Geospatial Wildfire Risk Assessment from Social, Infrastructural and Environmental Perspectives: A Case Study in Queensland Australia

Mahyat Shafapourtehrany

Kandilli Observatory and Earthquake Research Institute, Department of Geodesy, Bogazici University, Cengelkoy, Istanbul 34680, Turkey; mahyat.shafapour@boun.edu.tr

Abstract: Although it is hard to predict wildfires, risky areas can be systematically assessed and managed. Some of the factors for decision-making are hazard, vulnerability, and risk maps, which are the end product of wildfire mapping. This study deals with wildfire risk analysis in Queensland, Australia. A review of the previous studies focusing on each aspect has been done and used with wildfire records from 2011 to 2019 in Queensland, Australia, to compile the required input models to detect risky wildfire regions. Machine learning (ML) methods of Decision Tree (DT) and Support Vector Machine (SVM) were used to perform hazard assessment. The reason was to select the most accurate outcomes for the rest of the analysis. Among accuracy assessment techniques, the Area Under Curvature (AUC) method was used to evaluate the hazard maps. Prediction rates of 89.21% and 83.78% were obtained for DT and SVM, respectively. The DT prediction value showed that the DT-hazard map was more accurate than the SVM-hazard map. Vulnerability analysis was implemented by assigning weights to each factor according to the literature. Lastly, in order to create the wildfire risk map, the hazard and vulnerability indices were combined. The risk map showed that particularly dense urbanization regions are under future wildfire risk. To perform preliminary land use planning, this output can be used by local governmental authorities.

Keywords: wildfire; hazard; risk assessment; machine learning; geospatial; Australia

1. Introduction

For more than 350 million years, wildfires have been closely associated with our environment and society [1]. Their impacts can be grouped into damage to human lives and health [2], damage to infrastructure such as buildings and transportation [3], and damage to the ecosystem (carbon level and air pollution) [4]. Due to the severe social and ecological damage inflicted by wildfires, the demand for proper management and mitigation plans has grown [5]. Recently, the number of wildfires, especially in California, Australia, and Siberia, has increased [6]. Human activities have resulted in increased atmospheric carbon dioxide levels by around 50% relative to pre-industrial levels, which contributed to about a 1.0°C increase in global warming [7]. Australia is among countries that are a part of this unequivocal global warming trend. Harris and Lucas [8] investigated variations in Australian fire weather in the period from 1973 to 2017. Their study showed that at most observation stations there were linear increased trends in the Forest Fire Danger Index (FFDI). Research by Bowman, et al. [9] stated that due to climate change the current wildfire condition might get worse. According to their research, an increase of 20% to 50% in the number of days conducive to wildfire events is expected in the future. Consequently, by the mid-21st century (2041–2070), it is expected there will be increases in high fire danger, and the number of days per year of high fire danger is expected to be increased by 35% around the world. In general, during the summer, increases in temperature and decreases in humidity are the main reason for high fire danger days [10]. The Black Summer in Australia [11], which happened between December 2019 and February 2020, was caused by
both natural and accidental ignitions. According to the recorded data of 14 January 2020, about 18.6 million hectares were burnt, over 5900 buildings were demolished, and many people were killed [12]. This condition makes Australia one of the most prone areas to wildfire [13].

An interesting study by Moritz, Batllori, Bradstock, Gill, Handmer, Hessburg, Leonard, McCaffrey, Odion and Schoennagel [4] explains how to coexist with wildfire. They believe that social and ecological systems will face serious consequences without a wildfire management plan. As humans, we are responsible for wildfire increase in some regions by changing the natural form of the environment. Therefore, the optimal solution is to learn sustainable coexistence with wildfire [14]. Wildfire, and most its negative impacts, can be controlled and minimized by proper management strategies [15]. Increasing the number of wildfires and their damage has encouraged scientists to assess and develop new wildfire risk systems. Current risk is related to the possibility of loss under current conditions [16]. Although the predicted scenarios are not certain, it is possible to estimate their likelihood [5]. Risk is determined as a function of both hazard and vulnerability [17].

To measure risk in a region, hazard and vulnerability needs to be evaluated. According to Gervasi, et al. [18] hazard is a potentially destructive phenomenon that could cause damage to lives, properties, the economy, and the environment. Regarding vulnerability, this refers to the propensity of people and the environment to be affected by hazardous events [19]. These two factors together provide the risk conditions of a region.

