Article

Lost Person Search Area Prediction Based on Regression and Transfer Learning Models

Ljiljana Šerić [1,*], Tomas Pinjušić [1], Karlo Topić [1] and Tomislav Blažević [2]

[1] Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split,
R. Boškovića 32, 21000 Split, Croatia; Tomas.Pinjusic.00@fesb.hr (T.P.); Karlo.Topic.00@fesb.hr (K.T.)
[2] Croatian Mountain Rescue Service, Split Station, Šibenska 41, 21000 Split, Croatia; Tomislav.Blazevic@hgss.hr
[*] Correspondence: ljiljana.seric@fesb.hr

Abstract: In this paper, we propose a methodology and algorithms for search and rescue mission planning. These algorithms construct optimal areas for lost person search having in mind the initial point of planning and features of the surrounding area. The algorithms are trained on previous search and rescue missions data collected from three stations of the Croatian Mountain Rescue Service. The training was performed in two training phases and having two data sets. The first phase was the construction of a regression model of the speed of walking. This model predicts the speed of walking of a rescuer who is considered a well-trained and motivated person since the model is fitted on a dataset made of GPS tracking data collected from Mountain Rescue Service rescuers. The second phase is the calibration of the model for lost person speed of walking prediction with transfer learning on lost person data. The model is used in the simulation of walking in all directions to predict the maximum area where a person can be located. The performance of the algorithms was analysed with respect to a small dataset of archive data of real search and rescue missions that was available and results are discussed.

Keywords: search and rescue; machine learning; regression; transfer learning; cellular automata simulation

1. Introduction

Lost person search and rescue (SAR) activities are civil protection activities carried out by search and rescue teams. A subject may be lost under various circumstances—such as tourists wandering off the hiking trail, children wandering off into wildlands, or older people with dementia wandering from home. Only in Croatia, Croatian Mountain Rescue Service carried out over 6000 missions since its establishment 60 years ago [1]. When a person is lost, an incident of a lost person is reported, a *SAR team* assembles in a time-critical manner and carries out activities aimed towards finding the lost person as soon as possible. SAR team members have a task to search a certain area and the operation is directed and coordinated by the *SAR manager*. Each such incident and activity is different and has its own challenges, but the experience and intuition of the SAR team and manager can be of vital importance in critical situations. Experienced SAR managers will make decisions that will lead to the most effective completion of the task.

Specific SAR operation is characterised by *Initial Planning Point (IPP)*, which is the initial point around which a search for a lost person is planned. Usually, it is the same as the geographic location where the lost person is last seen. This location is often referred to as *Point Last Seen (PLS)*. This location is known from the report of the lost person since the person who reports a person is lost also gives the location where the subject is seen last time. The direction of the search, focus area and distance from IPP to be searched depends on different factors and are subject to an assessment of the SAR Manager. Experienced SAR manager will make decisions that will lead the team towards finding the lost person quickly. However, focusing only on the most probable locations, one can miss the true

location of the lost person, since not all behavior is the most common one. The research presented in this paper is aimed towards constructing a method and a software tool that will make these decisions easier.

Information and Communication Technologies (ICT) have infiltrated in all aspects of human activities [2]. Today, ICT tools are used not only in the improvement of productivity, communication, lifestyle, and traveling, but can also be found in the domain of government management and public safety [3]. In SAR activities, ICT tools, in its rudimentary form, are used as communication tools between team members. In its more sophisticated form, ICT tools can also serve as a *Decision Support System (DSS)*. DSS can be beneficial for optimal planning of SAR activities providing suggestions which are based on data, models and artificial intelligence.

This paper presents a methodology used and results achieved while developing an ICT-based tool whose intention is to be used by a search and rescue mission planning team for assessment of an area to be scanned by the team while searching the lost person. Our method makes suggestions of the direction of search, focus area and distance from IPP to be searched. In other words, the method suggests an irregularly shaped area to the SAR manager where a lost person should be searched for.

We describe the steps of the methodology for constructing algorithms—data pre-processing, development of regression models, transfer learning model calibration, simulation algorithm and construction of the proposed shape of the area that should be searched. We compared the results of the algorithms—shape of the proposed area with locations where the person was found from archived records and presented results.

1.1. Related Work

As already stated in the introduction, ICT tools can ease several tasks of organized SAR activities. In this work we will focus on the role of ICT systems in the task of determining the search area. *Search area* is the area searchers are screening to find out the new location of the lost person. There are various approaches in determining the search area.

First, we must distinguish that SAR can occur on the land and on the sea. When a person is lost at sea, the search area is determined with respect to sea currents, but the spatial features of the nearby coast can also be a valuable input for determining a more precise search area as proposed in [4].

In cases when SAR occurs on land, the prediction of the new location of the lost person depends on many factors that can be roughly separated to (a) features of the lost person and (b) spatial features of the surrounding area. In our work, we deal with the search for a lost person in non-urban environments, and that is often referred to as wilderness search and rescue [5]. Methodological search area prediction is based on a model, while subject of modelling can be the lost person or the area.

A model that describes lost person behaviour usually relies on an archive record of previous cases and statistics. The first documented attempt to analyze lost person behavior is when Father Lorenzo at the St. Gotthard Hospice, a monastery in Switzerland started recording missions of search and rescue in the Swiss Alps [6] in 1783. Since then, there are several records of lost person search and rescue archive databases. Statistics compiled in the book [7] was used as the first ground for search management. More recent archive database of previous SAR activities—International Search and Rescue Incident Database (ISRID) is the basis for lost person behavior analysis in [8].

Lost person behavior has been most thoroughly studied in [8]. In [8] the author proposes a model of lost person behaviour based on the statistics obtained from the ISRID database. The model uses Euclidian distance tables and proposes a search area using the point radius method around IPP.

However, for a case study of Yosemite National Park, the proposed model has shown poor results so new statistics for only this area is proposed in [9]. Evaluation of lost person behavior models was done in [10]. The authors compared Euclidean distance tables from [8] and watershed model from [9] and proposed a novel model based on combining the two

previous models. As we observe the differences between the model based on international statistics and the model based on local statistics, we can assume that in addition to the lost person behavior, the local characteristics of the terrain should be taken into account when estimating the search area.

