JOURNAL OF SPATIAL INFORMATION SCIENCE
Number 25 (2022), pp. 67–88 doi:10.5311/JOSIS.2022.25.204

RESEARCH ARTICLE

Modeling walkability by remote sensing as latent walking speed extracted from multiple digital trail maps

Ljiljana Šerić, Marina Tavra, Ivan Racetin, and Antonia Ivanda

Faculty of Electrical Engineering, Mechanical Engineering, and Naval Architecture, University of Split, Croatia
Faculty of Civil Engineering, Architecture, and Geodesy, University of Split, Croatia
Faculty of Civil and Geodetic Engineering, University of Ljubljana, Slovenia

Received: December 26, 2021; returned: March 15, 2022; revised: May 17, 2022; accepted: June 17, 2022.

Abstract: Coordinating and managing teams searching for missing persons in wilderness areas is challenging. Local terrain characteristics and environmental conditions strongly influence how searchers accomplish their search tasks. When making decisions, searchers consult various maps of the area. In this paper we proposed a methodology for mapping characteristics of the area that influence user behavior when walking the area, and define a walkability model of the terrain. We define walkability as a measure of how fast a person can walk through terrain. The observed walking speed depends on factors such as the fitness and motivation of a person walking through the terrain, as well as on assistive features and the configuration of the terrain. In our method, walkability is predicted only as a feature of terrain configuration. We used singular value decomposition (SVD) to transform datasets to extract latent features of the terrain and users from multiple Global Positioning System (GPS) trails. We define the walkability measure as a latent component of walking speed, which is a function of terrain features. Finally, we use a polynomial regression algorithm to build a model for predicting terrain walkability based on remote sensing imagery from the Sentinel-2 mission. The application of the proposed model is demonstrated in the Kozjak mountain region in the Republic of Croatia.

Keywords: digital trails, gpx, recommender system, SVD, search and rescue, DSS

1 Introduction

Spatial data play an important role in missing person search planning [7]. Decisions usually depend on the use of spatial data in combination with other data, such as movement speed [41] and personal characteristics, behavioral profiling [40] or the condition [21] of the missing person, which can greatly facilitate the planning of missing person searches included in search and rescue campaigns (SAR). This also applies to spatial data such as land use features or terrain interpretation. This problem is particularly pronounced in uninhabited and isolated terrain, which is often not mapped up to date in accordance with monthly natural changes in the environments, such as wilderness.

Coordinating and leading teams searching for missing persons in wilderness is challenging [15]. Local terrain features [10] and environmental conditions greatly influence how searchers accomplish their search tasks. The wilderness characteristics make such areas heterogeneous, and it is hardly possible to decide how much time straight forward is needed to walk over that terrain without spatial interpretation. Many studies today have focused on classifying the behavior of lost individuals based on the speed of their movements. However, one of the most common categorizations of movement in search and rescue (SAR) is walking speed by age [46]. Therefore, the index set for the movement distance factor uses age-specific walking speed as defined in [39].

There are numerous studies on the walkability of urban areas [42], for example in the context of urban policy, planning [31], or investment [11]. Here, the input parameters for calculating walkability are differ from those in the wild, i.e., more input data are available (infrastructure data, green areas, street design, sight distance etc.). The calculation of walkability in the wild is specific, i.e. no parameters such as terrain characteristics may be omitted. We can better estimate the movement of individuals in the wilderness by considering terrain characteristics such as slope [44] on typical terrain or even according to Naismith's rule [16]. However, information about how far a given person can walk in predefined time is critical when people are lost or have an accident. Therefore, SAR operations conducted by search and rescue teams are critical in these situations [25]. When a missing person incident is reported, a team critically assembles and conducts SAR activities to find the location of the lost person [45]. Each such activity is individual and has its challenges, but the experience of the SAR team and team leader [2] can be an advantage in critical situations [36]. SAR Team members are tasked with searching a specific area, and the team leader leads the effort based on simultaneously provided information [2].

One way to make information available is through the decision support system (DSS). DSS can be useful for faster and optimal disaster response planning by providing suggestions based on data, models, and artificial intelligence [33]. One of the SAR situations is the obvious benefit of using artificial intelligence in decision support systems related to missing person management [22]. In such situations, a quick response is required for the search to be successful, which can be supported by prepared models from large data sets [3]. Recommender systems are a machine learning-based technique commonly used in industry to match users with items that the system assumes they would like. In [28], the authors propose a method for a recommender system based on matrix factorization. This method is also the engine behind the award-winning Netflix algorithm [27]. However, this algorithms usefulness stretches beyond movie recommendations. An application of matrix factorization methods to spatial data is also described in [1]. The authors used SVD-based PCA to derive principal components for burned area assessment from remote sensing data.

The increasing availability of remote sensing data, such as satellite imagery, is being used to obtain highly detailed information about the terrain. Several studies have focused on the evaluation of environmental models, such as green spaces and urban areas that affect walkability. For example, in [32], the authors propose a method for walkability of urban green spaces using Advanced land observing satellite (ALOS) data, while in [43], the authors also used ALOS satellite imagery, but for implementing a user-friendly walkable area detection assistance system, where they calculate the green space index based on the spectra recorded in the satellite data. However, the available literature neglects the problem of mapping walkability in wilderness, even though a walkability map could be an important component of many GIS based decision support systems, not only for lost person search and rescue, but also for many other activities that take place in wilderness, such as hiking, hunting, afforestation planning, pasture planning, and the like.

According to SAR [9], the initial planning point (IPP) is the starting point around which a search is planned. If known, it is the point where the lost person was last seen. In the literature, it is also referred to as the point last seen (PLS) [2]. Knowledge of terrain features that affect walkability is necessary for assessing search zones and can be extremely useful in planning search operations. Knowledge of walkability can be beneficial in predicting the distance the missing person could have traveled in the time since they went missing. This distance is used to define the search radius. Knowing the time it will take searchers to traverse the area helps SAR managers plan activities and manage personnel. Both values—the distance the missing person could walk and the time it takes searchers to walk—are functions of walkability. However, the speed of walking is individual. If we consider the speed of movement GPS data [41] from past SAR campaigns and the characteristics of the terrain using satellite imagery, we could obtain a more complete and comprehensive picture of terrain accessibility or walkability as one of the parameters of SAR planning.

In this paper, we propose a novel approach to use multiple trails for evaluating the accessibility of a terrain. The mathematical tool of matrix factorization allows us to divide the walkability achieved on a terrain into two factors: Factor of the terrain and Factor of the user, which are latent values. The proposed method represents a novel algorithm that uses the behavior of multiple users on multiple instances of a terrain to describe the features of the terrain associated with the user's possible behavior. Unlike traditional approaches where authors rely on average user behavior, we extract latent values of the terrain that may be indicative of user behavior. In addition to proposing the new method, the contribution of this paper is a map of walkability values of the study area obtained by training a walkability prediction model using surface reflectance measured by satellite observations and topography data. Thus, in this paper, we propose an approach based on latent spatial transformation to evaluate the walkability of the terrain, which is crucial for determining the most likely search area using a decision support system. We used singular value decomposition (SVD) for a dataset transformation and a polynomial regression algorithm to predict the terrain walkability using remote sensing images from the Sentinel-2 mission.

This article is organized as follows. First, we introduce the terrain walkability characteristic of the study area and use a data set which consists of digital trails, DEM map and Sentinel-2 imagery. Second, we describe used methods related to GPS preprocessing, latent value extraction using the singular value decomposition and latent value prediction from remote sensing data. Finally, in the 4 and 5 sections, we summarize, analyze, and discuss the results for assessing the walkability of the terrain. The final section draws conclusions and presents plans for future work.

