification errors (higher ATES class predicted than actual one). Model RF1 has $1\,669$ overclassified pixels out of the total $13\,025$ in the test set ($12.81\ \%$) and $1\,026$ underclassified pixels ($7.9\ \%$). Model RF2 has $1\,242$ overclassified pixels ($9.54\ \%$) and $1\,061$ underclassified pixels ($8.15\ \%$). Model RF3 has $1\,497$ overclassified pixels ($11.49\ \%$) and $1\,053$ underclassified pixels ($8.08\ \%$). For each model, none of the *simple* class pixels in the test set were misclassified as *extreme*, and only three *extreme* class pixels were incorrectly predicted as *simple*, representing a very small proportion of the total misclassifications. Additionally, the confusion matrices show that nearly all classification errors involve a misclassification by only a single class level. The confusion matrix presents the raw class counts, which were used as the basis for deriving the subsequent evaluation metrics.

The overall accuracies for each model are $79.31\ \%$ for RF1, $82.32\ \%$ for RF2, and $80.42\ \%$ for RF3. These results indicate that model RF2 achieves the highest overall accuracy, followed closely by model RF3, which performs slightly better than model RF1 in terms of correct predictions.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

Table 5. Multi-class confusion matrices for Random Forest models RF1, RF2, and RF3. Absolute values for area overlap are provided in m^2, division by 100 yields number of pixels/predictions. Values in bold along the diagonals correspond to correct predictions; values below and above the diagnoals show under and overclassifications respectively.

	Predicted ATES Class			
	Simple	**Challenging**	**Complex**	**Extreme**
Actual ATES Class				
RF1				
Simple [m^2]	**404 600**	43 400	2 300	0
[%]	**89.85**	9.64	0.51	0
Challenging [m^2]	41 800	**225 400**	86 600	3 500
[%]	11.7	**63.08**	24.24	0.98
Complex [m^2]	1 700	45 600	**305 700**	31 100
[%]	0.44	11.87	**79.59**	8.1
Extreme [m^2]	300	100	13 100	**97 300**
[%]	0.27	0.09	11.82	**87.72**
RF2				
Simple [m^2]	**412 400**	36 700	1 200	0
[%]	**91.58**	8.15	0.27	0
Challenging [m^2]	43 600	**253 500**	56 900	3 300
[%]	12.2	**70.95**	15.92	0.92
Complex [m^2]	1 500	44 200	**312 300**	26 100
[%]	0.39	11.51	**81.31**	6.8
Extreme [m^2]	300	100	16 400	**94 000**
[%]	0.27	0.09	14.8	**84.84**
RF3				
Simple [m^2]	**407 600**	41 000	1 700	0
[%]	**90.52**	9.11	0.38	0
Challenging [m^2]	44 900	**231 300**	77 700	3 400
[%]	12.57	**64.74**	21.75	0.95
Complex [m^2]	1 500	41 500	**315 200**	25 900
[%]	0.39	10.8	**82.06**	6.74
Extreme [m^2]	300	0	17 100	**93 400**
[%]	0.27	0	15.43	**84.3**

Figure 4 displays the additional class-wise evaluation metrics for all models, calculated against the test set. Notably, for all three models, the precision, recall, and F1-scores for the *simple* terrain class are higher than for the other ATES classes, averaging around 90 %. The calculated metrics for the *challenging* class are consistently the lowest across all three models, while the metrics for the *complex* and *extreme* classes generally fall between those of the *simple* and *challenging* classes. Furthermore, for model RF1, the *extreme* class exhibits a high recall of 87.8 % on the test set, with similar recall values of around 85 % observed for the other models.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

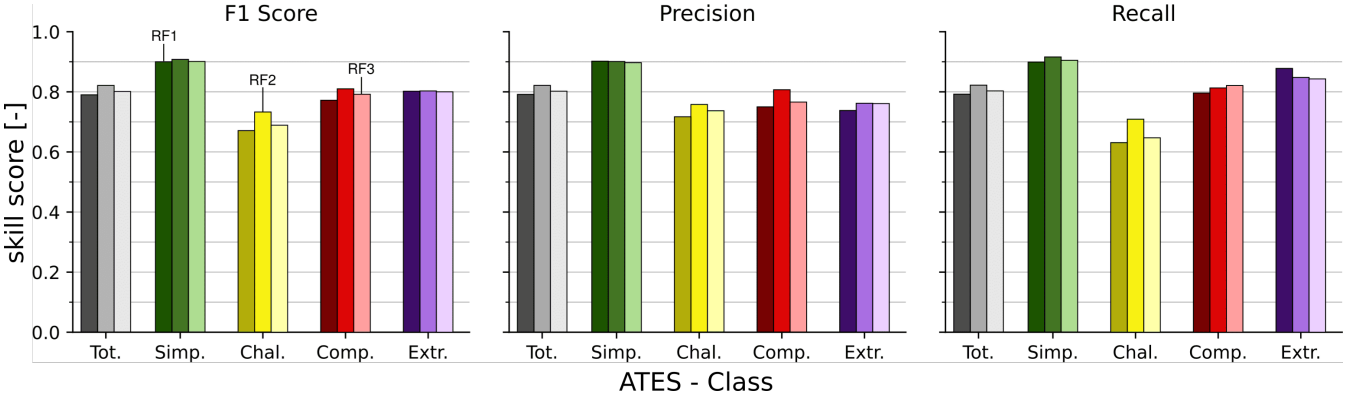

Figure 4. Per-class skill-scores for the three Random Forest models: RF1 (left), RF2 (middle), and RF3 (right), based on the test set. Total scores for the F1 score, precision and recall metrics are calculated as the weighted mean of class-specific scores. Tot.: *total* , Simp.: *simple*, Chal.: *challenging*, Comp.: *complex*, Extr.: *extreme*.

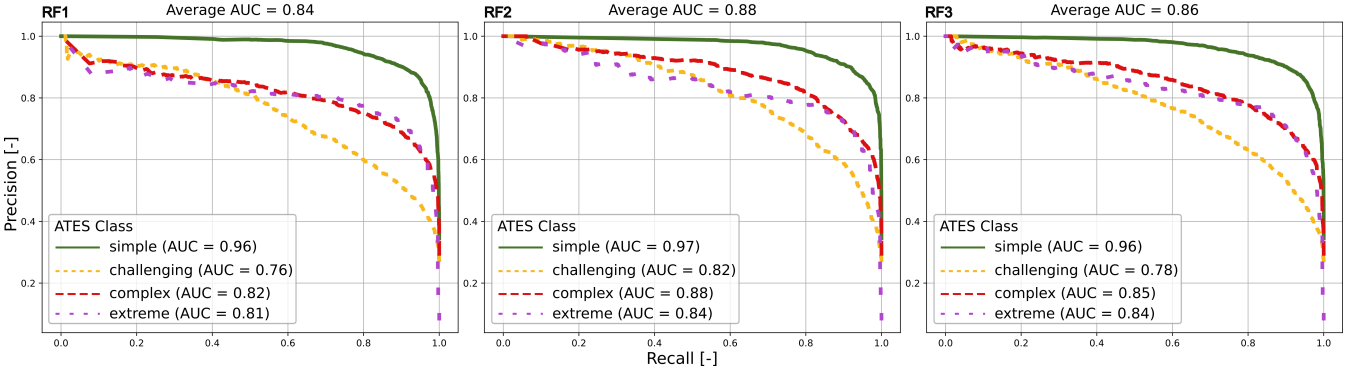

Figure 5. Precision-recall curves for each Random Forest model (RF1-3), shown per class, and the corresponding areas under the curve (AUC).

These general trends are also reflected in the precision-recall curves displayed in Fig. 5. In terms of this evaluation method, all three models exhibit the highest predictive capacities for the *simple* ATES class, while the *challenging* class consistently scores the lowest in computed AUC values. Predictive capacities for the remaining two ATES classes fall between the *simple* and *challenging* classes, with overall AUC values being highest for model RF2, followed by RF3 and RF1. Similar to the other utilised evaluation metrics, the variance between ATES classes is generally more pronounced than the differences between the three trained models.

3.2 Predictive confidence

Figure A2 shows maps of the predicted ATES classes for all three RF models (panels a-c) along with the prediction probabilities (panels d-f) for the entire study area. General patterns for both the ATES class predictions and the associated prediction