Analysis of the terrain is most effectively performed by using Geographical information systems (GIS). GIS systems effectively manipulate multiple information about terrain characteristics such as digital elevation model, land cover, roads, sightseeings, etc. In [11] authors present a GIS-based search and rescue decision support software. The software uses a model based on previous operations data and calculates the probability of a subject to be found in different segments of the search area. The output of the software is a heat map constructed by combining influencing features of the terrain. In [12] authors integrated aspects of the terrain and lost person and used *Bayesian approach* for predicting lost person behavior.

In contrast to similar systems, where the proposed search area is an area with the largest probability of finding the lost person having in mind the statistics of archived searches, we propose a new simulation-based approach. Our method is based on simulations of all possible behaviours and trajectories of walking and proposes the search area with all locations where the person can dwell after wondering from IPP.

The second novel aspect of our research is the usage of data science methods for modelling the speed of walking on non-urban terrain. Data science methods were already used for modelling the speed of walk-in urban areas. *Linear regression*, as a common machine learning technique was used in [13] for predicting the speed of walking. In [14] the authors exploited a *latent terrain model* to predict a traversal path of a subject moving. In [15] authors used transfer learning to predict urban crowd movement patterns. However, lost person movement in the wilderness is different and needs different approaches than urban movement modelling.

Transfer learning [16] has been successfully used in deep learning. With this approach, a neural network model is pre-trained with a large set of data and the learned features are used for the specification of the model for a particular domain where it is not possible to obtain a data set large enough for training the actual classifier or model. The same approach can be used for transfer learning of a linear regression model. In [17] a method for refining a linear regression model that is initially trained for one domain to be used on another domain is described.

Cellular automata (CA) [18] are simple mathematical models often used to investigate the summary effect of the collection of simple components. Their usefulness has been proved in many domains. Cellular automata have been used for traffic simulation [19] and simulation of pedestrians walking [20]. GIS-based cellular automata have been used for land-use change simulation [21] and fire spread simulation [22] . In the civil protection domain, cellular automata has been used to simulate evacuation routes in [23].

In [24] the authors used agent-based modelling to calculate the distribution of behaviors and compute the distributions of horizontal distances traveled in a fixed time.

1.2. Proposed Method

The novelty of our work can be noticed in two aspects. The first is a new, simulation-based approach for determining the search area. In similar systems search area is proposed as the most probable area where the lost person will be found based on the statistics of archived searches. Our prediction is based on simulation. We do not assume the lost person's behaviour will be statistically predictable, but rather simulate all possible behaviours and all possible trajectories of walking in order to achieve maximal area off all places where the person can dwell after wandering from IPP. The only parameter of the lost person's behaviour we assume is the predicted speed of walking on the terrain with various features.

The second novelty of our proposed method is the transfer learning-based regression model of the speed of walking. We do not have records of lost people walking so we cannot

create a machine learning model for predicting the speed of walking of the lost person. Thus, we use the available trajectories to create a model for predicting the speed of walking on a segment of terrain and use transfer it by scaling on a model for predicting the speed of walking of the lost person. The transfer learning-based approach enables us to create a lost person speed of walking model without a sufficient amount of data for an accurate machine learning model. We use the recorded data of walking on the same terrain we could obtain, and that is the GPX records of people searching for a lost person. We use a small set of records of lost person pairs of two locations—the initial point of search and point where the person is found and transfer model for the lost person.

2. Methodology

In this section, we will describe the data we used and the methodology of our work. Firstly, data collected from various sources that are expressed in several formats were pre-processed. Preprocessing was performed for data association and integration as depicted in Figure 1. The result of the preprocessing is a connected dataset that we used for training the model. The training was performed in two phases - pretraining model and calibration of the model. Finally, we describe algorithms we used for predicting search area.

2.1. Description and Sources of Data

The basis of our dataset was a set of files provided by the Croatian Mountain Rescue Service [1]. The set consisting out of 1908 GPX trails was made available for our research. The GPX trails are collected from three Mountain rescue service departments—Split, Karlovac and Dubrovnik. The trails were collected using different GPS devices held by different persons. The trails are recorded on wide areas of the three cities as shown in Figure 2. The trails were recorded and collected during real search missions on the past on incidents that occurred between 1999 and 2020. All data was anonymized. GPX [25] is GPS Exchange Format—XML format for exchanging GPS data. A GPX trail consists of a series of points each with associated geo-coordinates (longitude and latitude), elevation and time. This data set was enriched with spatial data collected from other sources describing the terrain of the segment, particularly vegetation—Corine land cover, CLC [26] and terrain—digital elevation model, DEM [27] and processed into the data set. CLC and DEM data were obtained in geotiff format [28], a format for storing georeferenced raster data.

2.2. Data Preprocessing

The complete diagram of pre-processing data is depicted in Figure 1. GPX set contains a set of GPX files. Each file consists of a trail of one person walking. A GPX trail is a record of a series of geographical points where a person wearing a GPS device was walking. Trails were processed and transformed into segments. A segment describes walking between two points on the earth's surface. Each segment is described with start and end points as well as start and end time. From this information, we can easily calculate the necessary features of a segment: distance length, the slope of the terrain and the speed of walking segment.

The distance length of a segment was calculated using haversine formula [29] for calculating the spherical distance between two points on the earth's surface. Even though the average length of a segment is only 6.7 m and the spherical distance is not necessary, we exploited the formula that is used as a common practice for distance calculation.

We assume that the person walked the distance between two points in a straight line having in mind the GPS devices used were precise enough and recorded points that are close enough. Terrain slope was calculated expressed as the absolute value of the tangent of the angle obtained as the elevation difference and length ratio. The absolute value is taken in accordance to [30]. In this work, the author proposed the model of hiking speed in hilly terrain. The resulting model showed that the speed depends on the absolute value of the slope rather than being significantly different for walking uphill and downhill. A similar relationship was discovered in our data set, so that decision is made that the variable of the slope of the terrain is expressed in the absolute value of the angle tangent.

Speed of walking is calculated as an average horizontal speed of walking towards the end point of the segment from the start point of the segment. The majority of the segment's length was under 10 m, the average being 6.7 m. In such a segment we assume that the walking was a straight line. The elevation difference is not taken into account while calculating the speed of walking. This simplification is also beneficial for final implementation because the model will be used on a two dimensional cell grid and we will need distance walked predicted from the model as distance on the map not on the sphere.