2 Materials

2.1 Study area

The study area for this research is the northern slope of the Kozjak mountain (Figure 1). Kozjak is one of the longest continuous ridges in the Republic of Croatia. This area is not inhabited, its slopes are not so steep and form a plateau that gradually descends towards the Dalmatian Zagora. Below the very steep and difficult to access southern slopes is the town of Kaštela with over 40,000 inhabitants [26].

Depending on which side we look at, different natural biotopes characterize Kozjak. In the north of Kozjak, widespread hazmophytic rock vegetation called Dalmatian limestone develops in the cracks of the dry limestone cliffs. The deciduous species of white hornbeam and black hornbeam, which characterize the sub-Mediterranean flora, are also prevalent in this area. On the southern slopes of Kozjak, the forest community of Aleppo pines and holm oaks, as well as broom and moss, is pronounced [8].

Near the study area are rural areas with sparse population and mostly abandoned agricultural land. The reason for selecting this particular area is its biodiversity and topographic diversity. In this area there are both uphill and downhill sections covered with different vegetation. The selected area is close to settlements, but is still natural and uninhabited, and is only occasionally visited by hikers and cyclists. In the recent past, there have been four missing persons cases and, accordingly, search and rescue operations have been carried out in this area, archiving the traces of the searchers GPS. However, we will not discuss the missing persons cases due to data confidentiality. The area was searched many times by multiple searchers and search dogs and we obtained dense walking trails. This makes the area suitable for analysis.

2.2 Walking trails

We collected a dataset made of a digital map provided by the Croatian Mountain Rescue Service (CMRS) and collected from random incidents of lost people SAR. We used a digital trail map based on GPS technology for this research. We used GPS tracking exported to GPX (GPS eXchange), the standard XML data format for digital trails. The anonymized data from real SAR missions consist of GPX trail: points, each with associated geocoordinates (latitude and longitude), altitude and time. The provided GPS dataset consists of 153 .gpx files with 126481 segments. The total walked length of the trails was 816377 m. The time span of SARs for a lost person refers to seven dates in 2018 and one in 2019. Each trail file was associated with a user based on the file name. The file name was manually assigned by the search and rescue team member responsible for mapping. The file name consists of the name, last name, or nickname of the person who owns the GPS device and a number to make the file unique. We've found that there are files with the same user name but created at different times.

2.3 Digital maps

Digital elevation maps (DEM) contain a variety of information to represent basic topographic features [29]. From untreated DEM, we can derive additional terrain features useful for defining search performance and mobility models [20]. We consider the slope within

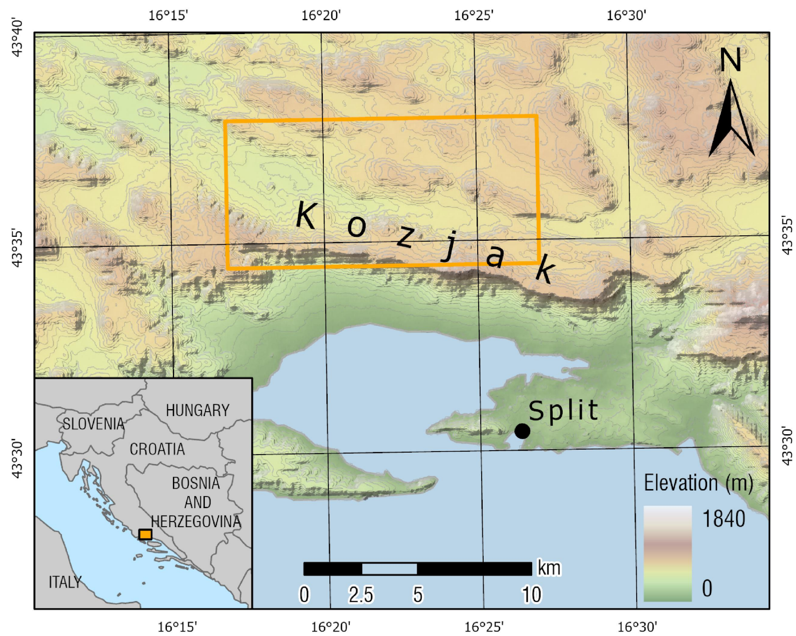

Figure 1: The northern slope of the Kozjak Mountains in the Republic of Croatia was selected as the study area for this study. The orange rectangle indicates the spatial extent of the results shown in Figure 7.

these derived features, computed as a function of the maximum rate of change between adjacent points. In our work we used European digital elevation model (EU-DEM) [13], version 1.1 which is available as a raster 32 bit GeoTiff file and with spatial resolution of 25 m.

2.4 Sentinel-2 imagery

The type of surface coverage of the terrain can be estimated from multispectral images. Various types of cover absorb various portions for each wavelength of the light, surface reflectance of multiple bands can be used to estimate land cover features. In this study, we selected freely available images captured by satellite mission Sentinel-2. Sentinel-2 is a European earth polar-orbiting satellite constellation consisting of two identical satellites Sentinel-2A and Sentinel-2B flying on a single orbit plane but phased at 180°. Each of these satellites hosts a multi-spectral instrument (MSI), which covers the visible to the shortwave infrared spectral range. This mission provides high-resolution imagery for the global and sustained monitoring of Earth land and coastal areas with a high revisit frequency of 5 days [19]. MSI sensor delivers 13 spectral bands which are listed in Table 1 with different central wavelengths, widths and spatial resolutions (10 m, 20m and 60m) [12]. According to [38] Sentinel-2 provides high-resolution satellite data which has been successfully applied to land cover/use, both for monitoring and classification. Furthermore, each spectral band has its own purpose where [14, 30]:

- B01 is often used for retrieving coastal aerosol,
- B02, B03, and B04 are standard blue,green and red channels, where e.g. B02 is sensitive to vegetation senescing, B03 is sensitive to total Chl-a in vegetation, and B04 is used for maximum Chl-a absorption,
- B05, B06, B07, B08,and B8a are vegetation bands, suitable for mapping shorelines, biomass content, detecting/analyzing vegetation and atmospheric corrections,
- B09 is used for water vapour absorption,
- B10 is used for detection of thin cirrus (atmospheric correction),
- B11 and B12 are useful for measuring the moisture content of soil and vegetation, also they are used for snow, ice, and cloud separation.

Spectral band	**B01**	**B02**	**B03**	**B04**	**B05**	**B06**	**B07**	**B08**	**B8a**	**B09**	**B10**	**B11**	**B12**
Central wavelength λ (nm)	443	490	560	665	705	740	783	842	865	945	1375	1610	2190
Spectral width λ(nm)	20	65	35	30	15	15	20	115	20	20	30	90	180
Spatial resolution (m)	60	10	10	10	20	20	20	10	20	60	60	20	20

Table 1: Sentinel-2 spectral bands specifications.

3 Methods

3.1 GPS preprocessing

Trails were processed and transformed into segments. A segment describes the walk between two points, and each segment is defined with start and end points and start and end times. Using this information, we can easily calculate the necessary features of a segment, such as the length of the trail, the slope of the terrain (if elevation above sea level is available for the start and end points), and the speed of walking the segment. In this work, however, the information about the slope was not calculated from the GPS trails, but by processing DEM.

The data was filtered for inconsistencies. In the cases where the device was turned off and on again, there was a long period of time between two points. This can also happen when the signal is lost due to the terrain of the test area. Therefore, we excluded all segments where the time difference between two points was more than one minute and the distance between two points was more than 30m. Each trail was saved as a file. The file name of the trail was composed of the date it was recorded, the name of the person holding the device, and a number that made the file name unique because there were multiple missions. If the device was attached to a search dog, the name also included the word "pas" (Croatian for dog). The numbers were not consecutive, but were randomly assigned, which only made the file name unique. There were 153 different files with unique names. We processed the string with the file name, removed the numbers from the string, and found 74 unique names. For a terrain instance, we calculated the speed score for each user by calculating the average speed of all segments within the terrain instance.