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

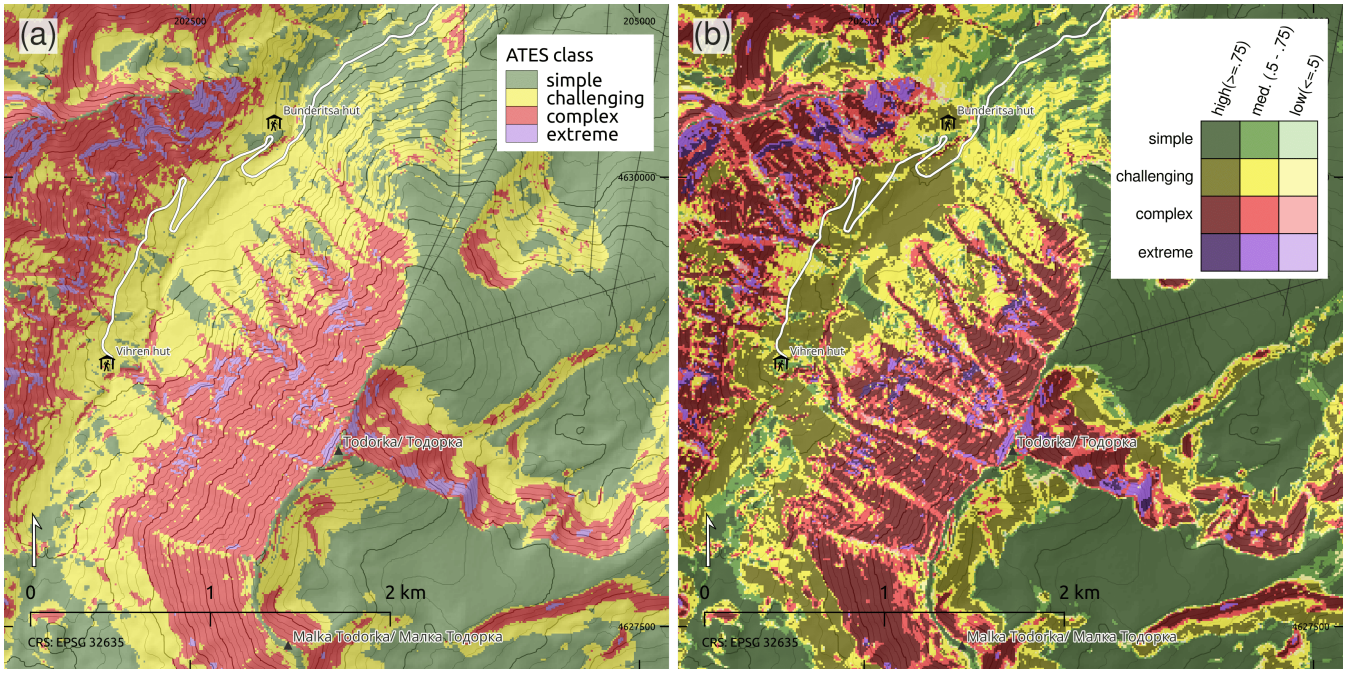

Figure 6. (a) Detail of ATES class predictions using Random Forest model RF2 for the area around Todorka Peak and Bunderitsa Valley. The steep west and north-east faces of Todorka are accurately classified as the complex and extreme ATES classes. **(b)** Example for additional visualisation of model prediction probabilities for RF2 in the same area. Model confidence is represented using colour shadings, with darker shades indicating higher prediction probabilities: high confidence ≥ 0.75, medium confidence $0.5 - 0.75$, low confidence ≤ 0.5.

probabilities appear to follow similar trends across all three RF models, with higher predictive confidence generally associated with the *simple* ATES class.

A closer look is provided in Fig.6, which zooms in on an area of the study region surrounding Todorka Peak and the top chairlifts of Bansko Ski Resort. Figure 6 (a) shows the raw 4-class predictions of the RF2 model, while Fig. 6 (b) additionally depicts the model's prediction probabilities for each ATES class, represented by intra-class colour shading. Darker colours indicate higher prediction confidence. In this example, generally high prediction confidence is associated with areas identified as *simple* terrain, but other trends can also be observed. Notably, there is a general pattern of higher predictive confidence at the center of individual terrain features associated with a specific ATES class, while predictive confidence tends to decrease near the boundaries between classes.

An alternative visualisation of the distributions of prediction probabilities for each ATES class across all three models is provided in Fig. 7. While this perspective confirms the general trends observed in the previously mentioned maps, it also reveals more nuanced variations in prediction probabilities between the three models. All three models clearly show the highest predictive confidence for *simple* terrain, with generally lower confidence for *challenging* terrain. Also, the confidence for the *challenging* class is slightly lower on average for models RF2 and RF3 as compared to model RF1. However, models RF2 and

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

435 RF3, in particular, exhibit relatively high confidence in predicting large proportions of *complex* terrain. Notably, the bimodal distributions of prediction probabilities for *complex* terrain in model RF1 and *extreme* terrain in model RF2 are characteristics that do not become apparent from the general evaluation metrics (Sect. 3.1) or the visual inspection of maps (Fig. 6, A2).

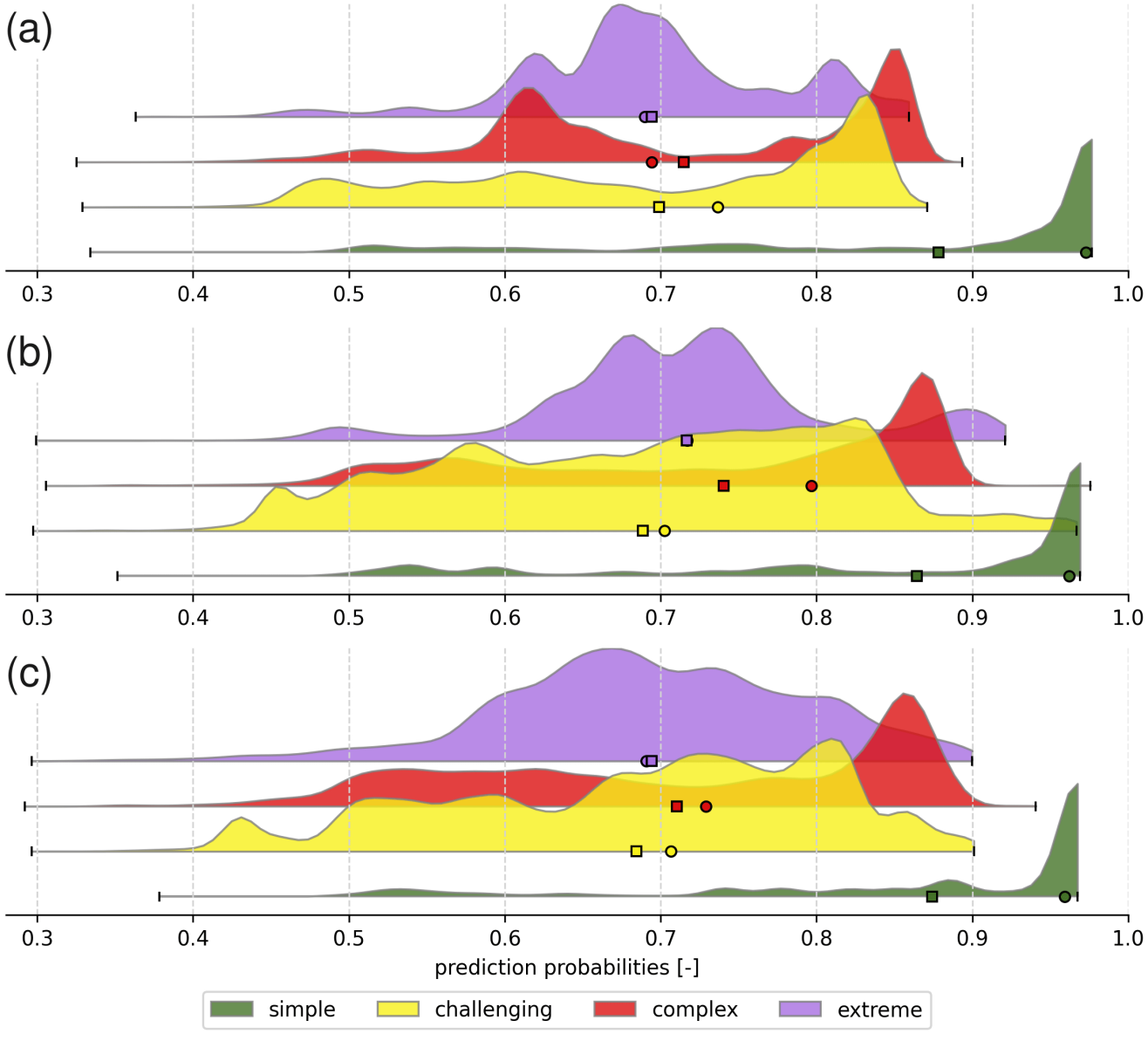

Figure 7. Distributions of prediction probabilities for each ATES class and Random Forest (RF) model across the entire study area. **(a)** RF 1 **(b)** RF2 **(c)** RF3. The means and medians of each distributions are indicated by the square and circle symbols respectively. The distributions for each ATES class in the panels are ordered from top to bottom as *extreme*, *complex*, *challenging*, and *simple*.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

3.3 Feature importance analysis

As outlined in the introduction, a defining feature of RFs is their ability to provide information on the relative importances of input features, alongside the model predictions and associated probabilities described in the previous sections. Figure 8 visualizes the relative feature importances for each of the three trained models.

Terrain inclination (slope) has been used as an input feature in all three RF models and consistently ranks among the top two in terms of feature importance. Similarly, the second input feature used in all models, local travel angle (γ), also ranks within the top three most important input features for all models, indicating that this information plays a significant role in model decision-making. The $\sum$routFlux output from `com4FlowPy` has only been included in models RF2 and RF3, but is considered the most important input feature for both models, albeit with a slight margin. At the opposite end of the spectrum, the PCC information consistently ranks last for feature importance among the models it has been used in. Also, the modelled avalanche intensities z^{δ}_{max} appear to have relatively minor importance for RF2.

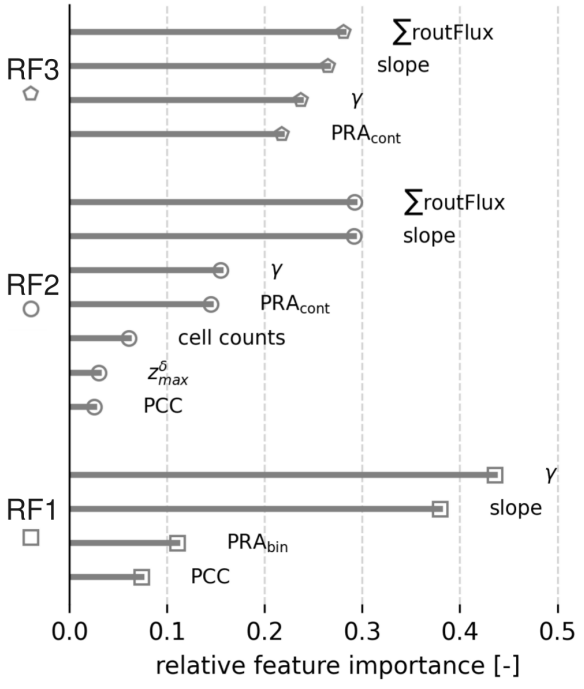

Figure 8. Relative feature importances for Random Forest models RF1, RF2 and RF3, highlighting the contribution of each input feature to the model's decision-making process.