Finally, the dataset is enriched with data about land cover from additional source. In our implementation we used the Corine Land Cover map obtained from Copernicus site [26].

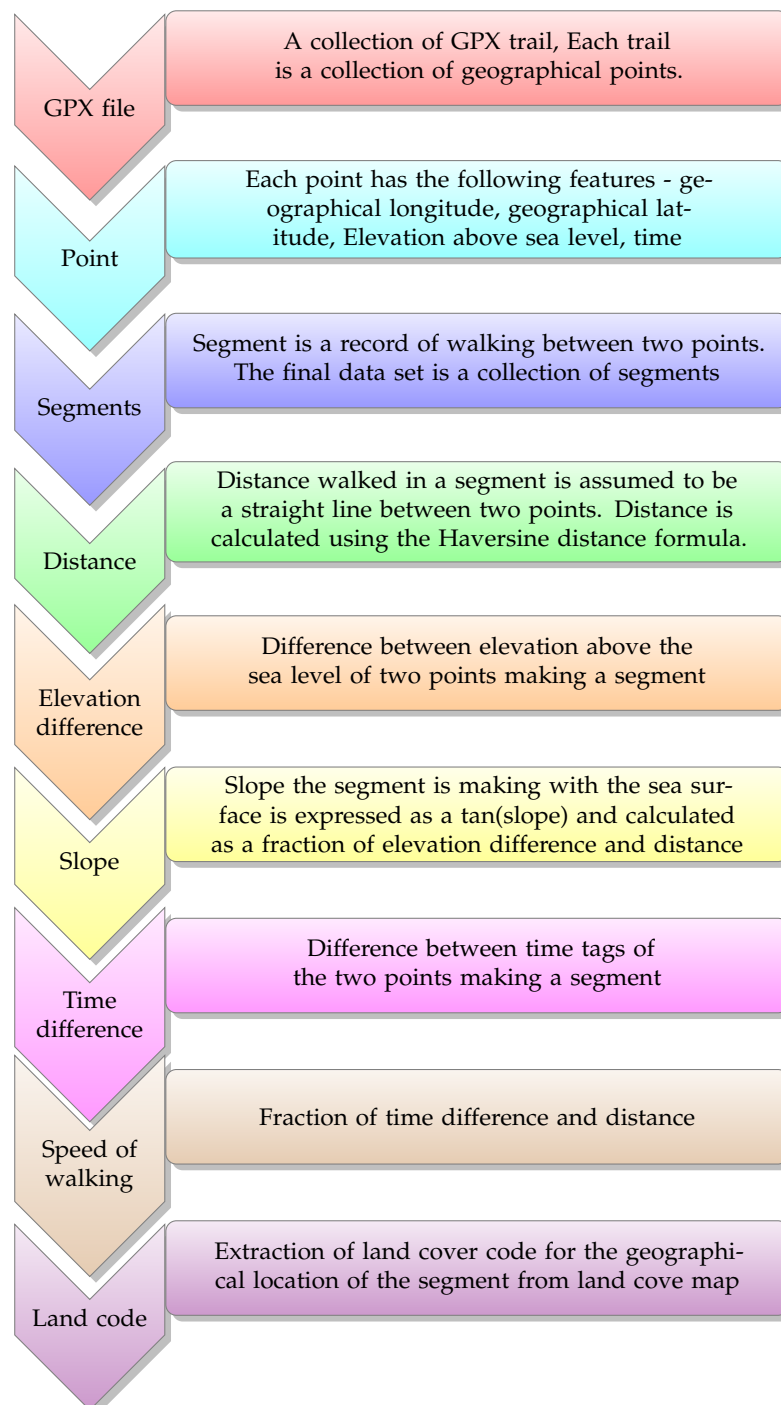Steps of data preprocessing used in this work are shown in Figure 1.

Figure 1. Steps of data preprocessing.

After preprocessing the whole dataset, we filtered the data that can bring confusion into the model, but can be rejected heuristically—such as calculated speed of walking was higher than 10 km/h, where time difference was larger than 20 s and similar. The dataset that was left consists of 1,432,740 segments of walking of various users on various terrains.

To better understand the volume and distribution of data in the dataset we visualized the data in a way that we presented the number of users walking the same segment as color and the result is shown in Figure 2. The darker the color of the line the more GPS trails are recorded on that segment.
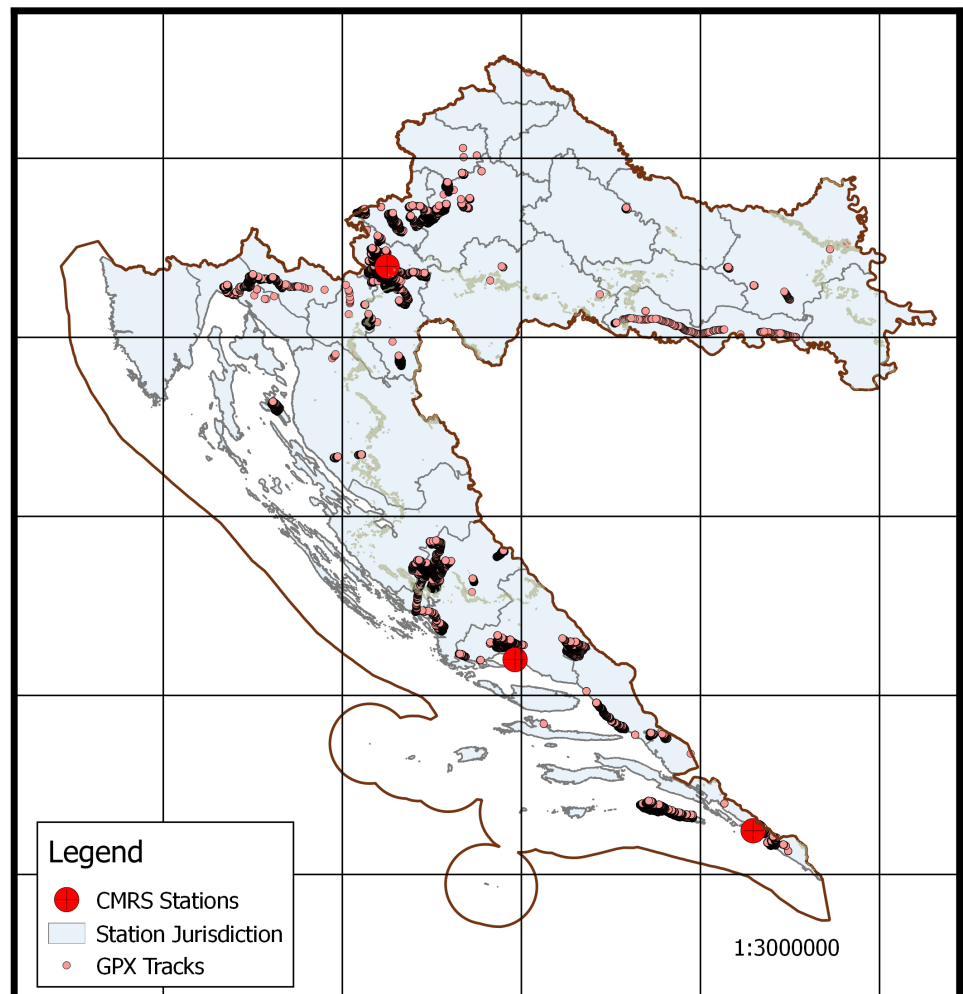