This study aims to evaluate the risky zones in Queensland, Australia is one of the most wildfire-hazardous regions, considering its fire history from 2011 to 2019. Social, environmental and economic factors were considered in assessing the vulnerability in this study. The reliability of each analysis is mainly affected by the data used and the methods applied. A comprehensive literature survey selected all factors related to hazard and vulnerability from previous sources. Wildfire management received considerable benefits from newly discovered and tested analytical tools that support a wide range of wildfire-related factors [20]. There are numerous methods available to perform hazard analysis. Each method has advantages and disadvantages. Some of the well-known methods in wildfire mapping are frequency ratio (FR) [21,22], the analytical hierarchy process (AHP) [23], the analytical network process (ANP) [24,25], logistic regression (LR) [26], evidential belief function (EBF) [27], index of entropy (IoE) [28], and ensemble methods [29].

Machine learning (ML) is a category of algorithms that are often used in natural hazard studies [30–32]. These methods are able to recognize data patterns and use them to predict possible future scenarios [33]. Compared to some other methods, ML algorithms are very suitable for dealing with natural hazard mapping such as wildfire [33] mapping. The reason is that in wildfire hazard mapping, many parameters with complex physical properties are involved [34]. ML is capable of handling such an analysis and is able to learn from data and perform assessments without expert rules, meaning that ML can circumvent physics-based simulation limitations [35]. A comprehensive review regarding the application of ML in wildfire mapping has been studied by Jain, Coogan, Subramanian, Crowley, Taylor and Flannigan [33]. Their research cites many references (71) that used various ML algorithms to map wildfire hazard, vulnerability, and risk. Due to the efficiency of the ML and its successful history in wildfire studies, Decision Tree (DT) and Support Vector Machine (SVM) were selected to be used in hazard assessment. The reason for considering two methods was to observe the performance of each and select the best to be used in risk assessment.

2. Study Area

Australian ecosystems, such as deserts and tropical forests, are constantly affected by wildfires [36]. Large regions of savannah in northern Australia have been devastated by intense wildfires [4]. Queensland is Australia’s second-largest state (715,000 square miles) and has been considerably affected by wildfire damages [37]. Regarding population, it is the third largest Australian state with an estimated 5.2 million residents (https://www.abs.gov.
Therefore, it was selected as an appropriate pilot area to perform wildfire analysis (Figure 1). The study area is located in the northeast of the continent (22.5752° S, 144.0848° E). Population and settlements in Queensland are concentrated along the east coast and in the southeast corner of the state [38]. The climate varies significantly across the state resulting in different climate zones [39]. Queensland has six climate zones: tropical, subtropical, temperate, grassland, equatorial, and desert [40]. Each of these climate zones is characterized by specific temperature and rainfall patterns. Figure 1 represents the study area and wildfire records from 2011 to 2019.

The Queensland government has been providing the fire history in Queensland since 1930 (Figure 2). The fires have been captured from field observations and digitization in a mapping application from reference datasets (e.g., topographic, cadastral data, satellite imageries). As can be seen in the Figure 2, there is an increasing trend in the occurrence of wildfires in this region over history.

Figure 1. Study area with wildfire locations from 2011 to 2019.

3. Methodology

Figure 3 shows the methodology flow chart of the study. The methodology starts by dividing the database into three sets: (a) wildfire hazard-related factors; (b) wildfire vulnerability-related factors, and (c) wildfire inventory. Dataset (a) consists of conditioning factors involved in wildfire hazard analysis, such as altitude, and land use/cover (LULC). Dataset (b), which includes vulnerability-related factors such as population and endangered locations, was used in vulnerability assessment. Dataset (c) is the wildfire records from 2011 to 2019 in Queensland, Australia which was divided into training and testing datasets to be used in hazard modeling and accuracy assessment, respectively. Each dataset is described in its related section. The flowchart also consists of three phases: (1) wildfire hazard analysis using DT and SVM; (2) wildfire vulnerability analysis, and (3) constructing a risk map. The following sections describe the details regarding each phase.