3.4 Qualitative comparison to field assessment

In order to assess the general ability of the trained RF models to generalise to previously unseen terrain, we qualitatively compared model predictions for areas outside the training region with local terrain assessments at selected locations.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

Figure 9 shows two examples comparing the ATES class predictions from the best-performing model, RF2, with photographs of actual winter terrain around Bezbog and Gazei Peaks. Annotated arrows highlight the corresponding terrain features between photographs and the ATES predictions overlaid on satellite imagery. In both examples, model predictions align well with on-site assessments of the avalanche terrain characteristics. The prominent terrain features visible in the photographs are also accurately recognised by the model. For instance, in the Bezbog Peak region, the broad, relatively flat ridge (indicated by arrow 2), which represents the standard winter ascent route, is correctly identified as *simple* terrain. Similarly, in the Gazei Peak area, prominent cliff bands (arrows 2 and 4) are correctly classified as *extreme* terrain in the ATES map. This visual comparison highlights the model's effectiveness in classifying avalanche terrain characteristics in previously unseen regions.

4 Discussion

4.1 Potential of machine learning (ML) classifiers in AutoATES model chains

The results obtained for RF model 1 (RF1) demonstrate that machine learning (ML) techniques can effectively perform the final classification task in automated ATES model chains, using the same input features as deterministic, expert-based classification methods (cf. Huber et al., 2023). This holds true even in data-sparse regions such as our study area, where high-resolution forest cover data are unavailable. In our study, we employed open-access $10\,\mathrm{m}$ resolution percent canopy cover (PCC) data from Copernicus for avalanche PRA, runout, and intensity modelling, as well as an input to the final classification step. These data proved sufficient for RF1 to achieve an accuracy level of $79.31\,\%$, which is comparable to accuracies reported by Sykes et al. (2024). In that study, the standard AutoATES 2.0 framework (Toft et al., 2024), optimised through fine-tuned parameters, was evaluated against expert-generated maps in two Canadian study areas, achieving classification accuracies of $74.5\,\%$ and $84.4\,\%$. It is important to note, however, that Sykes et al. (2024) assessed AutoATES 2.0 performance against continuous, manually generated maps covering the entire study regions, whereas our evaluation is based on a set of randomly sampled test pixels derived from expert-labelled polygons.

As shown in Fig. 4, RF1 performs best in distinguishing the *simple* and *extreme* ATES classes, while exhibiting greater difficulty in classifying the *challenging* and *complex* classes. The high recall for the *extreme* class indicates that a substantial proportion of pixels genuinely belonging to this class were correctly identified. This is particularly important in the context of avalanche risk mitigation, where failing to detect high-risk areas can have serious consequences. The precision for the *extreme* class is somewhat lower ($73.8\,\%$), meaning that $73.8\,\%$ of the pixels predicted as *extreme* terrain by the model were indeed labelled as such in the test dataset. For this ATES class, achieving high recall is more critical than maximising precision, as the primary objective is to ensure that hazardous terrain is not overlooked.

This pattern is consistent with previous research (Sykes et al., 2024; Spannring, 2024), which similarly reports that AutoATES model chains face the greatest classification challenges with the *challenging* and *complex* classes. This result is not unexpected, as the *simple* and *extreme* classes represent opposite ends of the classification spectrum, with well-defined characteristics that make them more easily distinguishable.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

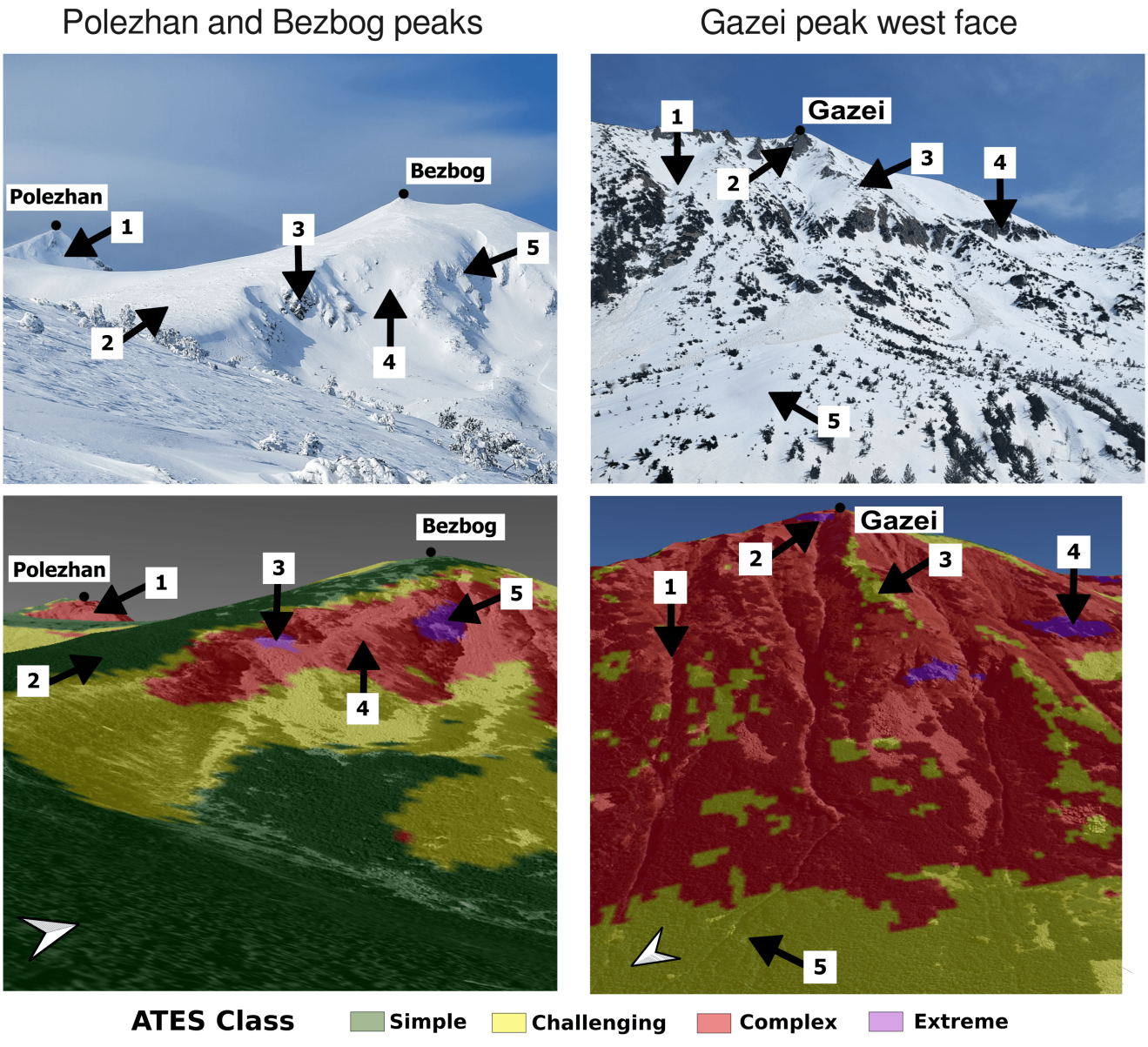

Figure 9. Exemplary extract of the ATES map from model RF2 for two separate areas outside the training region – Gazei Peak and Bezbog Peak – along with photographs of the winter terrain. The base imagery over which the ATES maps are overlaid is sourced from © Google Earth. The annotated arrows highlight the corresponding terrain features between the photographs and the ATES predictions.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

The precision-recall curve for model RF1, presented in Fig. 5, further confirms the model's effectiveness in the ATES classification task. The average area under the curve ($\mathrm{AUC} = 0.84$) is much closer to a perfect classifier ($\mathrm{AUC} = 1.0$) than a random one ($\mathrm{AUC} = 0.5$). The *simple* class exhibits the most favourable trade-off between precision and recall, reinforcing the consistency of our findings across multiple evaluation metrics that the *simple* terrain class is the most successfully classified. This suggests that, across all decision thresholds, the model maintains a high level of accuracy for the *simple* class, preserving both precision and recall. The model also demonstrates strong performance for the *complex* and *extreme* classes, whereas for the *challenging* class there is a more pronounced decline in the precision-recall curve. This drop indicates that as the decision threshold increases – requiring a higher confidence level for a prediction – the model's precision and recall for this class deteriorate more rapidly.

These results are consistent with expectations, as the *challenging* class encompasses a wide range of terrain types, from short, moderately steep slopes to very steep forested areas and even flat terrain within avalanche runout zones. Given this variability, misclassification with the *complex* or *simple* terrain classes is more likely, leading to lower precision and recall scores at higher decision thresholds (i.e., the right side of the precision-recall curve). This pattern is further reflected in the decreased precision, recall, and F1 scores for the *challenging* class in Fig. 4.

Another key finding is that the incorporation of additional input features (model RF2) enhances ATES classification performance. Machine learning models, such as RFs, excel at integrating diverse sources of information, leveraging each input feature to some extent to improve classification accuracy. A particular strength of RF is its ability to select the most informative feature at each decision split, ensuring that irrelevant features do not negatively influence the model (Breiman, 2001). Consequently, expanding the feature set logically contributes to improved classification outcomes.

Model RF2 achieves an overall accuracy improvement of around $3\,\%$ compared to model RF1. Figure 4 reveals improvements in precision, recall, and F1 score for model RF2 across all classes except *extreme* terrain, where performance remains largely unchanged. The most notable improvement is observed in the *challenging* class. The recall for the *challenging* class increases from $63.08\,\%$ in RF1 to $70.95\,\%$ in model RF2. Additionally, the average area under the precision-recall curve (Fig. 5) increases (in the case of the *challenging* class - from 0.76 to 0.82), and the curve for the *challenging* class exhibits a slower decline compared to model RF1.