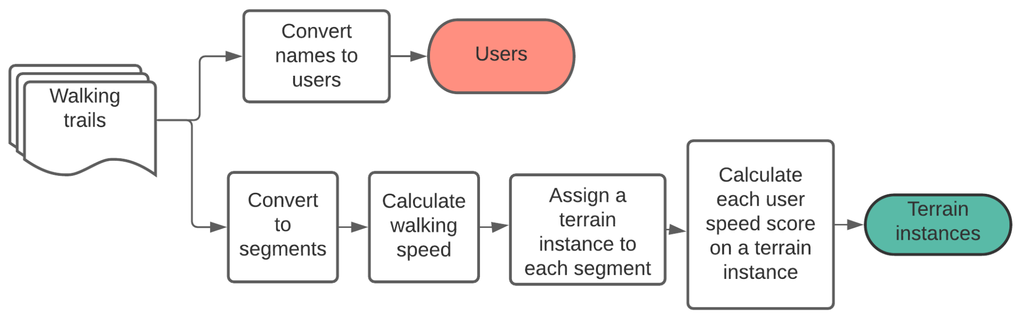

Figure 2: Pre-processing walking trails to obtain speed score for each user on visited terrain instances.

3.2 Latent value extraction with SVD

A mathematical construct that directly reveals the rank and corresponding ideal basis of a dataset is the singular value decomposition (SVD) [37]. The SVD of a matrix is a factorization of that matrix into three matrices.

$$A = U\Sigma V^T \tag{1}$$

Where A represents the matrix to factorize with m rows and n columns, U is a mxm matrix that consists of eigenvectors of AA^T, V matrix is a matrix nxn matrix of eigenvectors of A^TA and Σ is a diagonal matrix of singular values of A. Practical usage of SVD originates from the fact that for a dataset in n-dimensional space, for any k < n, the SVD will show the ideal basis for representing that data using only k dimensions. If the SVD reveals that the dataset is full rank and no feature reduction is possible along the calculated axes, then no axes exist for which a reduction is possible. Practical usage of SVD for the transformation of matrix A would include the following semantics: A= matrix of data to be transformed, rows of matrix A represents unique users columns of matrix A correspond to unique items elements of matrix A are denoted with $a_{u,i}$ and represent users u rating of item and i SVD transformation of a matrix. A will reveal such subspace, often referred to as latent space, where users and items can be described with fewer features. Particularly, matrix product $U\Sigma$ performs transformation of users to latent space and ΣV performs transformation of items in latent space.

SVD and matrix factorization is the basis for many recommender systems. A recommender system is a technique for predicting user rates on items based on previous user ratings. User rating matrix is a typical sparse matrix holding only a limited number of non-zero item ratings for each user. A matrix factorization-based recommender system builds a model for both—users and items. Finally, the recommender system uses reported ratings from similar users to recommend items to a user. The intuition behind this work is that variables can describe both users and items in latent space for building predictions.In our work, we use algorithms typically used for predicting user rating on movies [27], in this case for predicting users' walking speed on a terrain segment. In this work we will use implementation of the algorithm proposed by Funk [17] as the most common recommender system algorithm. Prediction of user ratings for the item (r_{ui}) of the algorithm is expressed as:

$$r_{ui} = \mu + b_u + b_i + q_i^T p_u \tag{2}$$

where μ is total bias, b_u is a user bias, b_i is item bias and q_i is a vector of an item represented in latent space and p_u is user latent space representation. In other words, b_u and p_u are parameters associated with the user model, while b_i and q_i are parameters associated with the item model. To implement the latent space model for the terrain, we expressed our data set in the form of matrix A. Rows of matrix A represent the users walking the terrain, and columns are different segments of 30m length. The reason for selecting 30m length of a segment is a trade off between resolution of the resulting walkability map we are aiming to build and resolution of input data. As shown in Table 1 resolution of satellite data ranges from 10m to 60m, while resolution of DEM data is 30m. Thus 30m was selected to be the resolution of the walkability map.

Values of matrix A, instead of ratings , are filled in with the walking speed score. In most cases, GPS track contains more precise data and records points closer than 30m that compose a terrain instance, so we calculate the median speed of walking of the observed user. Speed score is the median speed of walking of the observed user in the terrain instance. The constructed matrix is sparse. We use the implementation of the SVD prediction algorithm provided by [23]. We used a prepared data set to train all factors from equation (2) and that is μ, b_u, p_u, b_i, and q_i. The calculated factors are used for prediction of ratings of known users for known items and for filling the sparse matrix A with unknown values. However, the algorithm we used was critiqued by [18] for its instability, mostly due to a phenomenon called 'popularity bias' where popular items are rated more often, while unpopular items are rarely rated. However, due to the nature of our dataset, we can plausibly excluded the popularity bias. Unlike hikers and climbers who only walk on trails, rescuers walk on both accessible trails and inaccessible terrain in search of the lost person.

To test the stability of the results, we performed iterative training cycles in which we varied the parameters of the algorithm and compared the resulting factors. We did not find any significant differences between the factors as a function of the initialization values.

3.2.1 Definition of walkability

From the five factors obtained from the dataset, we select only those associated with the terrain instance features. We define walkability as the sum of values of the μ, b_i, and q_i factors.

- The factor μ can be understood as an average value for walking speed, to which we can add or subtract the terrain and user components of walking speed. This factor is the same for the entire data set—regardless of terrain instance and user.

- The factor b_i is the item bias (often referred to as intercept) of the speed of walking on the terrain instance. b_i factor is a feature of the terrain instance representing the average difference (can be positive or negative) of the speed of walking on the specific terrain instance and average speed of walking of the whole dataset μ.
- The q_i factor is the factor that needs to be multiplied by the user factor in order to predict the speed of the specific user on the terrain instance.

To define a value that quantitatively describes the walking speed possible in the terrain independent of the user, we calculate the walkability value as the sum of μ, b_i, and q_i.

$$w = \mu + b_i + q_i \tag{3}$$

It's important to understand that the walkability value is not the average walking speed, but the part of the possible speed that depends only on the terrain features. Depending on the user and terrain characteristics, the walking speed may be increased on certain sections of the terrain for one user and decreased on others. We do not assume which features of the terrain define walkability and which features of the users are described in the latent space. Therefore, we consider walkability as the walking speed of an unspecified user. In this part of the methodology, we're concerned with measuring the value of walkability using several existing trails. Our evidence for walkability is the walking speed of multiple users. We extract from multiple users' trails only the part of walkability that depends on the terrain. Some users walk faster on the terrain, others slower, but this depends on the user characteristics. The user characteristics are neglected in the rest of the analysis because we're only concerned with predicting the walkability of the terrain. However, for the prediction of the user's walking speed, we'll also use the user's biases and factors.