Figure 2. Spatial distribution of the GPX files data. The GPX trails are collected from three Mountain rescue service departments—Split, Karlovac and Dubrovnik. The trails are collected in a wide area of the three cities.

Additionally, to enrich the dataset, we described each segment with information about the land cover. Initially, we used the standard land cover classification, Corine Land Cover [26]. Corine land cover code is a three digit code describing the class and subclasses of the terrain. When observing the average speed of walking the terrain, those codes did not suffice the purpose of linearity, so additional pre-processing was made. After observation of the speed of walking on particular terrains, we constructed a translation table from the original CLC codes into our Land cover identification codes (LC id). CLC codes were translated into an LC id using the translation table shown in Table 1.

Table 1. Translation table for Corine land cover translation to Land Cover id.

LC id	CLC Code	Label
22	244	Agro-forestry areas
23	311	Broad-leaved forest
24	312	Coniferous forest
25	313	Mixed forest
26	321	Natural grasslands
27	322	Moors and heathland
28	323	Sclerophyllous vegetation
29	324	Transitional woodland-shrub

Finally, we constructed a data set where each row represents a segment walked by a particular person. A segment in a row is described with the following features:

- id—unique identifier of the data sample
- LC id—land cover type identifier as described in Table 1
- DEM—value read from the digital elevation model file associated with the start point denoting elevation above sea level in meters,
- abs slope—absolute value of the slope tangent, calculated as a fraction of vertical elevation difference (in meters) and horizontal distance (in meters).
- dist wgs—distance length of the segment between two geographical points in World Geodetic System in meters,
- d from start—distance, i.e., position of the segment in the collection from the start of the GPX trail
- speed 2d kmh—average speed of walking on the segment by the particular user expressed in km/h

A sample of data from the dataset is shown in Table 2.

Table 2. A sample of data used for regression model training.

id	LC id	DEM	abs Slope	Elev d	Dist Wgs	d from Start	Sped 2d kmh
0	29	68.39	0.000137	0.96	7.0102	0	2.80408
3	29	68.39	0.000203	0.96	4.721129	3	1.545097
7	29	67.91	0.000424	0.96	2.26302	7	0.581919
11	29	68.87	−0.000168	−0.96	5.731096	11	1.375463

2.3. Linear Regression Model

Linear regression is a machine learning technique used to predict the value of a continuous dependent variable, often referred to as output variable, modelled as linear combinations of independent variable values, referred to as explanatory variables or input variables [31]. A simple regression equation has on the right hand side an intercept and an explanatory variable with a slope coefficient. A multiple regression has multiple explanatory variables on the right hand side, each with its own slope coefficient. Equation for prediction is shown in Equation (1):

$$y = q_0 + q_1 x_1 + q_2 x_2 + ... + q_n x_n \tag{1}$$

where y is the value being predicted, q_0 is bias or intercept , $q_1, q_2, ..q_n$, are slope coefficients for explanatory variables $x_1, x_2, ..x_n$ respectively.

Training a linear regression model comes down to adjusting the values of slope coefficients so that the model will fit the best into training data. In this work, we trained a linear regression model to predict the time taken for walking a segment of the terrain as a linear combination of values:

- land cover id value, determined as described in the previous section,

- terrain slope,
- distance length a segment,
- difference in elevation of end and start point,
- elevation above the sea level.

The output variable is the time of walking the terrain segment. We rather predict time than speed since the time of walking will be used later in the simulation. However, these two variables are correlated if we have a fixed distance length. After performing gradient descent for k-folds fitting we received a model with a score on the train set 0.4127 and on the test set 0.4120.

Although the score value is not optimal, we assume that this is due to the simplification we made by neglecting the variable that describes a person walking the segment and his or her characteristics. Due to the anonymization of data the aspect of information regarding a person could not be included in the model using the method described in this paper. However, we assume that the produced model incorporated averaging of the speed of walking, which can be used as a basis for calibration with transfer learning.

2.4. Model Calibration with Transfer Learning

Detailed information about lost people tracking is not available. We cannot assume the dynamics of the direction change by a subject wondering between the IPP and the location of finding. The only available information we could obtain for research purposes was the initial point—IPP or PLS and the location where the lost person is found. Both of these points are described with geographical coordinates. Quite a small number of point pairs was collected—only 20 samples of data. All collected locations are situated in forests or rural areas. By analyzing this data set we came to the conclusion that 50% of lost persons are found within 1 km distance from the initial point, while 75% of lost persons are found within 2 km distance from the initial point.