These findings indicate that incorporating additional features significantly benefits classification using RF models, particularly for the *challenging* class. *Challenging* terrain frequently includes flat areas susceptible to avalanches originating from higher elevations. Thus, introducing features such as *cell counts* and $\sum routFlux$, which quantify overhead hazard and flow thickness-dependent terrain trap potential, proves advantageous in refining classification accuracy. This observation aligns with discoveries in recent studies, which also included overhead exposure in AutoATES to improve classification. (Toft et al., 2024; Sykes et al., 2024).

As described in Sect. 3.1, Table 5 shows that across all three models, the majority of misclassifications are overclassifications. This type of error is preferable in potential hazard classification problems such as ATES, where it is safer to overestimate rather than underestimate risk. Additionally, the confusion matrices indicate that nearly all misclassifications differ by only a single class level, suggesting that the models are generally making reasonable and conservative predictions.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

Our qualitative expert evaluation (Fig. 9) confirms that the RF2 model generalises effectively to regions not included in the training data, suggesting that Random Forests could likely be applied to similar mountainous terrain in other regions and achieve reasonable results.

4.2 Additional benefits of machine learning (ML)

4.2.1 Feature importances

One key benefit of the Random Forest algorithm is that it allows for the assessment of feature importances after the model has been trained – a capability not available in the traditional, expert-based AutoATES approach.

As illustrated in Fig. 8, percent canopy cover (PCC) has a very low feature importance in both models RF1 and RF2. This indicates that it is involved in a minimal percentage of the splits and can likely be excluded as a feature. Since PCC is already accounted for in the PRA calculation and `com4FlowPy` avalanche runout and intensity modelling, it does not contribute additional information in the ATES classification step. Furthermore, removing PCC from current AutoATES classifiers would simplify the process, as thresholding for this feature has been a particularly challenging task to fine-tune, often leading to inconsistencies, with different thresholds being applied in various studies (Huber et al., 2023; Panayotov et al., 2024b; Toft et al., 2024).

In model RF3, PCC was removed as input feature and only the top four features by importance from model RF2 – $\sum routFlux$, slope, travel angle (γ), and PRA as a continuous raster – were retained. As shown in Table 5 and Fig. 4 and 5, the model continues to perform well in the classification task, with performance slightly lower than RF2. This suggests that restricting the feature set to just four features and excluding the PCC would still yield better results than using the four features from RF1. This constitutes an interesting discovery that could be used for future improvements to AutoATES. However, additional investigation is warranted to assess whether using a higher-resolution PCC layer, or applications to different study regions, would yield the same low feature importance for PCC in the AutoATES classifier. This would help to confirm that its minor importance is not merely an effect of data quality or peculiarities of the study region.

The feature importance analysis also reveals that incorporating new features, such as $\sum routFlux$, can enhance model performance. In addition to improving classification accuracy in model RF2, $\sum routFlux$ emerges as the most important feature in that model, followed closely by slope. This is not unexpected, as the $\sum routFlux$ output of `com4FlowPy` combines information about the size of overhead PRAs with potential flow and deposit thickness of modelled avalanches. As such this finding is in alignment with recent studies that also suggested including input features characterising overhead hazard in the AutoATES classifier (Sykes et al., 2024; Toft et al., 2024).

Slope, widely recognised as one of the most critical factors in avalanche modelling, consistently ranks as a key feature across all models. Similarly, travel angle (γ), which indicates the proximity to PRAs and the average track inclination at different terrain points, has been included in all previous AutoATES versions and applications (e.g. Larsen et al., 2020; Schumacher et al., 2022; Huber et al., 2023; Toft et al., 2024), and remains a highly relevant predictor, as confirmed by its feature importance ranking.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

Moreover, during the process of iteratively expanding the feature set, as described in Sect. 2.5.2, the first step was to construct a model using the same input features as RF1, but replacing the binary PRA (PRA_{bin}) with its continuous counterpart (PRA_{cont}). In this model, the PRA's feature importance increased from 0.11 in binary form (RF1) to 0.21 in continuous form, highlighting another advantage of ML techniques such as RF, which can leverage continuous input features to their full potential. By allowing the classifier to learn optimal thresholds from data, rather than relying on predefined classification criteria, these methods can enhance predictive efficiency and model flexibility.

4.2.2 Practical advantages of machine learning for avalanche terrain mapping

It is also worth highlighting the demonstrated flexibility in training data format. Our results show that a continuous ATES map is not a prerequisite for effective model training. Instead, machine learning models can be successfully trained using isolated, expert-labelled polygons, as demonstrated in this study. This approach facilitates confident labelling of small, confined regions, whereas drawing continuous ATES maps is often more challenging, particularly in delineating boundaries between ATES classes, which is a task that can be a difficult even for experienced mappers. Therefore, this methodology significantly reduces the burden on domain experts when constructing training datasets and reinforces the practicality and accessibility of ML-based approaches for avalanche terrain classification.

4.2.3 Confidence in predictions

Another valuable output of the Random Forest classifiers is the confidence associated with their predictions for each class, as shown in Fig. 7. Across all three models, the highest confidence is consistently observed for the *simple* class, which is expected given that *simple* terrain is generally the easiest to distinguish.

Notably, the average confidence levels for the *challenging* class decrease slightly in models RF2 and RF3 compared to RF1. This suggests that these models are better calibrated and realistic, as they do not assign high confidence indiscriminately when predicting the *challenging* class. The observed behaviour may stem from the inclusion of additional features in RF2 and more informative features in RF3, enabling the models to better capture the complexity of certain cases and appropriately temper their prediction confidence. In contrast, RF1 exhibits higher average confidence for this class, but its lower precision, recall, and F1 scores indicate that this confidence may be misleading – indicating a degree of overconfidence. However, model RF2 also predicts a larger number of pixels with very high confidence for the *challenging* class, as indicated by the extended right-hand tail in Fig. 7(b).

As shown in Fig. 7, the confidence distribution for the *complex* class in model RF1 exhibits two peaks, with a lower one around 0.6. In model RF2, this secondary peak disappears, and the median shifts to a higher value, close to 0.8. This shift suggests that adding more features increases the model's average confidence in its predictions for the *complex* class.

The confidence map for model RF2 shown in Fig. 6 reveals high-certainty predictions for *simple* terrain in flat areas situated far from potential avalanche runout zones – such as within the ski resort – where terrain characteristics are well-defined. Similarly, the steep western face of Todorka peak is predominantly classified as *complex* terrain with high confidence, while the runout zones below are identified as *challenging* terrain with similarly strong confidence. In contrast, classification certainty

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

is lower in the transition zones between ATES classes, particularly between *challenging* and *simple* terrain, which is expected, as defining the precise boundaries between ATES classes is inherently challenging. In these regions, feature values are likely less distinctly associated with a single ATES class and instead fall between classes, leading to reduced classifier confidence.

Higher uncertainty is also evident in forested regions, particularly east of Bunderitsa Hut and west of the ski lift, where the model predominantly predicts *challenging* terrain with moderate confidence. Small patches of *simple* terrain with low confidence appear within the forest, indicating classification uncertainty in these areas. This is consistent with results from previous applications of AutoATES in the same region, where the delineation of ATES classes in the same spots has been particularly challenging due to the inherent limitations of the available forest layers, as well as challenges associated with parametrisation of the PCC contribution to the AutoATES classifier (Panayotov et al., 2024b).

In contrast to the purely deterministic output provided by classical AutoATES classifiers, the prediction confidence provided by the RF classifier allows for a more detailed interpretation of model predictions. For example, confidence scores produced by the RF models alongside the model predictions might be utilised in the process of manual quality control and fine-tuning of automatically produced ATES maps. As such, they can provide a more nuanced perspective on model predictions and aid in identifying areas that warrant special attention before final map publication. Moreover, information on the prediction confidence could serve as a basis for a more explicit visualisation and communication of uncertainties associated with discrete class borders, as previously suggested in the fuzzy ATES data model proposed by Sharp et al. (2023).

4.3 Limitations and potential for improvement

In this study, we systematically evaluated the performance of automated ATES classification using different RF models, based on a held-out test set and multiple evaluation metrics. While overall accuracy provides a general measure of model performance, class-specific metrics – such as precision, recall, and F1 score – offer deeper insight into per-class behaviour, helping to identify areas for future improvement. Although these methods of evaluation successfully approximate model generalisation to unseen areas, several limitations must be acknowledged.

One major limitation is that RF models were evaluated exclusively on a held-out test set. While this approach is widely accepted in ML for assessing generalisation performance to unseen terrain (Varoquaux and Colliot, 2023), the test set only includes pixels from within the training region. We did not formally assess model performance using quantitative metrics outside the training region. Instead, for the entire study area, we conducted initial evaluations through visual comparisons and qualitative validation (see Fig. 9).

As with any supervised learning approach, the RF classifier is highly dependent on the quality and construction of the training dataset, including subjective decisions made by the experts annotating it. This is particularly relevant when the training data is created by a single expert, as was the case in our study. Substantial modification to the training data could significantly influence both the classification outcomes and the resulting feature importance rankings.

Although mountainous terrain tends to share broad characteristics globally, our training data was limited to a specific region of the Pirin Mountains. Expanding the training dataset and study areas to include a broader range of mountain regions, both within and outside Bulgaria, could improve model generalisability. Such a dataset would further support the development

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

of more universal models for application to previously unmapped areas, while localised training could yield region-specific models tailored to distinct mountain environments. A formal comparison of locally vs. universally trained ML models across different study regions could provide further insights into the advantages and disadvantages of these two approaches.

Another limitation is that the RF classifier operates on a per-pixel basis. While this is consistent with previously proposed AutoATES classifiers, it does not account for spatial autocorrelation effects. Incorporating information from a local neighbourhood window or region for predictions at a given location could further improve classification performance and produce smoother, more coherent results. One approach that leverages this spatial context and could be explored in future work is a convolutional neural network (CNN), which applies a sliding window mechanism across the input data.