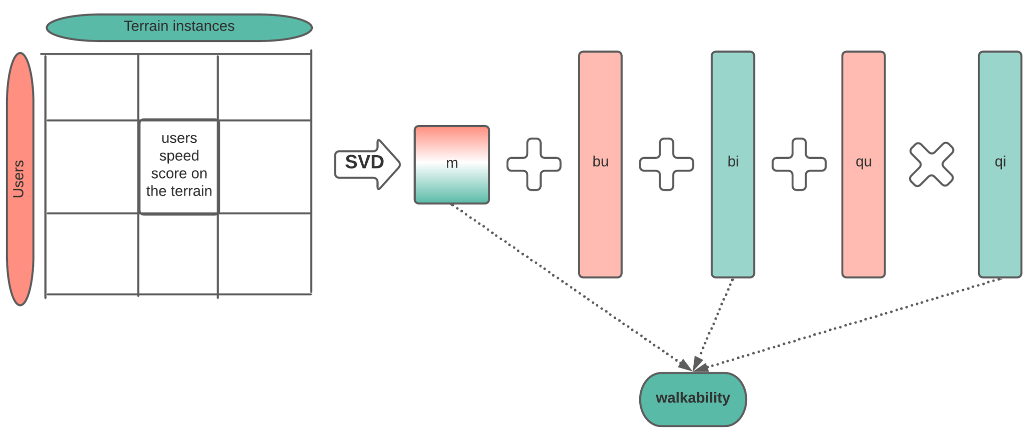

Figure 3: Extracting walkability values as latent values from the speed score matrix of users and terrain instances.

The values of the extracted walkability are shown in Figure 4. The value of walkability is expressed as a numerical value without units. Since it represents the latent value of the walking speed matrix, in order to predict the speed of walking the terrain, walkability needs to be multiplied with users' latent value. Thus, we expect the value of walkability to be higher on smooth flat terrain and lower on the rough inclined terrain. From the figure, it can be seen that the roads have a higher walkability value, while the mountain peaks have a low walkability value.

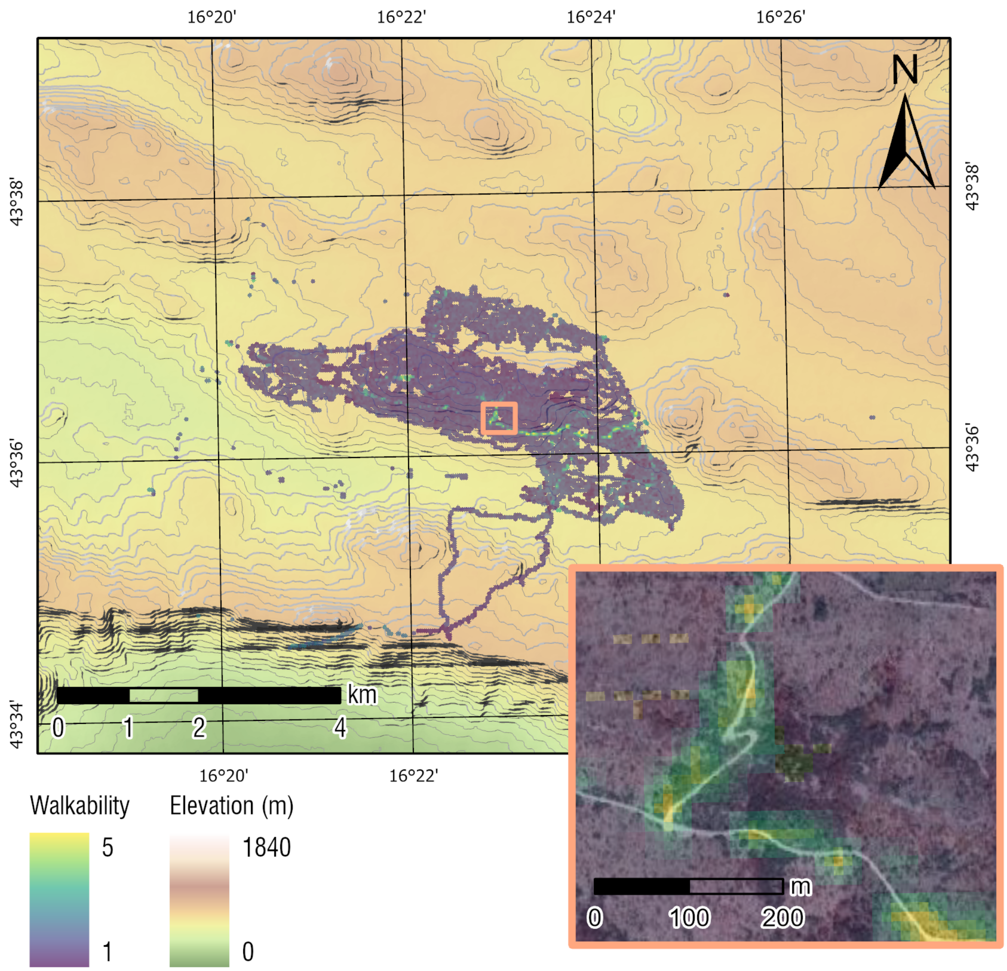

Figure 4: Walkability values shown on map.

Walkability is extracted only for terrain instances covered by trails. We see that the walkability value is higher on roads than on dense terrain.

3.3 Latent value prediction from remote sensing data

Using the method described earlier, we extracted the quantitative walkability value of trails. As can be seen in Figure 4, the known values are assigned to the walked segments of the terrain, but only for the parts of the terrain that were walked by the users whose trails were available to us.

To obtain the map of the entire terrain, we need a walkability prediction model that can be applied to the entire map. Our hypothesis is that the walkability of the terrain depends mainly on the surface conditions and the topography of the terrain. Therefore, we use two available data sources that can be associated with the two sets of features.

We downloaded high-quality band images for date 02-04-2021 from the Sentinel Hub EO browser, except for the band image for B10, which was not available at the time of writing. Cloud cover for the scene is 15.7%, but there are no clouds over the study area. All images are in *.tiff* format. For each terrain, we extracted 12 associated values—one value for each band. In addition, we extracted the context of each cell by calculating the average value of the band of surrounding cells with radius three.

The terrain topography was extracted from the digital elevation model. We calculated the slope of the terrain for each cell under study—a terrain instance and used the value as an additional feature of the terrain. The slope value was not taken as an angle, but in accordance with [44] we calculated the slope factor:

$$slopefactor = e^{-3.5|\frac{dh}{dx}+0.05|} \tag{4}$$

where *dh* is the maximum difference in height above sea level between start and end point in the cell, and *dx* is the length of the segment.
Finally, we obtained the dataset constructed from:
12 band values | 12 average band values | slope factor | w

3.3.1 Polynomial regression model

The relationship between the selected and measured walkability was not linear, so we used polynomial regression as a model to predict walkability based on 12 band values, 12 averaged band values, and the slope factor value.

Polynomial regression [35] is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is assumed to be curvature, so it is modeled as an n-th degree polynomial in x. In multiple polynomial regression a dependent variable y depends on multiple independent variables x_1, x_2 âĂę x_n.

Multiple polynomial regression is performed by calculating polynomial features and then applying linear regression for predicting the dependent variable from multiple independent variables.

Polynomial features are calculated as n-th product of each variable, i.e., for the two variables x_1 and x_2 polynomial features of order two would be: (x_1,x_2) –> (x_1,x_2,x_1^2,x_2^2, x_1*x_2)

Polynomial features of order two were calculated from 25 available features. With this expansion of polynomial features, we are able to develop a more sophisticated model and fit the coefficients of polynomial regression to the obtained values of walkability.

The formula for predicting the walkability value of a terrain instance depends on the Sentinel-2 reflectance of all bands and the topography. After fitting the formula, the feature-dependent values can be easily obtained for the whole area, not only for the terrain instances for which we have walkability values.

By applying the raster calculation formula to the entire map, we obtain a map of walkability, including the parts of the area that have not been walked.

4 Results

The dataset obtained from the search and rescue authority contained 153 walking trails inside the study area. The files were analysed, and we identified 74 different subjects in the filenames. The trails were processed, and transformed into the list of segments of walking between two points, and we calculated the speed of walking between the two points in the segment. This resulted in 126481 segments.