We used the model from the previous section to predict the distance to which a person would walk in all directions. The time of simulation was intuitively selected to be four hours to support the initial search. The distance walked is calculated using the same cellular automata simulation procedure as described in the next section. The simulation result is visualized as isochrones—lines connecting the distances that a person modeled would be able to walk in any direction from the initial point at the same time. The result of one of the simulation is shown in Figure 3, where the initial point of the search is labeled with a star, location where the person is found is labeled with a cross, and isochrones connecting every 30 min steps are shown in red lines. After inspecting the results of the simulation for all data from the archive dataset, comparing the resulting isochrones with the location where the lost person is found, and neglecting the outliers we adjusted the parameters of the resulting model so that the location where the person is found is inside the simulation isochrones for 75% of the data. The described process of building the final model is depicted in Figure 4.
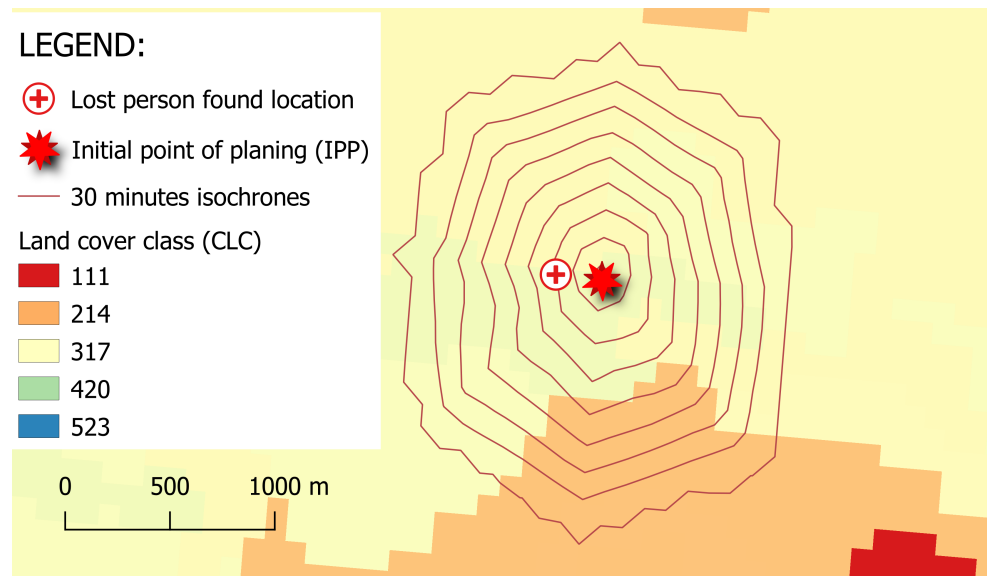
Figure 3. An example of simulation and comparison of lost person end point and distance walked predicted with pre trained model.
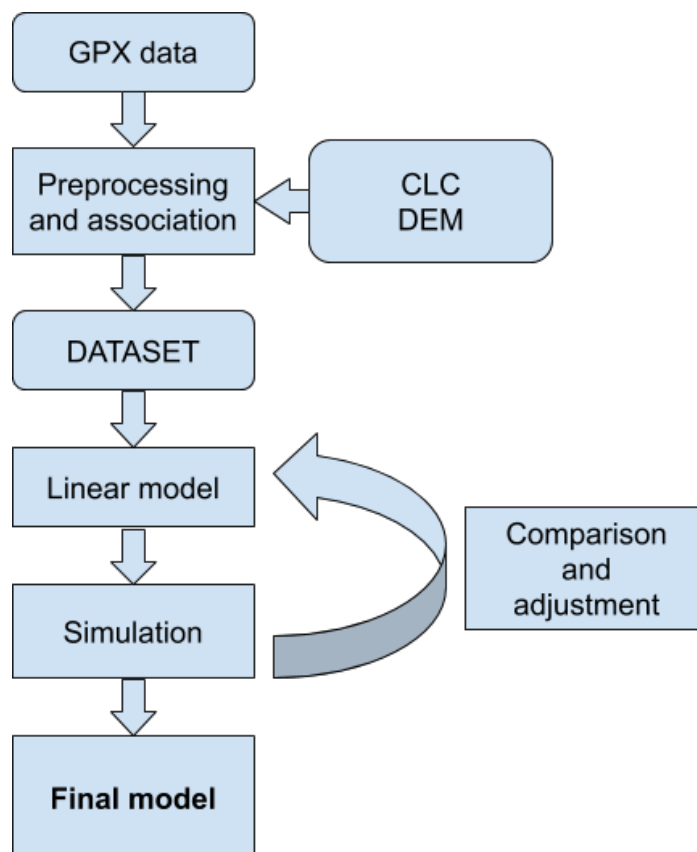

Figure 4. The process of building the model.

2.5. Search Area Prediction

As already mentioned, predicting the search area is performed by running a simulation of walking the terrain surrounding the initial point of search. Before running the simulation we prepare a cell grid of the terrain features surrounding the initial point. Each cell covers 5 m × 5 m area, since this is the most precise resolution of the data we use. We create a grid with 400 cells in each direction (up, down, left and right) of the initial point. The reason for choosing exactly 400 cells is to cover the maximum radius of 2 km distance. Each

cell is described with elevation above the ground (dem) and Land Cover id (LC id) read from GIS data. The simulation is performed in time ticks. Initially, a person is located in the initial point—IPP located in the center of the grid. The cell in the center is assigned a visited state and is labeled *active* for calculation. For all 8 surrounding cells as shown in Figure 5 we calculate the time taken to reach the cell. The time needed to reach the cell is time for walking 5m for up, down, left and right cells, and 7.25 m for diagonal cells. When assigning the distance, we do not calculate the distance that takes into account the difference in elevation, since the elevation difference is taken into account in the other features—slope and elevation difference. The time of walking the segment is predicted using the model in the form of Equation (1)