Differences in mapping scales between human- and machine-generated ATES maps also present a challenge (Toft et al., 2024; Sykes et al., 2024). Human cartographers tend to generalise terrain features, resulting in smoother, more abstract representations that omit finer-scale slope variations. In contrast, pixel-based classifiers, such as RF or traditional AutoATES models, generally operate at a finer spatial resolution, capturing small-scale terrain features in greater detail. For practical applications ATES maps typically require a level of cartographic generalisation and smoothing to enhance readability and reduce noise. In this study we did not analyse post-processing techniques such as those proposed by Larsen et al. (2020) or Toft et al. (2024). However, additional cartographic generalisation could help bridge the gap between machine-generated classifications and human-drawn maps, improving harmonisation between the two approaches.

5 Conclusions and outlook

This study demonstrates that ML methods, such as the RF algorithm, can be successfully applied in the final classification step of recently proposed AutoATES model chains. ML techniques present a promising alternative to previously developed AutoATES classifiers, which rely on deterministic thresholds and rules that require adaptation when applied to new regions. The overall classification accuracies achieved in our study are comparable to those reported for more traditional AutoATES approaches. Beyond its demonstrated effectiveness for the investigated classification task, the RF algorithm also offers additional methodological advantages.

One of the key benefits of RF models is their ability to provide feature importances rankings, which enhance the understanding of the relative influence of different input features in automated ATES mapping. Consistent with existing AutoATES algorithms, the feature importance rankings found in our study confirm that slope and local travel angle (γ) are particularly important in determining ATES classes. Moreover, our results indicate that including additional features related to overhead exposure and potential terrain traps can further improve classification accuracy. Our findings also suggest that the role of forest layers, such as PCC, merits further investigation in future AutoATES applications. In our study, PCC was not identified as a significant feature for the final AutoATES classification step by any of the RF models. Since PCC is already included in the PRA and runout modelling steps of the model chain, we hypothesize that excluding this forest layer from the final ATES classification step may enhance overall model performance. This consideration may be relevant not only for ML-based approaches but also for existing decision tree-based AutoATES classifiers.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

Another advantage of the RF methodology is its ability to provide prediction confidence levels. This information can support human interpreters during manual quality control and in making revisions to map products prior to their release to end users. Moreover, prediction confidence may contribute to ATES data models and map representations that aim to communicate mapping uncertainties more explicitly.

While this study offers promising initial insights into the use of ML techniques for automated avalanche terrain classification, some limitations must be acknowledged. These include the geographically restricted training and study area, the exclusive focus on RF modelling, and the limited assessment of model generalisation and smoothing techniques. Addressing these aspects through more comprehensive investigations will be important for future research. Specifically, expanding the training data set to include additional regions and testing alternative ML techniques – such as Support Vector Machines, Gradient Boosting, or Neural Networks – may lead to more generalisable results. Ultimately, future research should not only focus on improving the final ATES classifier but also address earlier steps in the AutoATES model chain, including input data quality, PRA calculations, and avalanche runout modelling.

In conclusion, this study establishes an important foundation for a novel approach to automated ATES classification. With further refinement and optimisation, the proposed methodology has the potential to facilitate more accessible and accurate ATES classification across diverse mountain regions.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

Appendix A

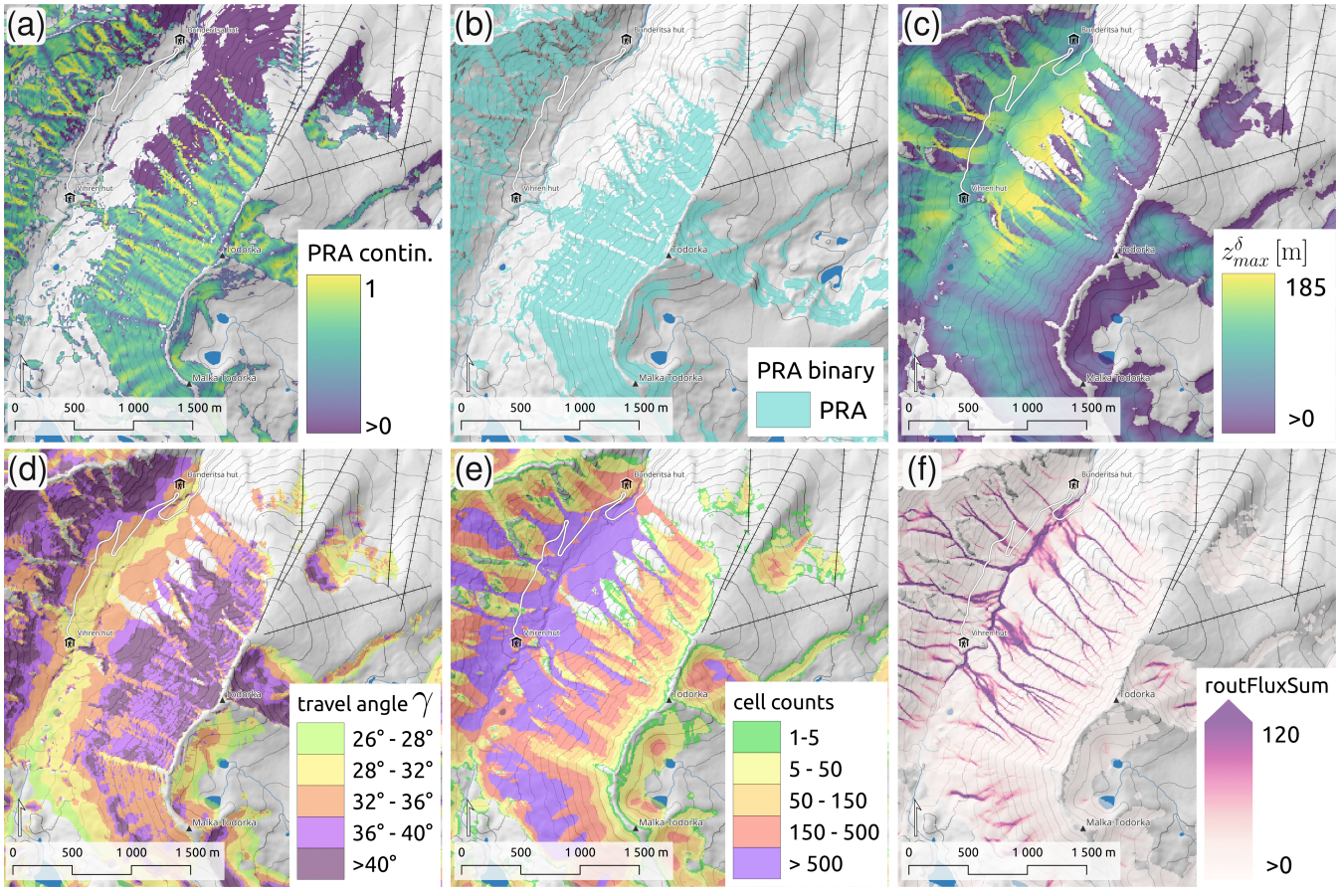

Figure A1. Outputs of the PRA model (panels a, b) and avalanche runout modelling with `com4FlowPy` (panels c-f). **(a)** Continuous PRAs PRA_{cont}, **(b)** binary PRAs PRA_{bin} obtained with a threshold of 0.3, **(c)** maximum modelled energy line height z^{δ}_{max}, **(d)** maximum local travel angle (γ), **(e)** cell counts, and **(f)** sum of routing flux ($\sum \mathrm{routFlux}$).