We created a grid of terrain instances with 30m cell size (30m30m). Each segment belonged to one terrain instance. We calculated for each subject its median speed (expressed in km/h) inside the terrain instance cell to represent the speed score. This resulted in 32619 instances of speed scores each associated to one user on one terrain instance.

With this data we performed SVD in order to calculate parameters of the SVD model: μ, b_u, p_u, b_i, and q_i.

The trained SVD model was evaluated using root mean square error (RMSE) and mean absolute error (MAS). The RMSE score of the trained model was 1.378 and MAE score 0.804. We calculated walkability from the trained model parameters as defined in Equation 3. With our analysis we obtained the measure of walkability for 21139 terrain items with 30m cell size. Walkability values, calculated according to the equation 3, ranged from 0 to 7, with the majority of segments having walkability between 1 and 3. The distribution of values is shown in Figure 5. The lower values of walkability are assigned to terrain that is barely walkable, and the higher values are assigned to accessible terrain.

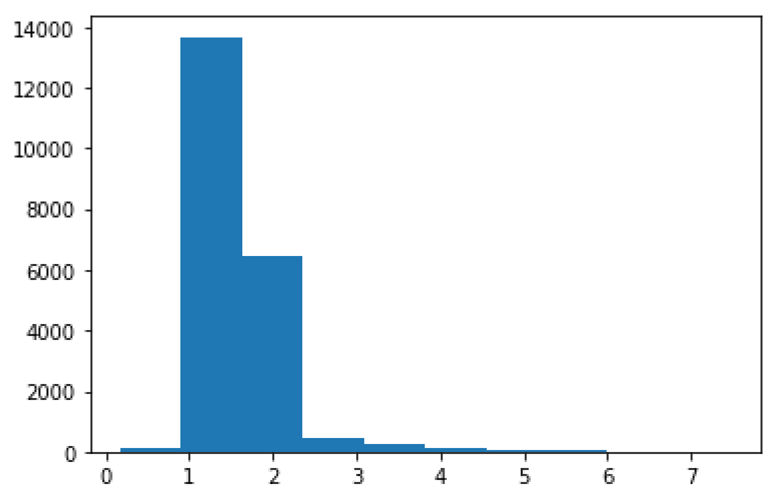

Figure 5: Distribution of walkability values calculated from walking trails in the study area.

When we compared the calculated walkability on terrain instances with a map, we noticed the walkability is higher in the cells with roads, and lower on slopes of the mountain, as can be seen on Figure 4.

In the next step we created a dataset for training a predictive model of walkability. For each terrain instance with cell size 30m we extracted the values of the 12 bands captured with Sentinel-2 instruments. Besides the central pixel value, we extracted the average value of the band in the surrounding cells with radius three. For each cell we calculated the maximal slope from the digital elevation model. We calculated polynomial features of order two to prepare a dataset for polynomial regression.

We split the dataset randomly into train and test portions with ratio 80:20. We trained a polynomial regression model to predict the walkability from the Sentinel-2 bands, averaged values of Sentinel-2 bands and a value calculated from the slope as described in Equation 4. The model was trained on 80% randomly selected data from the dataset.

For the trained model we calculated evaluation measures. Namely we used R^2 score, and RMSE.

The R^2 score or coefficient of determination [5] evaluates the performance of a regression model. It is equal to the proportion of the dependent variable variance that can be predicted from the independent variables and can be calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(x_i - y_i)^2}{\sum_{i=1}^{m}(\overline{y} - y_i)^2} \tag{5}$$

where y_i represents actual value of each sample, $\overline{y}$ is the mean value of actual variable y_i and x_i represents predicted values.

RMSE is a metric for showing uncertainty in a data set. It can be interpreted as average mean square error between the predicted value and actual value [24]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2} \tag{6}$$

where n represents the number of samples, y_i is the actual value of each sample and x_i is predicted value.

The trained model results: R^2 score for train set 0.165, and R^2 score for test set 0.099. RMSE measure for the train set was 0.301 and 0.388 for the test set. The R^2 score is positive, but not close to the ideal value of 1, so the results of the R^2 value do not show a strong correlation between the measured and predicted values. On the other hand, the RMSE value is relatively low, showing that the error of the predicted values has, on average,a small difference from the observed values. Looking at the error plot shown in Figure 6 we can see that the model underestimates the walkability of the terrain instance where the walkability is higher than 3 and predicts values from the range between 1 and 3, which is the range where most of the observed values were. However, we can see that the prediction follows the same trend as the observed values. This means that the prediction model predicts values that are incorrect in absolute values, but in the majority of cases predicts higher values for walkable segments, and low values for inaccessible segments. To test this hypothesis, we compared the prediction of the model with different classes of land cover of the study area.

Finally we used the obtained model for predicting walkability on the whole map of the study area. Resulting map is shown in Figure 7. The map shows the predicted walkability

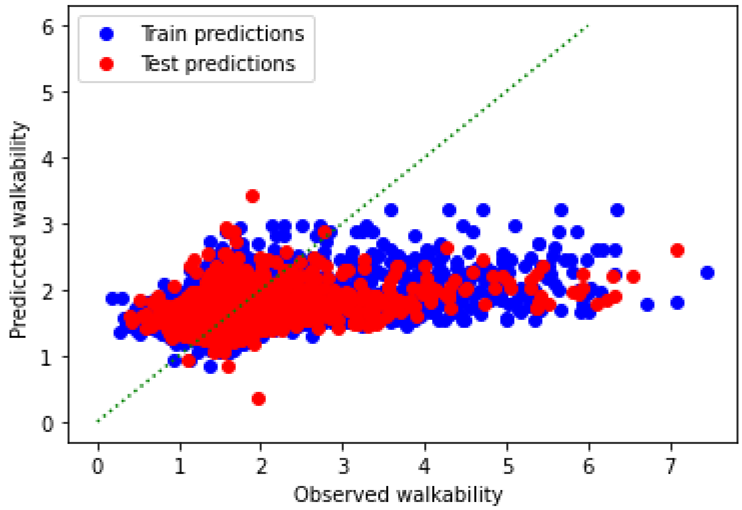

Figure 6: Error plot showing the error of the model between observed walkability and walkability predicted by our model for the obtained dataset.

for each pixel of the terrain. We can clearly distinguish roads and paths in the forest. The areas near mountain tops and with large slopes have lower value of walkability.

For assessment of the produced walkability model, a land cover map was derived from Sentinel-2 and OpenStreetMap (OSM) [34] data (Figure 7). Supervised random forest [4] classification algorithm was used to classify the satellite data in six classes: grassland, sparse vegetation, dense vegetation, soil, bare rock, and built-up. In order to improve the satellite-derived land cover map, roads, and pathways were extracted from the OSM dataset, rasterized, and overlayed to create the final map (Figure 7). We matched walkability values with land cover classes for each pixel of the two maps, and the resulting descriptive statistics are shown on (Figure 8).

The minimum walkability values were noted for sparse (0.9) and dense (0.8) vegetation, while maximum walkability for built-up (4.4) and sparse vegetation (4.6) classes. The highest average walkability for the analysed area was for built-up (1.8) and bare rock (2.0) classes. Sparse and dense vegetation exhibited the lowest average walkability of (1.6.). The lowest range of walkability values had bare rock (1.4) and grassland (2.1) classes, and the highest ranges were seen for built-up (3.3) and sparse vegetation (3.7) classes. Built-up class had the highest standard deviation of walkability (0.3) and grassland had the lowest (0.1).

5 Discussion

In the previous section, we presented a novel approach to using multiple digital trail maps to assess the walkability of the terrain. The results are based on walking trails data obtained from actual searches of lost people and terrain features obtained from remote sensing by the Sentinel-2 mission.