Surrounding cells become visited and are labeled *active* for activation in the next step. *Active* cells are assigned a value of the time taken to reach them from the IPP. The simulation is done in iterations, calculating the time taken to reach each surrounded cell of all *active* cells. If a cell can be reached from more than one active cell, the cell is assigned the value of the lowest time to reach the IPP. The simulation runs until all cells are visited. The dynamic of this process is depicted in Figure 6 where figure (a) depicts the state of the cells after 30 ticks and (b) after 300 ticks.
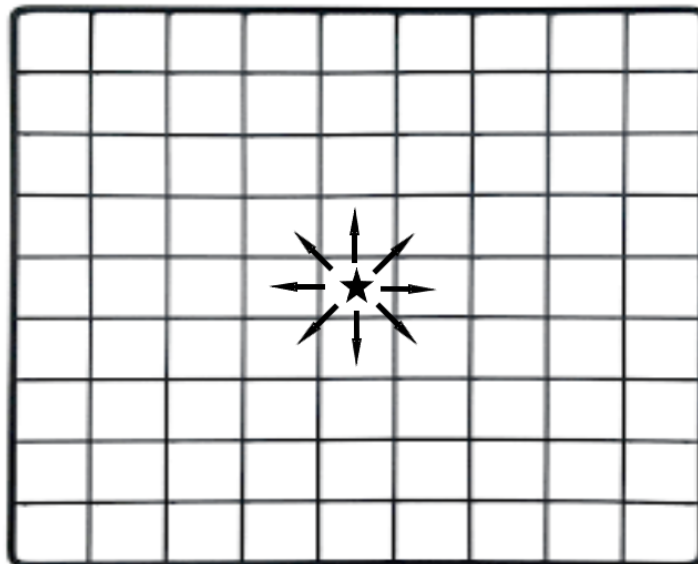

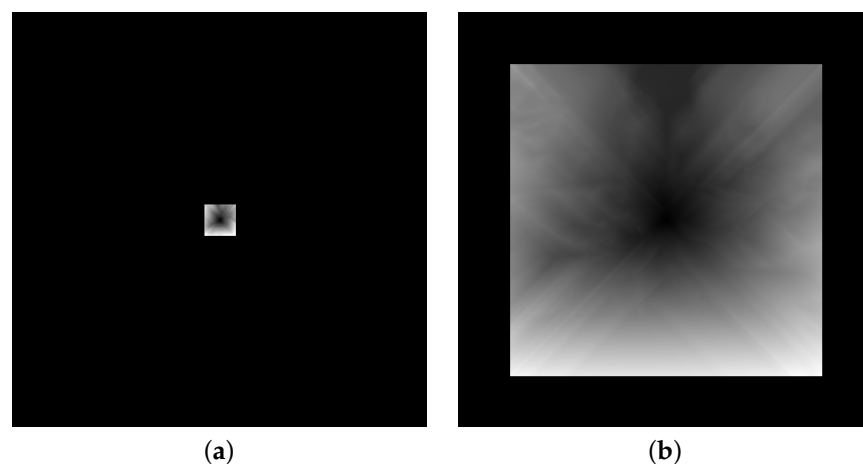**Figure 5.** Search for surrounding cells in the initial tick of simulation.

(**a**)   (**b**)

Figure 6. Visualization of time taken to reach the cells in (**a**) 30 ticks and (**b**) 300 ticks.

After the simulation is done, each cell of the grid has a value assigned—a positive number denoting the time taken to reach the cell from the initial cell. We use gdal-contour [32]

utility for transforming the obtained results to a shapefile of isochrones reached every 30 min. The produced shapefile is used for visualization of results in any standard GIS software, such as QGIS [33].

3. Results and Discussion

Pre-training of a linear regression model was done with a training set that is obtained by processing a whole set of GPX tracks into segment items. Furthermore, the dataset was associated with DEM and Corine land cover data. The final set is split into training and testing sets in the ratio 0.67/0.33. Several instances of regression models were tested. Polynomial regression with 2 degree polynomial produced slightly better scores on the training and test set. We also performed an experiment with other regression models on the same train and test dataset. The resulting scores on the test set are compared in Table 3.

Table 3. Comparison of test score of various regression models on the same dataset.

Model	Test Score
Linear Regression	0.412
Polynomial Regression	0.454
Decision Tree Regressor	0.647
Bayesian Ridge Regression	0.412
Elastic Net Regression	0.410
Ridge Regression	0.401

Decision Tree Regressor model for speed of walking scored the best results on a test set. However, the gain in the score results was not significant enough to justify using a more complex model, so we decided to proceed with the more simple linear regression model. The motivation for using linear regression is because this simple model is easily transferred between the domains of different subjects — searchers and the lost person. We obtained a linear regression model for predicting time for walking a segment as shown in Equation (2):

$$time_{pt} = q_0 + q_1 * LC_{id} + q_2 * dem + q_3 * absslope + q_4 * elevd + q_5 * dist \qquad (2)$$

where:

q_0	4.786872732767515
q_1	0.013315859975442301
q_2	0.0019411657191748125
q_3	-16.319148163916193
q_4	-0.026739066247719285
q_5	0.5717657455052271
and	
LC_{id}	land cover id value
dem	elevation above sea level in m
absslope	absolute value of segment slope tangent
elev d	difference in elevation between and and start point
dist	distance walked in m

In this equation, the pre-trained prediction of time for walking segment ($time_{pt}$) is calculated as a linear combination of explanatory values. We provide slope coefficient precise values for the reproducibility of the model. After calibration of the equation for predicting the lost person speed of walking as explained in Section 2.4, we obtained following factors for lost person speed of walking on a segment of terrain:

q_0	1.1967181831918787
q_1	0.0033289649938605752
q_2	0.0004852914297937031
q_3	-4.079787040979048
q_4	-0.006684766561929821
q_5	0.14294143637630677

Equation (2) together with corrected factors shown in the table above is exploited in a simulation. We developed a script written in Python programming language [34] for running the simulation, while several ticks and the extension of the area that is observed can be adjusted as a parameter. We used Rasterio library [35] for transforming the resulting grid cell into a georeferenced tiff file and *gdal-contour* [32] for vectorizing the results into a shapefile. The resulting shapefile gives us the area where a lost person can obtain depending on the time elapsed since the person is seen in the initial point. An example of simulation and search area prediction is shown in Figure 7.
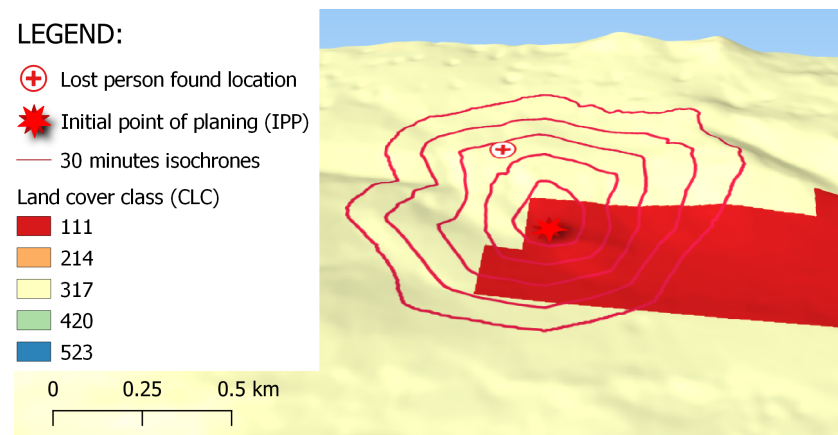

Figure 7. Result of simulation with isochrones showing maximal reached area in every half an hour. IPP is taken from archive of SAR activities, and location where subject is found is denoted with red cross.