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

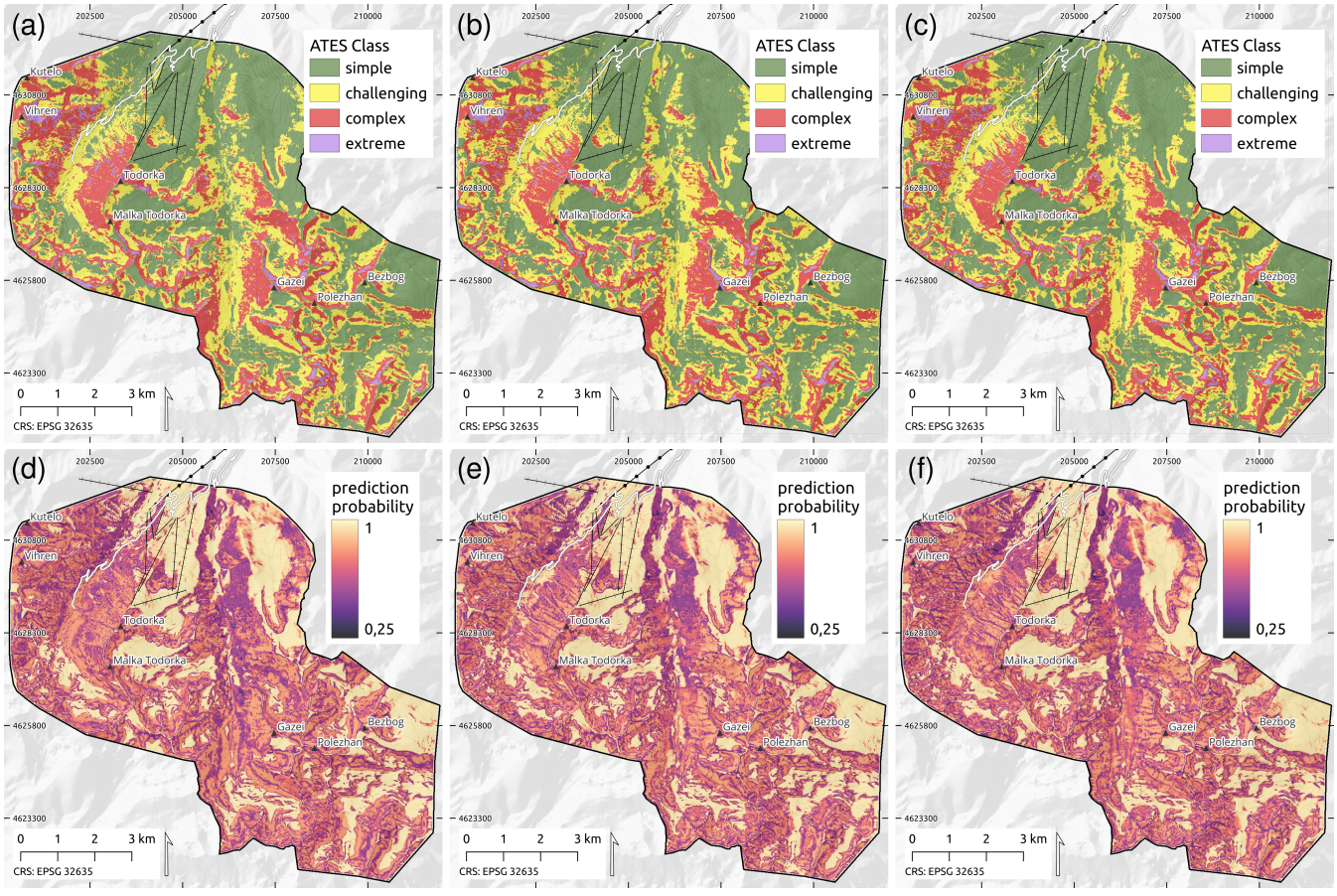

Figure A2. ATES maps produced by the investigated Random Forest (RF) models for the entire study area: **(a)** RF1, **(b)** RF2, and **(c)** RF3. Additionally, prediction probabilities output by the models for the ATES class prediciton: **(d)** RF1, **(e)** RF2, and **(f)** RF3.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

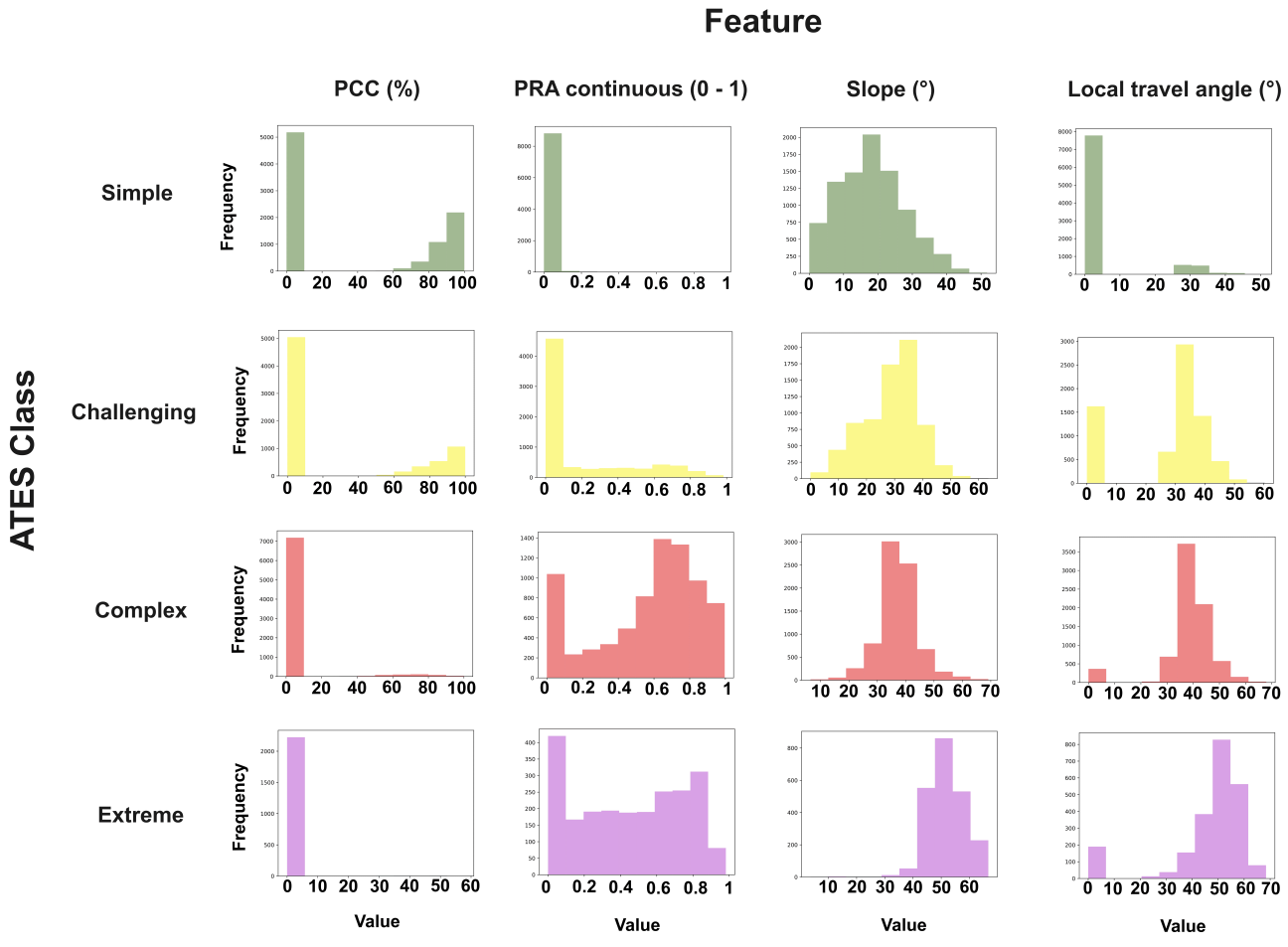

Figure A3. Distributions of input feature values for pixels in the training data, grouped by ATES class. *PCC* stands for *percent canopy cover*, while *PRA* represents the *potential release areas* as a continuous raster, where 0 denotes no possibility of avalanche release, and 1 corresponds to a very high likelihood of release.

Code availability. All Python code and trained model files can be found in our online Git repository: https://doi.org/10.5281/zenodo.15310357 (Markov, 2025) For future work, the GitHub repository: https://github.com/kalinmarkov95/machine-learning-auto-ates/tree/main will be maintained (Markov, 2025). Please refer to the repository for updates and potential additional data uploads.

Data availability. The digital terrain model (DTM) of the study region was obtained from Pirin National Park and the private geospatial data provider Geopolyomorphic Cloud specifically for this study. It is not available for public use and is not included in the Git repository. If you would like to access the DTM, please contact the author, Kalin Markov, or the data providers directly.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

Author contributions. KM: conceptualization, methodology, software, formal analysis, visualization, and writing – original draft preparation. AH: data curation, methodology, investigation, visualization, and writing – original draft preparation. MP, CH, PS, JTF, and MT: writing – review and editing. MP and MT: supervision, project administration, and funding acquisition.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The study was conducted as part of the project "Modelling of avalanches in forested and alpine terrain in Bulgaria and Austria - similarities and differences" funded by the National Science Fund of Bulgaria (Project KP-06-Austria/4/2023) and the Austrian Agency for International Cooperation in Education and Research OeAD (Project number BG 01/2023). Additional financial support was provided through the Open Avalanche Framework AvaFrame (https://www.avaframe.org), a cooperation between the Austrian Research Centre for Forests (Bundesforschungszentrum für Wald; BFW) and the Austrian Service for Torrent and Avalanche Control (WLV), under the auspices of the Austrian Federal Ministry of Agriculture and Forestry, Climate and Environmental Protection, Regions and Water Management (BMLUK), as well as the Interreg Alpine Space project MOSAIC, co-funded by the European Union (grant number ASP0100014). We would like to acknowledge the use of the Copernicus Tree Cover Density product. DTM data from Pirin National Park and Geopolymorphic Cloud were obtained for the needs of project KP-06-N31-3/2019 "Avalanche-forest interactions in the Pirin Mountains, Bulgaria" and also for a project supported by the Bulgarian Extreme and FreeSkiing Association. We acknowledge the use of AI tools that assisted in improving the readability in parts of the initial version of this manuscript.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

References

Acharya, A., Steiner, J. F., Walizada, K. M., Ali, S., Zakir, Z. H., Caiserman, A., and Watanabe, T.: Review article: Snow and ice avalanches in high mountain Asia – scientific, local and indigenous knowledge, NAT HAZARD EARTH SYS, 23, 2569–2592, https://doi.org/10.5194/nhess-23-2569-2023, publisher: Copernicus GmbH, 2023.

Ahsan, M., Khan, A., Khan, K. R., Sinha, B. B., and Sharma, A.: Advancements in medical diagnosis and treatment through machine learning: A review, EXPERT SYST, 41, e13 499, https://doi.org/10.1111/exsy.13499, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/exsy.13499, 2024.

Amin, G., Imtiaz, I., Haroon, E., Saqib, N. u., Shahzad, M. I., and Nazeer, M.: Assessment of Machine Learning Algorithms for Land Cover Classification in a Complex Mountainous Landscape, Journal of Geovisualization and Spatial Analysis, 8, 34, https://doi.org/10.1007/s41651-024-00195-z, 2024.

Avis, C. D. V., Sykes, J. M., and Tutt, B.: DEVELOPMENT OF LARGE SCALE AUTOMATED AVALANCHE TERRAIN EXPOSURE SCALE (ATES) RATINGS IN COLLABORATION WITH LOCAL AVALANCHE EXPERTS, in: Proceedings, International Snow Science Workshop 2023, Bend, Oregon, 2023.

Bhattacharyya, S.: Confidence in Predictions from Random Tree Ensembles, in: 2011 IEEE 11th International Conference on Data Mining, pp. 71–80, https://doi.org/10.1109/ICDM.2011.41, iSSN: 2374-8486, 2011.