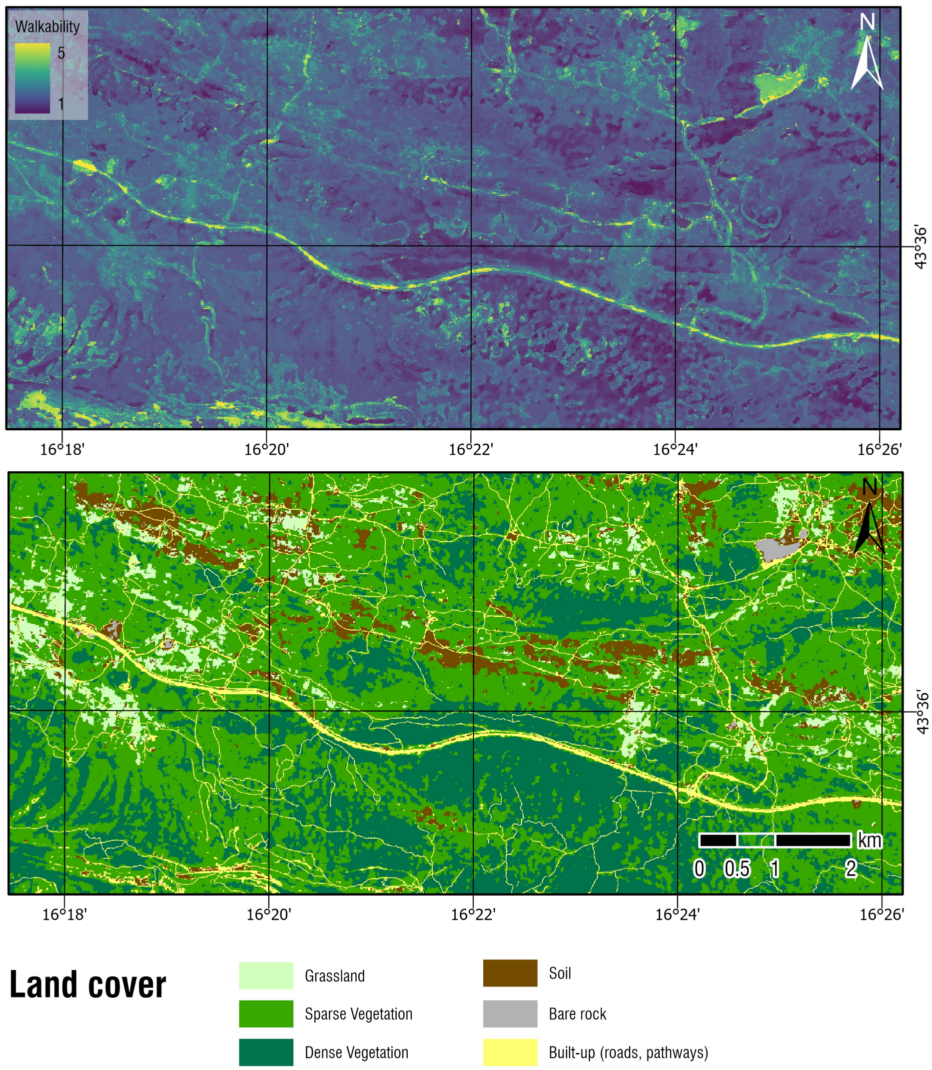

Figure 7: Walkability map for the Kozjak Mountains study area predicted by our model (top image). The thin gray lines represent the 25-meter elevation contours. Land cover map from Sentinel-2 and OpenStreetMap data (lower image).

We proposed a method to extract the walkability value as a latent value of the speed score matrix obtained by SVD. We proposed an equation to define the walkability as the sum of the terrain instance factors obtained by matrix factorization and calculated the walkability for known terrain instances. The polynomial regression model was trained with the values of the walkability of the known terrain to predict the walkability of the unknown terrain. The final results are evaluated by calculating two common evaluation measures: R^2 score and RMSE. The evaluation of the regression model showed a small R^2 value, which

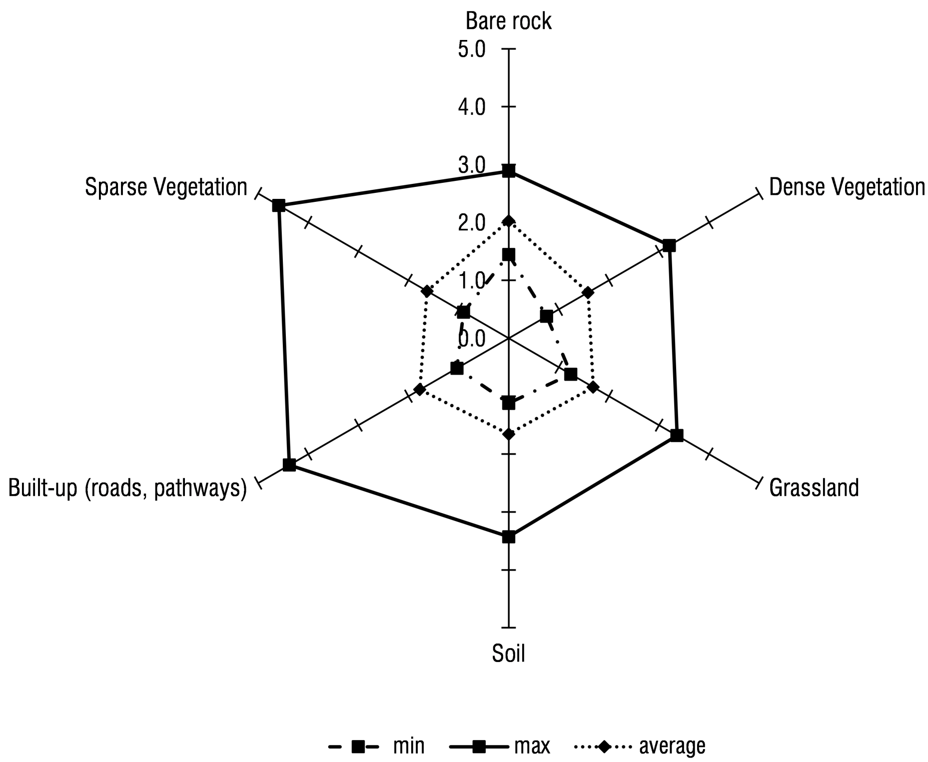

Figure 8: Chart with descriptive statistics of walkability values according to the land cover classes shown in the Figure 7.

does not indicate a high correlation between the predicted and measured values. However, the RMSE has a relatively small value, which means that, on average, the predicted value and the observed value do not differ much in absolute value. The error plot shows that our model sometimes underestimates the walkability value for observers with high walkability, but generally predicts larger values from the range where most walkability values are observed.

Even-though evaluation measures did not show our model has high accuracy, we do not think this is the reason to reject this approach for future research. As we know, both the SVD and polynomial models introduce errors into the prediction—the SVD model predicts the speed score with an RMSE of 1.37, and the polynomial model predicts the walkability with an RMSE of 0.30. If we also take into account that GPS devices measure the location with precision that is not optimal, we must acknowledge that we introduce errors on three levels (GPS measurement error, SVD transformation error, and polynomial regression error) and therefore cannot expect an accurate model. Since walkability, as we define it, is not a precisely measurable feature, it is expected that the proposed model will not be quantitatively accurate. Walkability is defined as a latent value of speed score, and as such cannot be measured directly. Speed score is calculated using GPS data, which is not precise. The first model (SVD model) attempts to measure walkability using speed scores from multiple users. The second model (polynomial model) measures the same feature using satellite data. Both models have errors, and therefore do not accurately predictthe values that were measured. However, when we compare the map with the final results to what we expected, we can observe that the predicted walkability value can distinguish between several classes of walkability.