3.1. Data Used

3.1.1. Inventory Dataset

Collecting and managing the datasets are important steps before the generation of natural hazard maps [41]. The GIS database collection starts with creating an inventory map regarding the disasters that occurred in the past [42]. The current study’s inventory dataset represents the wildfire records between 2011 and 2019. Figure 1 shows that wildfires mostly occurred in the far north and the northeast. The inventory dataset was received in polygon format; therefore, a random point selection was applied, and 300 wildfire locations were selected. Following the literature, 70% of the wildfire inventory points were randomly allocated to the training set, and the rest, 30%, were used for testing [43].
3.1. Data Used

3.1.1. Inventory Dataset

Collecting and managing the datasets are important steps before the generation of natural hazard maps [41]. The GIS database collection starts with creating an inventory map regarding the disasters that occurred in the past [42]. The current study’s inventory dataset represents the wildfire records between 2011 and 2019. Figure 1 shows that wildfires mostly occurred in the far north and the northeast. The inventory dataset was received in polygon format; therefore, a random point selection was applied, and 300 wildfire locations were selected. Following the literature, 70% of the wildfire inventory points were randomly allocated to the training set, and the rest, 30%, were used for testing [43].

3.1.2. Hazard-Related Factors

The most effective physical properties that impact wildfire spread are fuel composition, the area’s topography, and weather [33]. Factors influencing wildfire hazards are not limited to the mentioned parameters. A good review of these factors was implemented by Pourtaghi, et al. [44]. According to the literature, there is no fixed dataset creation framework for natural hazard analysis [45]. Therefore, the literature [46], expert knowledge [23], and data availability [45] are the main influential factors for this purpose. The traditional literature-based approach was used in this research to select the primary influencing factors. To apply the wildfire hazard assessment, a spatial database of 14 wildfire hazard-related factors was designed and constructed (Figure 4). Table 1 lists the source for each factor.
Figure 4. Cont.
Figure 4. Cont.
Figure 4. Influencing factors consisting of (a) altitude, (b) slope, (c) aspect, (d) curvature, (e) TWI, (f) NDVI, (g) distance to road, (h) distance to river, (i) rainfall, (j) wind, (k) forest types, (l) LULC, (m) geology, and (n) soil.

<table>
<thead>
<tr>
<th>Table 1. Data source related to each dataset.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forest Fire Extent</strong></td>
</tr>
<tr>
<td><strong>Source</strong></td>
</tr>
<tr>
<td>Altitude 5-m spatial resolution (produced from LiDAR data)</td>
</tr>
<tr>
<td>Slope Derived from altitude</td>
</tr>
<tr>
<td>Aspect Derived from altitude</td>
</tr>
<tr>
<td>Curvature Derived from altitude</td>
</tr>
<tr>
<td>Topographic wetness index (TWI) Derived from altitude</td>
</tr>
<tr>
<td>Wind Global Wind Atlas [<a href="https://globalwindatlas.info/area/Australia">https://globalwindatlas.info/area/Australia</a>, accessed on 1 January 2020]</td>
</tr>
<tr>
<td>Distance from rivers Queensland Government Wetlands Info Website [<a href="https://www.qld.gov.au/">https://www.qld.gov.au/</a>, accessed on 1 January 2020]</td>
</tr>
<tr>
<td>Distance from roads OpenStreetMap</td>
</tr>
<tr>
<td>Soil types (1:250,000 scale) [<a href="https://www.asris.csiro.au/themes/Atlas.html">https://www.asris.csiro.au/themes/Atlas.html</a>, accessed on 1 January 2020]</td>
</tr>
<tr>
<td>LULC Queensland Land Use Mapping Program (QLUMP)</td>
</tr>
<tr>
<td>Forest types Department of Agriculture, Fisheries and Forestry (ABARES)</td>
</tr>
<tr>
<td>NDVI Landsat imagery</td>
</tr>
</tbody>
</table>
Table 1. Cont.

<table>
<thead>
<tr>
<th>Forest Fire Extent</th>
<th>Vulnerability-related factors</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Residential building density</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Commercial building density</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Universities, colleges, and schools</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Populated places</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Power generation facilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Protected areas</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Tourist points</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children and elderly population density</td>
<td>The Humanitarian Data Exchange Website (<a href="https://data.humdata.org/">https://data.humdata.org/</a>, accessed on 1 January 2022)</td>
<td></td>
</tr>
</tbody>
</table>