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A.-L., Deng, D., and Lindauer, M.: Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges, WIREs Data Mining and Knowledge Discovery, 13, e1484, https://doi.org/10.1002/widm.1484, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1484, 2023.

Breiman, L.: Random Forests, MACH LEARN, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Campbell, C. and Gould, B.: A Proposed Practical Model for Zoning with the Avalanche Terrain Exposure Scale, in: Proceedings, International Snow Science Workshop 2013, Grenoble, France, 2013.

Cetinkaya, S. and Kocaman, S.: IMPACT OF LEARNING SET AND SAMPLING FOR SNOW AVALANCHE SUSCEPTIBILITY MAPPING WITH RANDOM FOREST, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-1-2023, 57–64, https://doi.org/10.5194/isprs-archives-XLVIII-M-1-2023-57-2023, conference Name: 39th International Symposium on Remote Sensing of Environment (ISRSE-39) “From Human needs to SDGs” - 24–28 April 2023, Antalya, Türkiye Publisher: Copernicus GmbH, 2023.

Contreras, P., Orellana-Alvear, J., Muñoz, P., Bendix, J., and Célleri, R.: Influence of Random Forest Hyperparameterization on Short-Term Runoff Forecasting in an Andean Mountain Catchment, ATMOSPHERE-BASEL, 12, 238, https://doi.org/10.3390/atmos12020238, number: 2 Publisher: Multidisciplinary Digital Publishing Institute, 2021.

Copernicus: Tree Cover Density 2018 (raster 10 m and 100 m), Europe, 3-yearly, https://doi.org/10.2909/486f77da-d605-423e-93a9-680760ab6791, 2020.

D’Amboise, C., Teich, M., Hormes, A., Steger, S., and Berger, F.: Protective forests as Ecosystem-based solution for Disaster Risk Reduction (ECO-DRR), chap. Modeling Protective Forests for Gravitational Natural Hazards and How It Relates to Risk-Based Decision Support Tools, IntechOpen, London, https://doi.org/10.5772/intechopen.99510, 2021.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

D'Amboise, C. J. L., Neuhauser, M., Teich, M., Huber, A., Kofler, A., Perzl, F., Fromm, R., Kleemayr, K., and Fischer, J.-T.: Flow-Py v1.0: a customizable, open-source simulation tool to estimate runout and intensity of gravitational mass flows, GEOSCI MODEL DEV, 15, 2423–2439, https://doi.org/10.5194/gmd-15-2423-2022, publisher: Copernicus GmbH, 2022.

Dutta, S., Arunachalam, A., and Misailovic, S.: To Seed or Not to Seed? An Empirical Analysis of Usage of Seeds for Testing in Machine Learning Projects, in: 2022 IEEE Conference on Software Testing, Verification and Validation (ICST), pp. 151–161, IEEE, Valencia, Spain, ISBN 978-1-66546-679-0, https://doi.org/10.1109/ICST53961.2022.00026, 2022.

Engeset, R. V., Pfuhl, G., Landrø, M., Mannberg, A., and Hetland, A.: Communicating public avalanche warnings – what works?, NAT HAZARD EARTH SYS, 18, 2537–2559, https://doi.org/10.5194/nhess-18-2537-2018, publisher: Copernicus GmbH, 2018.

G, S. G. C. and B.Sumathi: Grid Search Tuning of Hyperparameters in Random Forest Classifier for Customer Feedback Sentiment Prediction, International Journal of Advanced Computer Science and Applications (IJACSA), 11, https://doi.org/10.14569/IJACSA.2020.0110920, number: 9 Publisher: The Science and Information (SAI) Organization Limited, 2020.

Gavaldã, J., Moner, I., and Bacardit, M.: Integrating the ATES into the Avalanche Information in Aran Valley (Central Pyrenees), in: Proceedings, International Snow Science Workshop 2013, Grenoble, France, 2013.

Gong, Y., Liu, G., Xue, Y., Li, R., and Meng, L.: A survey on dataset quality in machine learning, INFORM SOFTWARE TECH, 162, https://doi.org/10.1016/j.infsof.2023.107268, 2023.

Gowda V, D., Pathak, D., Prasad, K. D. V., Srinivas, V., Y M, M., and Sudhakar Reddy, N.: Scalable Machine Learning Frameworks for Large-Scale Multimodal Image and Speech Signal Processing, in: 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), pp. 1693–1699, https://doi.org/10.1109/I-SMAC61858.2024.10714812, iSSN: 2768-0673, 2024.

Harvey, S., Schmudlach, G., Bühler, Y., Dürr, L., Stoffel, A., and Christen, M.: Avalanche Terrain Maps for Backcountry Skiing in Switzerland, in: Proceedings, International Snow Science Workshop, Innsbruck, Austria, 2018, pp. 1625 – 1631, Innsbruck, Austria, https://arc.lib.montana.edu/snow-science/objects/ISSW2018_O19.1.pdf, 2018.

Harvey, S., Christen, M., Bühler, Y., Hänni, C., Boos, N., and Bernegger, B.: Refined Swiss Avalanche terrain mapping CATV2 / ATHV2, in: Proceedings, International Snow Science Workshop, Tromsø, Norway, 2024, pp. 1637 – 1644, Tromsø, Norway, https://arc.lib.montana.edu/snow-science/item/3363, 2024.

Hesselbach, C.: Adaptation and Application of an Automated Avalanche Terrain Classification in Austria, Master's thesis, University of Natural Resources and Life Sciences, Vienna, 2023.

Horn, B. K. P.: Hillshading and the Reflectance Map, P IEEE, 69, 14 – 47, https://doi.org/doi:10.1109/proc.1981.11918, 1981.

Huber, A., Hesselbach, C., Oesterle, F., Neuhauser, M., Adams, M., Plörer, M., Stephan, L., Toft, H., Sykes, J., Mitterer, C., and Fischer, J.-T.: AutoATES Austria - Testing and application of an automated model-chain for avalanche terrain classification in the Austrian Alps, in: Proceedings, International Snow Science Workshop, Bend, OR, USA, 2023, pp. 1272 – 1278, Innsbruck, Austria, http://arc.lib.montana.edu/snow-science/item/2989, 2023.

Huber, A., Saxer, L., Spannring, P., Hesselbach, C., Neuhauser, M., D'Amboise, C., and Teich, M.: REGIONAL-SCALE AVALANCHE MODELING WITH com4FlowPy – POTENTIAL AND LIMITATIONS FOR CONSIDERING AVALANCHE-FOREST INTERACTION ALONG THE AVALANCHE TRACK, in: Proceedings, International Snow Science Workshop, Tromsø, Norway, 2024, pp. 587 – 594, https://arc.lib.montana.edu/snow-science/item.php?id=3194, 2024.

Jaiswal, J. K. and Samikannu, R.: Application of Random Forest Algorithm on Feature Subset Selection and Classification and Regression, in: World Congress on Computing and Communication Technologies (WCCCT), 2017.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

Japkowicz, N.: Concept-Learning in the Presence of Between-Class and Within-Class Imbalances, in: Advances in Artificial Intelligence, edited by Stroulia, E. and Matwin, S., pp. 67–77, Springer, Berlin, Heidelberg, ISBN 978-3-540-45153-2, https://doi.org/10.1007/3-540-45153-6_7, 2001.

Körner, H. J.: The energy-line method in the mechanics of avalanches, J GLACIOL, 26, 501 – 505, https://doi.org/10.3189/S0022143000011023, 1980.

Larsen, H. T., Hendrikx, J., Slåtten, M. S., and Engeset, R. V.: Developing nationwide avalanche terrain maps for Norway, NAT HAZARDS, 103, 2829–2847, https://doi.org/10.1007/s11069-020-04104-7, 2020.

Markov, K.: kalinmarkov95/machine-learning-auto-ates: Implement Random Forest (RF) for automated ATES classification, https://doi.org/10.5281/zenodo.15310358, 2025.

Markov, K. and Ivanov, I.: Avalanche hazard assessment and modeling in Pirin National Park, LOPS Foundation, Sofia, Bulgaria, 2021.

Markov, K. and Panayotov, M.: Avalanches in the Bunderitsa valley in Pirin Mountains, chap. Defining of avalanche terrain with ATES modelling for the region of Bansko ski resort in Pirin, Intel Entrans, Sofia, 2024.

McClung, D. and Gauer, P.: Maximum frontal speeds, alpha angles and deposit volumes of flowing snow avalanches, COLD REG SCI TECHNOL, 153, 78–85, https://doi.org/10.1016/j.coldregions.2018.04.009, 2018.

Nasteski, V.: An overview of the supervised machine learning methods, HORIZONS, 4, 51–62, https://doi.org/10.20544/HORIZONS.B.04.1.17.P05, 2017.

Oesterle, F., Wirbel, A., Fischer, J.-T., Huber, A., and Spannring, P.: avaframe/AvaFrame: 1.11, https://doi.org/10.5281/zenodo.14893015, 2025.

Panayotov, M., Menteshev, B., and Mihaylova, S.: Avalanches in Bulgaria, BEFSA, 2021.

Panayotov, M., Christen, M., Bebi, P., and Bartelt, P.: Avalanches in the Bunderitsa valley in Pirin Mountains, chap. Simulation of avalanches in Bunderitsa valley with the RAMMS software, Intel Entrans, Sofia, 2024a.

Panayotov, M., Markov, K., Tsvetanov, N., Huber, A., Hesselbach, C., and Teich, M.: AVALANCHE HAZARD MAPPING USING THE AVALANCHE TERRAIN EXPOSURE SCALE (ATES) IN THE HIGH MOUNTAIN RANGES OF BULGARIA, in: Proceedings, International Snow Science Workshop 2024, Tromsø, Norway, 2024b.

Pandey, S., Chand, S., Horkoff, J., Staron, M., Ochodek, M., and Durisic, D.: Design pattern recognition: a study of large language models, EMPIR SOFTW ENG, 30, https://doi.org/10.1007/s10664-025-10625-1, 2025.