Despite the less than ideal evaluation measures, the walkability map calculated for the entire study area gives us confidence that the model can be used to analyze the terrain (Figure 7). On the map, we can clearly identify roads and trails in the wilderness as areas with higher walkability. Since we assume that such areas are more walkable, we can say that this map can be used to evaluate the walkability of the terrain. Even though walkability is not used to represent the speed of walking, but only the difficulty of walking over a given terrain, and that is what this map does.

The walkability model can be used not only for planning the search for a missing person, but also for a variety of experts and researchers using spatial data for their work, without and with DSS integration. Walkability is an important feature to consider when planning hiking, hunting, afforestation, pasture, firefighting, and other activities that involve movement in wilderness.

As mentioned earlier, several studies aim to classify lost person behavior in terms of the speed of their movement, and this is the practice used to plan lost person search [46]. When searching in wilderness and non-urban areas, it is important to consider the nature of the terrain when planning the search [6] and determining the possible movement of the missing person.

In contrast to traditional approaches, where authors rely on average user behavior, we used the mathematical tool of matrix factorization, which allows us to divide the walking speed achieved on a terrain into two factors: Factor of the terrain and Factor of the user. The proposed method represents a novel algorithm that uses the behavior of multiple users on multiple terrain instances to describe the characteristics of the terrain associated with the possible behavior of the users.

Therefore, predictive models cannot be used to predict the location of the lost person, but only as a probabilistic tool to determine the area of the terrain to be searched where the probability of being found is high enough. In the context of this research, we consider walkability as a measure of the nature of the terrain in a given direction of movement from one point to another. Thus, the walkability of the terrain is a feature associated with a portion of the earth's surface that quantifies the potential walking speed.

When looking at the data from the digital trail maps used to calculate these maps, one should consider the different movement speeds under different conditions. For example, Figure 4 with the data from GPS shows that we had most of the data in locations with convenient to walk, e.g., roads, where walking speed is likely to be highest, which then correlated well with the maps from the satellite imagery. Therefore, the proposed model can be integrated into DSS's, but it should be considered that additional parameters should be included in certain terrains. Apart from the fact that the proposed model can be used for various GIS analyzes, it can also be implemented for other purposes and used for various applications. However, it is necessary to test the application in terms of the level of detail of the segments in order to improve the level of detail of the output maps.

6 Conclusions

Using the problem of lost persons in the wilderness as an example, we developed a model to improve search planning. We applied singular value decomposition (SVD) to a dataset transformation and polynomial regression algorithm to predict terrain walkability using remote sensing imagery from the Sentinel-2 mission.

We integrated a dataset created by collecting traces of different people who walked through the terrain during the search for missing persons, the terrain slope, and Sentinel-2 remote sensing data. The application of the proposed model is presented in the Kozjak mountain region in the Republic of Croatia. The results show that this approach can adequately model the terrain. The application of maps created with this modeling approach can be used for various purposes such as planning, forestry, mountaineering, scouting, decision support systems, etc. Specifically for search and rescue operations, the walkability map provides decision makers with information about the level of roughness of the terrain in order to plan how fast the lost person might walk on that terrain and how fast searchers can walk it. Walkability maps should be used in combination with the lost person model at SAR to predict the lost person's walking speed and behavior.

In this paper, we described a method to determine the latent value of the earth's surface from human behavior records. In addition, this paper proposed the methodology of dividing the influence of terrain and user characteristics into latent values by using SVD decomposition. The same method can be applied to other domains where user behavior varies from user to user and on different terrain instances. In evaluating the proposed model, we identified three levels at which errors occur and affect the final results. In future work, we plan to focus on improving the model by addressing all three sources of error. GPS signal variance estimation will be calculated to reduce the GPS error. To improve the walkability measurement, we'll include trails from different user groups such as scouts and hunters. We'll also look for additional terrain features to improve the walkability prediction model.

Acknowledgments

We would like to thank all the people and organizations who provided data, especially the Croatian Mountain Rescue Service (CMRS). This research was partially supported through project KK.01.1.1.02.0027, co-financed by the Croatian Government and the European Union through the European Regional Development FundâĂŤthe Competitiveness and Cohesion Operational Programme. The authors would like to thank ESA for providing Sentinel-2 through Copernicus Open Access Hub.

References

[1] ALEXANDRIS, N., KOUTSIAS, N., AND GUPTA, S. Remote sensing of burned areas via PCA, Part 2: SVD-based PCA using MODIS and Landsat data. *Open Geospatial Data, Software and Standards 2*, 1 (2017), 1–26. doi:10.1186/s40965-017-0029-0.

[2] AUSTRALIAN NATIONAL SEARCH AND RESCUE COUNCIL. ANSRC - The National Search and Rescue Manual - 2022 EDITION Version 1.

https://www.amsa.gov.au/national-search-and-rescue-council/manuals-and-publications/national-search-and-rescue-manual. Last Accessed May 2, 2022.

[3] AYMERICH VERDAGUER, E. *Applications of machine learning in decision analysis for rescuing lost people*. PhD thesis, ETSI_Informatica, 2021.

[4] BREIMAN, L. Random forests. *Machine learning 45*, 1 (2001), 5–32. doi:10.1023/A:1010933404324.

[5] CHICCO, D., WARRENS, M. J., AND JURMAN, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science 7* (2021), e623. doi:10.7717/peerj-cs.623.

[6] CIESA, M., GRIGOLATO, S., AND CAVALLI, R. Analysis on vehicle and walking speeds of search and rescue ground crews in mountainous areas. *Journal of Outdoor Recreation and Tourism 5* (2014), 48–57. doi:10.1016/j.jort.2014.03.004.

[7] CIOLLI, M., MENGON, L., VITTI, A., ZATELLI, P., AND ZOTTELE, F. A GIS-based decision support system for the management of SAR operations in mountain areas. *Dipartimento di Ingegneria Civile e Ambientalee, Università degli Studi di Trento Italia* (2006).

[8] CROATIAN AGENCY FOR THE ENVIRONMENT AND NATURE. National Habitat Classification of the Republic of Croatia (4th version). http://www.haop.hr/sites/default/files/uploads/dokumenti/2017-12/Nacionalna%20klasifikacija%20stanista_IVverzija.pdf, 2017. Last Accessed November 26, 2021.

[9] CROATIAN MOUNTAIN RESCUE SERVICE. CMRS - Croatian Mountain Rescue Service. http://www.hgss.hr/. Last Accessed December 15, 2020.

[10] DOHERTY, P. J., GUO, Q., DOKE, J., AND FERGUSON, D. An analysis of probability of area techniques for missing persons in Yosemite National Park. *Applied Geography 47* (2014), 99–110. doi:10.1016/j.apgeog.2013.11.001.

[11] D'ORSO, G., AND MIGLIORE, M. A GIS-based method for evaluating the walkability of a pedestrian environment and prioritised investments. *Journal of transport geography 82* (2020), 102555. doi:10.1016/j.jtrangeo.2019.102555.

[12] DU, Y., ZHANG, Y., LING, F., WANG, Q., LI, W., AND LI, X. Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band. *Remote Sensing 8*, 4 (2016), 354. doi:10.3390/rs8040354.

[13] EUROPEAN ENVIRONMENT AGENCY (EEA). European Digital Elevation Model (EU-DEM), version 1.1, https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1/view Jun 2020.

[14] EUROPEAN SPACE AGENCY. GMES Sentinel-2 Mission Requirements Document. https://esamultimedia.esa.int/docs/GMES/Sentinel-2_MRD.pdf, 2010. Last Accessed March 21, 2022.