In general, the topography of the region is one of the most critical factors in wildfire creation, spread, and extent [47]. Topography itself influences sun exposure, temperature, precipitation, and wind [48]. Three factors of altitude (Figure 4a), slope (Figure 4b), and aspect (Figure 4c) affect the level of solar radiation and fuel moisture content [49]. Fuel preheating spread direction, and slope factors influence the speed of the wildfire. This factor also affects vertical fuel continuity and, therefore, the upward spread of fire [50]. Faster combustion takes place in areas that have sharp slopes as these areas preheat and dry up-slope fuels [51]. Vegetation cover is indirectly affected by aspect as it controls the solar radiation received [52]. In the southern hemisphere, north-facing slopes receive higher levels of solar radiation that cause the fuels to dry out more quickly [53]. Therefore, these areas have less vegetation and lighter fuel loads. On the contrary, south-facing slopes have more vegetation and greater fuel quantities. They dry more slowly due to shadows, which cause more severe wildfires.

The morphology of the topography can be measured by curvature values (Figure 4d). Curvature consists of positive values (upwardly convex), zero (flat), and negative values (upwardly concave) [54]. In wildfire studies, the wetness of the ground is essential; therefore, TWI (Figure 4e) was measured, calculated based on Equation (1) [55]:

\[ TWI = \ln \frac{\alpha}{\tan \beta} \]  

(1)

where the slope angle at the point is \( \tan \beta \) and the cumulative up-slope area draining through a point is represented by \( \alpha \).

The NDVI factor (Figure 4f) has been used to measure vegetation cover. It was calculated using Landsat-8 images according to Equation (2):

\[ NDVI = \frac{NIR - R}{NIR + R} \]  

(2)

where \( NIR \) and \( R \) values are the infrared and red bands, respectively [56].

The continuity of fire in the area can be interrupted by linear features such as rivers and roads [57]. Therefore, distance to the road (Figure 4g) and distance to the river (Figure 4h) factors were used in the assessment. Rainfall (Figure 4i) and wind (Figure 4j) are two of the most influential wildfire conditioning factors which affect wildfire generation and extension [58,59]. Rainfall affects fuel moisture, and wind controls the direction and extent of the wildfire. Cumming [60] stated that two determinants of fire behavior are weather and spatial variation in forest types. Hence, forest types (Figure 4k) across QLD were included. LULC, geology and soil, as other wildfire conditioning factors, are illustrated in Figure 4l–n, respectively. The dataset was prepared by converting all the factors to raster format with a
5 \times 5 \text{ m} \text{ cell size. Regarding the scale of soil and geology, although the original pixel sizes of soil and geology layers were greater than 5 m they covered large areas with no sudden changes in those classes. Therefore, they were resampled to 5 \times 5 \text{ m} \text{ cell size as well in order to keep the consistency in the dataset.}

In studies such as natural hazards, in which a variety of factors are involved, assessment of multicollinearity among conditioning factors is an essential task [61]. It is defined as multicollinearity when having near-linear relationships among wildfire conditioning factors [62]. Performing the analysis before solving the multicollinearity problem could lead to inaccurate outcomes. To handle multicollinearity, it is important to recognize its sources. Tolerance (TOL) and variance inflation (VIF) factors are the two most popular and highly effective methods to justify multicollinearity between the various factors [63]. In this study, the VIF and TOL methods were used to check for multicollinearity of the conditioning factors. In multicollinearity analysis, a VIF greater than 5, or a tolerance smaller than 0.2 shows the existence of a multicollinearity problem [64]. Therefore, the factors with these values should be removed from the dataset [42].

3.1.3. Vulnerability-Related Factors

Figure 5 is a schematic presentation of the wildfire vulnerability-related factors. By overlaying the hazard map over these factors, their relation to fire danger can be estimated. First, each factor was ranked according to the literature from zero to five, with five being the highest and zero being the lowest vulnerability. This dataset includes children and elderly population density (Figure 5a), residential building density (Figure 5b), commercial building density (Figure 5c), education (Figure 5d), tourist points (Figure 5e), emergency service facilities (Figure 5f), populated places (Figure 5g), power generation facilities (Figure 5h) and protected areas (Figure 5i). The “children and elderly population density” factor was prepared using children younger than five years old and elderly above sixty years old. The factors of “children and elderly population density”, “residential building density,” and “commercial building density” were created using the kernel density tool in the ArcGIS environment. A higher density represents greater vulnerability. Figure 5d represents the schools, colleges, and universities in QLD. Figure 5d and the rest of the figures were produced using the Euclidean distance tool. The closer to these features the higher the vulnerability. Similar to wildfire hazard-related factors, all factors were generated with 5 \times 5 \text{ m} \text{ cell size.}

![Figure 5. Cont.](image-url)
3.2. Hazard Assessment

The general assumption in wildfire prediction is that future wildfires will occur under the same conditions as past wildfires. According to this assumption, the correlation between wildfires happening in a region and wildfire hazard-related factors needs to be evaluated. In this research, two well-known ML methods have been used to perform the hazard analysis which will be described below.