Parks Canada Agency, G. o. C.: Avalanche Terrain Exposure Scale - Avalanche Terrain Exposure Scale, https://parks.canada.ca/pn-np/bc/glacier/visit/hiver-winter/ski/ates, last Modified: 2024-12-30, 2017.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, J MACH LEARN RES, 12, 2825–2830, 2011.

PirinNationalPark: PIRIN NATIONAL PARK – WORLD HERITAGE PROPERTY OF UNESCO, https://www.pirin.bg/?p=5591#more-5591, 2022.

Prasetiyowati, M. I., Maulidevi, N. U., and Surendro, K.: Determining threshold value on information gain feature selection to increase speed and prediction accuracy of random forest, Journal of Big Data, 8, 84, https://doi.org/10.1186/s40537-021-00472-4, 2021.

Probst, P., Wright, M., and Boulesteix, A.-L.: Hyperparameters and Tuning Strategies for Random Forest, WIREs Data Mining and Knowledge Discovery, 9, e1301, https://doi.org/10.1002/widm.1301, arXiv:1804.03515 [stat], 2019.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

Pugliese, R., Regondi, S., and Marini, R.: Machine learning-based approach: global trends, research directions, and regulatory standpoints, Data Science and Management, 4, 19–29, https://doi.org/10.1016/j.dsm.2021.12.002, 2021.

Python Software Foundation: Python Language Reference, version 3.9.21, https://www.python.org/.

Rainio, O., Teuho, J., and Klén, R.: Evaluation metrics and statistical tests for machine learning, SCI REP-UK, 14, 6086, https://doi.org/10.1038/s41598-024-56706-x, publisher: Nature Publishing Group, 2024.

Ramadhan, M., Sitanggang, I., NASUTION, F., and GHIFARI, A.: Parameter Tuning in Random Forest Based on Grid Search Method for Gender Classification Based on Voice Frequency, DEStech Transactions on Computer Science and Engineering, https://doi.org/10.12783/dtcse/cece2017/14611, 2017.

Rimal, Y., Sharma, N., and Alsadoon, A.: The accuracy of machine learning models relies on hyperparameter tuning: student result classification using random forest, randomized search, grid search, bayesian, genetic, and optuna algorithms, MULTIMED TOOLS APPL, 83, 74 349–74 364, https://doi.org/10.1007/s11042-024-18426-2, 2024.

Rogers, J. and Gunn, S.: Identifying Feature Relevance Using a Random Forest, in: SLSFS: International Statistical and Optimization Perspectives Workshop "Subspace, Latent Structure and Feature Selection", edited by Saunders, C., Grobelnik, M., Gunn, S., and Shawe-Taylor, J., pp. 173–184, Springer, Berlin, Heidelberg, ISBN 978-3-540-34138-3, https://doi.org/10.1007/11752790_12, 2006.

Salman, H. A., Kalakech, A., and Steiti, A.: Random Forest Algorithm Overview, Babylonian Journal of Machine Learning, 2024, 69–79, https://doi.org/10.58496/BJML/2024/007, 2024.

Saraswat, P.: Supervised Machine Learning Algorithm: A Review of Classification Techniques, Smart Innovation, Systems and Technologies, 273, 477–482, https://doi.org/10.1007/978-3-030-92905-3_58, 2022.

Sarker, I. H.: Machine Learning: Algorithms, Real-World Applications and Research Directions, SN Computer Science, 2, 160, https://doi.org/10.1007/s42979-021-00592-x, 2021.

Schmudlach, G. and Köhler, J.: Method for an automatized Avalanche Terrain Classification, in: Proceedings, International Snow Science Workshop, Breckenridge, Colorado, 2016, pp. 729 – 736, Breckenridge, Colorado, https://arc.lib.montana.edu/snow-science/objects/ISSW16_P2.04.pdf, 2016.

Schumacher, J., Toft, H., McLean, J. P., Hauglin, M., Astrup, R., and Breidenbach, J.: The utility of forest attribute maps for automated Avalanche Terrain Exposure Scale (ATES) modelling, SCAND J FOREST RES, 37, 264–275, https://doi.org/10.1080/02827581.2022.2096921, 2022.

Schweizer, J. and Lüschg, M.: Characteristics of human-triggered avalanches, COLD REG SCI TECHNOL, 33, 147 – 162, https://doi.org/10.1016/S0165-232X(01)00037-4, 2001.

Sharp, A. E. A.: Evaluating the exposure of heliskiing ski guides to avalanche terrain using a fuzzy logic avalanche susceptibility model, Masters' thesis, University of Leeds, School of Geography, 2018.

Sharp, E., C., C., Statham, G., and Schroers, B.: A standardized, multiscale, fuzzy spatialdata model for a Avalanche Terrain Exposure Scale mapping, in: Proceedings, International Snow Science Workshop, Bend, Oregon, 2023, pp. 838 – 845, http://arc.lib.montana.edu/snow-science/item/2974, 2023.

Spannring, P.: Comparison of two avalanche terrain classification approaches: Avalanche Terrain Exposure Scale - Classified Avalanche Terrain, Master's thesis, University of Innsbruck, Faculty of Geo- and Atmospheric Sciences, 2024.

Statham, G. and Campbell, C.: The Avalanche Terrain Exposure Scale (ATES) v.2, NAT HAZARD EARTH SYS, 25, 1113–1137, https://doi.org/10.5194/nhess-25-1113-2025, publisher: Copernicus GmbH, 2025.

https://doi.org/10.5194/egusphere-2025-2143
Preprint. Discussion started: 2 June 2025

Statham, G., McMahon, B., and Tomm, I.: The Avalanche Terrain Exposure Scale, in: Proceedings, International Snow Science Workshop 2006, Telluride, Colorado, 2006.

Sykes, J., Toft, H., Haegeli, P., and Statham, G.: Automated Avalanche Terrain Exposure Scale (ATES) mapping – local validation and optimization in western Canada, NAT HAZARD EARTH SYS, 24, 947 – 971, https://doi.org/10.5194/nhess-24-947-2024, 2024.

Techel, F. and Zweifel, B.: Recreational avalanche accidents in Switzerland: Trends and patterns with an emphasis on burial, rescue methods and avalanche danger, in: Proceedings International Snow Science Workshop 2013, Grenoble, France, https://arc.lib.montana.edu/snow-science/item.php?id=1844, 2013.

Techel, F., Jarry, F., Kronthaler, G., Mitterer, S., Nairz, P., Pavšek, M., Valt, M., and Darms, G.: Avalanche fatalities in the European Alps: long-term trends and statistics, Geographica Helvetica, 71, 147–159, https://doi.org/10.5194/gh-71-147-2016, publisher: Copernicus GmbH, 2016.

Toft, H., Sykes, J., Schauer, A., Hendrikx, J., and Hetland, A.: AutoATES v2.0: Automated Avalanche Terrain Exposure Scale mapping, NAT HAZARD EARTH SYS, 24, 1779 – 1793, https://doi.org/10.5194/nhess-24-1779-2024, 2024.

Tsvetanov, N.and Panayotov, M.: Avalanches in the Bunderitsa valley in Pirin Mountains, chap. Dendrochronologicalreconstruction of large magnitude avalanches in Bunderitsavalley, Intel Entrans, Sofia, 2024.

Ul Hassan, C. A., Khan, M. S., and Shah, M. A.: Comparison of Machine Learning Algorithms in Data classification, in: International Conference on Automation and Computing (ICAC), 2018.

Varoquaux, G. and Colliot, O.: Machine Learning for Brain Disorders, chap. Evaluating Machine Learning Models and Their Diagnostic Value, Springer US, 2023.

Veitinger, J., Purves, R. S., and Sovilla, B.: Potential slab avalanche release area identification from estimated winter terrain: a multi-scale, fuzzy logic approach, NAT HAZARD EARTH SYS, 16, 2211 – 2225, https://doi.org/doi:10.5194/nhess-16-2211-2016, 2016a.

Veitinger, J., Sovilla, B., and Purves, R. S.: Model code of release area algorithm, https://github.com/jocha81/Avalanche-release, 2016b.

Venkadesh, P., S. V., D., Marymariyal, P., and Keerthana, S.: Predicting Natural Disasters With AI and Machine Learning, IGI Global Scientific Publishing, 2024.

Viallon-Galinier, L., Hagenmuller, P., and Eckert, N.: Combining modelled snowpack stability with machine learning to predict avalanche activity, The Cryosphere, 17, 2245–2260, https://doi.org/10.5194/tc-17-2245-2023, publisher: Copernicus GmbH, 2023.

von Avis, C. D., Sykes, J. M., and Tutt, B.: Development of large scale automated Avalanche Terrain Exposure Scale (ATES) ratings in collaboration with local avalanche experts, in: Proceedings, International Snow Science Workshop, Bend, Oregon, 2023, pp. 982–988, http://arc.lib.montana.edu/snow-science/item/2998, 2023.

Wang, S., Zhuang, J., Zheng, J., Fan, H., Kong, J., and Zhan, J.: Application of Bayesian Hyperparameter Optimized Random Forest and XGBoost Model for Landslide Susceptibility Mapping, FRONT EARTH SCIFRONT EARTH SCI, 9, https://doi.org/10.3389/feart.2021.712240, publisher: Frontiers, 2021.

Werners, B.: Mathematical Models for Decision Support, vol. 48 of *NATO ASI Series*, chap. Aggregation models in mathematical programming, pp. 295 – 305, Springer, Berlin, Heidelberg, 1988.

Yolanda, A., Adnan, A., Ell Goldameir, N., and Rizalde, F.: The Comparison of Accuracy on Classification Data with Machine Learning Algorithms (Case Study: Human Development Index by Regency/City in Indonesia 2020), in: AIP Conference Proceedings, 2023.