The resulting area is irregularly shaped, and the shape depends on the configuration of the surrounding terrain. This means that in directions where the terrain is configured so that one must walk slower, the area to be searched is smaller. Search area defined in such a way is more accurate than the traditional approach—determination of the radius of a circle around the initial point of the search where every direction is probably the same, while still we do not reject area where the probability of finding the lost person is low.

To evaluate the results, we ran the simulation for 20 locations from the archive data of the lost person SAR and compared the location where the person is found and the isochrones resulting from the simulation. Out of 20 cases, four people were found outside the area predicted with our method. For nine simulations the lost person is found within the first isochrone, in six cases within 2. isochrone, and for 1 case the person is found within 3. isochrone line. This is summed up in Table 4.

Table 4. Evaluation of the algorithm on real lost person SAR database.

Number of Isochrones within the Person Is Found	Number of Cases
1	9
2	6
3	1
outside the predicted area	4

This method is based on many approximations and neglects several aspects such as lost person physical and psychological features, circumstances under which the person is lost, auxiliary conditions such as weather and visibility. However, the model and prediction technique can help SAR managers to better understand the influence of features, such as land cover and terrain slope, in lost person movement and help them decide about the search area shape and extent while still relying on experience and intuition. The predicted area can be further analyzed in GIS software.

In the scope of this paper, we do not address the problem of calculation time and complexity since we focus primarily on predicting the shape and size of the search area that can be done offline.

4. Conclusions and Future Work

This paper proposes and demonstrates a method for building a software system which can be used as a part of a decision support system in search and rescue operative actions management. We present the software we created using the proposed method. The goal of the system is to propose a search area—an area of land within what the lost person most probably will be found. We speculate that the best proposed search area does not have a regular shape, but its irregularity depends on surrounding terrain configuration as well as land cover. We propose a linear regression model trained on GPS tracking data collected from previous search and rescue missions by tracking the movement of rescuers. The speed of walking the segment of land is modeled as a linear regression with a score on the train set 0.4127 and a score on the test set 0.4120. The linear regression model was calibrated to predict the speed of walking of the lost person by introducing the scaling factor of the original coefficients.

An algorithm for predicting the search area is based on cellular automata simulation of walking from the initial point and determining the maximum distance a lost person could walk in a predefined period of time. We calculate the times for reaching the locations within a 2 km distance from the initial point and create isochrone lines for every 30 min. Each isochrone shows the maximal area where the subject may be found if he/she has walked an additional 30 min. The resulting model predicts the search area based on terrain configuration described with land cover class, elevation above the sea level, and slope of the terrain in the direction of walking. The resulting area can be included in GIS-based decision support software and further analyzed with respect to roads, watersheds, and place-marks for which the lost person may have an interest.

Other features could be taken into account in the future refining of the model. Remote sensing data and data from other sources (such as weather data) will be used to enrich the training dataset in an attempt to achieve better precision of the predictive model.

A model that would take into account the features of the lost person could significantly improve the accuracy of the model. One of the directions of future work is to extract the features of the person's behavior from obtained GPS tracks by using latent space transformation. More sophisticated modelling techniques for predicting the movement of lost persons will be examined in future work.

Additionally, more attention will be given to optimizing the performance of the simulation in order to achieve usability in real-time with faster results.

Author Contributions: Data curation, T.B.; Methodology, L.Š.; Resources, T.B.; Software, L.Š., T.P. and K.T.; Supervision, L.Š.; Writing—original draft, L.Š., T.B. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

Abbreviations

The following abbreviations are used in this manuscript:

SAR	Search and Rescue
IPP	initial planning point
PLS	Point Last Seen
ICT	Information and Communication Technologies
DSS	Decision Support System
CA	Cellular Automata
ISRID	International Search and Rescue Incident Database
GIS	Geographical Information System
GPS	Global Positioning System
GPX	GPS Exchange Format
XML	eXchangable Markup Language
CLC	Corine Land Cover
DEM	Digital Elevation Model