[15] FERGUSON, D. GIS for wilderness search and rescue. In *ESRI federal user conference* (2008), vol. 2012, p. 10.

[16] FRITZ, S., CARVER, S., AND SEE, L. New GIS approaches to wild land mapping in Europe. In *Proceedings of the Wilderness science in a time of change conference. USDA Forest Service, Missoula, Montana* (2000), pp. 120–127.

[17] FUNK, S. Netflix Update: Try This at home . http://sifter.org/~simon/journal/20061211.html. Accessed: 2022-03-15.

[18] GABBOLINI, G., D'AMICO, E., BERNARDIS, C., AND CREMONESI, P. On the instability of embeddings for recommender systems: the case of matrix factorization. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing* (2021), pp. 1363–1370. doi:10.1145/3412841.3442011.

[19] GATTI, A AND BERTOLINI, A. Sentinel-2 products specification document. https://sentinel.esa.int/documents/247904/349490/S2_MSI_Product_Specification.pdf, 2013. Last Accessed November 25, 2021.

[20] GONZALEZ, R., JAYAKUMAR, P., AND IAGNEMMA, K. Generation of stochastic mobility maps for large-scale route planning of ground vehicles: A case study. *Journal of Terramechanics 69* (2017), 1–11. doi:10.1016/j.jterra.2016.10.001.

[21] HARRINGTON, K., BROWN, M., PINCHIN, J., AND SHARPLES, S. Decision-making within missing person search. *Cognition, Technology & Work 20*, 4 (2018), 665–680. doi:10.1007/s10111-018-0515-x.

[22] HETH, C. D., AND CORNELL, E. H. A geographic information system for managing search for lost persons. In *Applied Spatial Cognition*. Psychology Press, 2020, pp. 267–284.

[23] HUG, N. Surprise: A Python library for recommender systems. *Journal of Open Source Software 5*, 52 (2020), 2174. doi:10.21105/joss.02174.

[24] JIERULA, A., WANG, S., OH, T.-M., AND WANG, P. Study on Accuracy Metrics for Evaluating the Predictions of Damage Locations in Deep Piles Using Artificial Neural Networks with Acoustic Emission Data. *Applied Sciences 11*, 5 (2021), 2314. doi:10.3390/app11052314.

[25] KELLEY, D. E. *Mountain search for the lost victim*. Montrose, California, 1973.

[26] KOREN, T., IVONA, B., GLAVAN, G., AND VEROVNIK, R. Contribution to the knowledge of the butterfly fauna (Lepidoptera: Papilionoidea) of Mt Kozjak, Split, Croatia. *Natura Croatica: Periodicum Musei Historiae Naturalis Croatici 28*, 1 (2019), 21–34. doi:10.20302/NC.2019.28.2.

[27] KOREN, Y. The bellkor solution to the netflix grand prize. *Netflix prize documentation 81*, 2009 (2009), 1–10.

[28] KOREN, Y., BELL, R., AND VOLINSKY, C. Matrix factorization techniques for recommender systems. *Computer 42*, 8 (2009), 30–37. doi:10.1109/MC.2009.263.

[29] KWEON, I. S., AND KANADE, T. Extracting topographic terrain features from elevation maps. *CVGIP: image understanding 59*, 2 (1994), 171–182. doi:10.1006/ciun.1994.1011.

[30] LI, S., GANGULY, S., DUNGAN, J. L., WANG, W., AND NEMANI, R. R. Sentinel-2 MSI radiometric characterization and cross-calibration with Landsat-8 OLI. *Advances in Remote Sensing 6*, 02 (2017), 147. doi:10.4236/ars.2017.62011.

[31] LO, R. H. Walkability: what is it? *Journal of Urbanism 2*, 2 (2009), 145–166. doi:10.1080/17549170903092867.

[32] LWIN, K. K., AND MURAYAMA, Y. Modelling of urban green space walkability: Eco-friendly walk score calculator. *Computers, Environment and urban systems 35*, 5 (2011), 408–420. doi:10.1016/j.compenvurbsys.2011.05.002.

[33] OGIE, R. I., RHO, J. C., AND CLARKE, R. J. Artificial intelligence in disaster risk communication: A systematic literature review. In *2018 5th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)* (2018), IEEE, pp. 1–8. doi:10.1109/ICT-DM.2018.8636380.

[34] OPENSTREETMAP CONTRIBUTORS. Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org, 2017.

[35] OSTERTAGOVÁ, E. Modelling using polynomial regression. *Procedia Engineering 48* (2012), 500–506. doi:10.1016/j.proeng.2012.09.545.

[36] PHILLIPS, K., LONGDEN, M. J., VANDERGRAFF, B., SMITH, W. R., WEBER, D. C., MCINTOSH, S. E., AND WHEELER III, A. R. Wilderness search strategy and tactics. *Wilderness & environmental medicine 25*, 2 (2014), 166–176. doi:10.1016/j.wem.2014.02.006.

[37] PHILLIPS, R. D., WATSON, L. T., WYNNE, R. H., AND BLINN, C. E. Feature reduction using a singular value decomposition for the iterative guided spectral class rejection hybrid classifier. *ISPRS Journal of Photogrammetry and Remote Sensing 64*, 1 (2009), 107–116. doi:10.1016/j.isprsjprs.2008.03.004.

[38] PHIRI, D., SIMWANDA, M., SALEKIN, S., NYIRENDA, V. R., MURAYAMA, Y., AND RANAGALAGE, M. Sentinel-2 Data for Land Cover/Use Mapping: A Review. *Remote Sensing 12*, 14 (2020). doi:10.3390/rs12142291.

[39] SATOH, E., YOSHIKAWA, T., AND YAMADA, A. Investigation of converted walking distance considering resistance of topographical features and changes in physical strength by age model for location planning of regional facilities considering topographical condition and aging society part 1. *Journal of Architecture and Planning(Transactions of AIJ)*, 610 (2006), 133–139. doi:10.3130/aija.71.133_2.

[40] SAVA, E., TWARDY, C., KOESTER, R., AND SONWALKAR, M. Evaluating lost person behavior models. *Transactions in GIS 20*, 1 (2016), 38–53. doi:10.1111/tgis.12143.

[41] ŠERIĆ, L., PINJUŠIĆ, T., TOPIĆ, K., AND BLAŽEVIĆ, T. Lost Person Search Area Prediction Based on Regression and Transfer Learning Models. *ISPRS International Journal of Geo-Information 10*, 2 (2021), 80. doi:10.3390/ijgi10020080.

[42] TELEGA, A., TELEGA, I., AND BIEDA, A. Measuring walkability with GIS—Methods overview and new approach proposal. *Sustainability 13*, 4 (2021), 1883. doi:10.3390/su13041883.

[43] THAPA, R. B., AND MURAYAMA, Y. Evaluating walkability in Tsukuba using remote sensing and GIS. In *Papers and Proceedings of the Geographic Information Systems Association* (2009), vol. 18, pp. 549–552.

[44] TOBLER, W. *Three Presentations on Geographical Analysis and Modeling: Non- Isotropic Geographic Modeling; Speculations on the Geometry of Geography; and Global Spatial Analysis (93-1)*. UC Santa Barbara: National Center for Geographic Information and Analysis, 1993.

[45] WYSOKIŃSKI, M., MARCJAN, R., AND DAJDA, J. Decision support software for search & rescue operations. *Procedia Computer Science 35* (2014), 776–785. doi:10.1016/j.procs.2014.08.160.

[46] YOO, H. J., AND LEE, J. Developing a Method to Define Mountain Search Priority Areas Based on Behavioral Characteristics of Missing Persons. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography 37*, 5 (2019), 293–302. doi:10.7848/ksgpc.2019.37.5.293.