References

1. Croatian Mountain Rescue Service. CMRS—Croatian Mountain Rescue Service. Available online: http://www.hgss.hr/ (accessed on 15 December 2020).
2. Lu, Y. Industry 4.0: A survey on technologies, applications and open research issues. *J. Ind. Inf. Integr.* **2017**, *6*, 1–10. [CrossRef]
3. Jansen, A. The Understanding of ICTs in Public Sector and Its Impact on Governance. In *Electronic Government*; Scholl, H.J., Janssen, M., Wimmer, M.A., Moe, C.E., Flak, L.S., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 174–186.
4. Zhou, X.; Cheng, L.; Zhang, F.; Yan, Z.; Ruan, X.; Min, K.; Li, M. Integrating island spatial information and integer optimization for locating maritime search and rescue bases: A case study in the south China sea. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 88. [CrossRef]
5. Phillips, K.; Longden, M.J.; Vandergraff, B.; Smith, W.R.; Weber, D.C.; McIntosh, S.E.; Wheeler, A.R., III. Wilderness search strategy and tactics. *Wilderness Environ. Med.* **2014**, *25*, 166–176. [CrossRef] [PubMed]
6. Profound Journey. 13 Fascinating Facts about Lost Person Behaviour. Available online: https://profoundjourney.com/13-facts-lost-person-behaviour/ (accessed on 15 December 2020).
7. Kelley, D.E. *Mountain Search for the Lost Victim*; David E Kelley: Montrose, CO, USA, 1973.
8. Koester, R.J. *Lost Person Behavior: A Search and Rescue*; dbS Productions LLC: Charlottesville, VA, USA, 2008.
9. Doke, J. Analysis of Search Incidents and Lost Person Behavior in Yosemite National Park. Ph.D. Thesis, University of Kansas, Lawrence, KS, USA, 2012.
10. Sava, E.; Twardy, C.; Koester, R.; Sonwalkar, M. Evaluating Lost Person Behavior Models. *Trans. GIS* **2016**, *20*, 38–53. [CrossRef]
11. Wysokinski, M.; Marcjan, R.; Dajda, J. Decision support software for search & rescue operations. In Proceedings of the 18th Annual International Conference on Knowledge-Based and Intelligent Information and Engineering Systems (KES), Pomeranian Sci & Technol, Gdynia, Poland, 15–17 September 2014; Volume 35, pp. 776–785. [CrossRef]
12. Lin, L.; Goodrich, M.A. A Bayesian approach to modeling lost person behaviors based on terrain features in wilderness search and rescue. *Comput. Math. Organ. Theory* **2010**, *16*, 300–323. [CrossRef]
13. Keijsers, N.; Stolwijk, N.; Renzenbrink, G.; Duysens, J. Prediction of walking speed using single stance force or pressure measurements in healthy subjects. *Gait Posture* **2016**, *43*, 93–95. [CrossRef] [PubMed]
14. Feng, A.; Gordon, A.S. Latent terrain representations for trajectory prediction. In Proceedings of the 1st ACM SIGSPATIAL International Workshop on Computing with Multifaceted Movement Data, Chicago, IL, USA, 5 November 2019; pp. 1–4.
15. Wang, L.; Geng, X.; Ma, X.; Liu, F.; Yang, Q. Cross-city transfer learning for deep spatio-temporal prediction. *arXiv* **2018**, arXiv:1802.00386.
16. Torrey, L.; Shavlik, J. Transfer learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*; IGI Global: Hershey, PA, USA, 2010; pp. 242–264.
17. Xinshun, L.; Xin, H.; Hui, M.; Jing, L.; Weizhong, L.; Qingwen, Y. Automatic Cross-Domain Transfer Learning for Linear Regression. *arXiv* **2020**, arXiv:2005.04088.
18. Chopard, B.; Droz, M. *Cellular Automata*; Springer: Berlin/Heidelberg, Germany, 1998; Volume 1.
19. Vasic, J.; Ruskin, H.J. Cellular automata simulation of traffic including cars and bicycles. *Phys. A Stat. Mech. Its Appl.* **2012**, *391*, 2720–2729. [CrossRef]
20. Yue, H.; Guan, H.; Zhang, J.; Shao, C. Study on bi-direction pedestrian flow using cellular automata simulation. *Phys. A Stat. Mech. Its Appl.* **2010**, *389*, 527–539. [CrossRef]

21.	Hu, X.; Li, X.; Lu, L. Modeling the Land Use Change in an Arid Oasis Constrained by Water Resources and Environmental Policy Change Using Cellular Automata Models. *Sustainability* **2018**, *10*, 2878. [CrossRef]

22.	Bodrožić, L.; Stipaničev, D.; Šerić, M. Forest fires spread modeling using cellular automata approach. *CEEPUS Summer Sch. Mod. Trends Control* **2006**, 23–33.

23.	Han, T.; Zhao, J.; Li, W. Smart-Guided Pedestrian Emergency Evacuation in Slender-Shape Infrastructure with Digital Twin Simulations. *Sustainability* **2020**, *12*, 9701. [CrossRef]

24.	Hashimoto, A.; Abaid, N. An Agent-Based Model of Lost Person Dynamics for Enabling Wilderness Search and Rescue. In *Dynamic Systems and Control Conference*; American Society of Mechanical Engineers: New York, NY, USA, 2019; Volume 59155, p. V002T13A005.

25.	GPX the GPS Exchange Format. 2002. Available online: http://www.topografix.com/gpx.asp (accessed on 15 December 2020).

26.	Corine Land Cover (CLC) 2018 Version 2020 20u1 European Environment Agency. Available online: https://land.copernicus.eu/pan-european/corine-land-cover/clc2018 (accessed on 15 December 2020).

27.	European Digital Elevation Model (EU-DEM), Version 1.1. Available online: https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1/view (accessed on 15 December 2020).

28.	Ritter, N.; Ruth, M.; Grissom, B.B.; Galang, G.; Haller, J.; Stephenson, G.; Covington, S.; Nagy, T.; Moyers, J.; Stickley, J.; et al. Geotiff format specification geotiff revision 1.0. *SPOT Image Corp* **2000**, *1*. Available online: http://geotiff.maptools.org/spec/geotiffhome.html (accessed on 15 December 2020).

29.	Van Brummelen, G. *Heavenly Mathematics: The Forgotten Art of Spherical Trigonometry*; Princeton University Press: Princeton, NJ, USA, 2012.

30.	Tobler, W. *Three Presentations on Geographical Analysis and Modeling: Non-Isotropic Geographic Modeling Speculations on the Geometry of Geography Global Spatial Analysis*; Technical Report; National Center for Geographic Information and Analysis: Santa Barbara, CA, USA, 1993; Volume 93.

31.	Freedman, D.A. *Statistical Models: Theory and Practice*; Cambridge University Press: Cambridge, UK, 2009.

32.	GDAL Development Team. *GDAL—Geospatial Data Abstraction Library, Version 2.2.3*; Open Source Geospatial Foundation: San Michele all'Adige, Italy, 2011.

33.	QGIS Development Team. *QGIS Geographic Information System*; Open Source Geospatial Foundation: San Michele all'Adige, Italy, 2009.

34.	Van Rossum, G.; Drake, F.L., Jr. *Python Tutorial*; Centrum voor Wiskunde en Informatica Amsterdam: Amsterdam, The Netherlands, 1995; Volume 620.

35.	Gillies, S.; Ward, B.; Petersen, A.S. Rasterio: Geospatial Raster I/O for Python Programmers. 2013. Available online: https://github.com/mapbox/rasterio (accessed on 15 December 2020).