3.2.1. Decision Tree (DT)

One of the most popular supervised learning algorithms is DT [65], which is able to perform both classification and regression analysis [66]. A DT consists of if-then-else rules with multiple branches joined by decision nodes and terminated by leaf nodes. The decision node is where the tree splits into various branches. Each branch is related to the specific decision made by the algorithm, and leaf nodes show the output of the model [67]. This could be a label for a classification problem or a continuous value in the case of a regression problem. DT divides the dataset into groups so that each group holds more or less homogeneous states of the target variable (wildfire). At each splitting step, all attributes
related to the conditioning factors are evaluated for their influence on wildfire occurrence. The recursive process continues and once completed, the final decision tree is shaped [68]. Chi-square automatic interaction detection (CHAID) [69] is one of the DT algorithms that perform Dodge separation using the chi-square test (categorical target variable) or F-test (continuous target variable). Pearson’s chi-square statistic or likelihood ratio chi-square statistic is used in the CHAID algorithm as a separation criterion in case the target variable is categorical (current case of wildfire). This algorithm was used to perform DT and consists of three steps: merging, splitting, and stopping.

3.2.2. Support Vector Machines (SVMs)

Another popular supervised learning algorithm is SVM [70]. SVM is a classifier that determines the hyperplane, a decision boundary in an n-dimensional space separating the boundary of each class, for data in n dimensions [71]. The optimal hyperplane is based on the distance between the nearest point of each class, and the decision boundary is maximized. If it is possible to separate the data by a line, then the hyperplane can be defined as $W^T x + b = 0$, where the $w$ is the weight vector, $x$ is the input vector, and $b$ is the bias. The distance of the hyperplane to the closest data point $d$, called a support vector, is defined as the margin of separation. The main objective is to recognize the optimal hyperplane that minimizes the margin. If the data are not linearly separable, kernel SVM methods first apply a set of transformations to the data to a higher dimensional space where finding this hyperplane is easier. The classification decision function is expressed as:

$$g(x) = \text{sign} \left( \sum_{i=1}^{n} y_i \alpha_i K(x_i, x) + b \right)$$

where $K(x_i, x_j)$ is the kernel function. Radial basic function (RBF) was used to perform SVM, which is one of the kernels most often used due to its proficiency and precision.

3.3. Vulnerability Assessment

As described in Section 3.1.3, vulnerability-related factors were ranked and prepared in an index format from 0 to 1. Table 2 listed the factors and their assigned weights. In the case of density factors such as residential building density, the denser the area, the higher the vulnerability. Regarding distance-related factors, for all except “emergency services facilities,” the vulnerability decreases with increasing distance. The vulnerability stage was assessed by overlaying factors that weighted the areas according to the level of vulnerability.

Table 2. Ranking the vulnerability related factors.

<table>
<thead>
<tr>
<th>Vulnerability Related Factors</th>
<th>Vulnerability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children and elderly population density</td>
<td>1 indicates the highest vulnerability</td>
</tr>
<tr>
<td>Residential building density</td>
<td>1 indicates the highest vulnerability</td>
</tr>
<tr>
<td>Commercial building density</td>
<td>1 indicates the highest vulnerability</td>
</tr>
<tr>
<td>Universities, colleges, and schools</td>
<td>0 indicates the highest vulnerability</td>
</tr>
<tr>
<td>Emergency services facilities</td>
<td>1 indicates the highest vulnerability</td>
</tr>
<tr>
<td>Populated places</td>
<td>0 indicates the highest vulnerability</td>
</tr>
<tr>
<td>Power generation facilities</td>
<td>0 indicates the highest vulnerability</td>
</tr>
<tr>
<td>Protected areas</td>
<td>0 indicates the highest vulnerability</td>
</tr>
<tr>
<td>Tourist points</td>
<td>0 indicates the highest vulnerability</td>
</tr>
</tbody>
</table>

3.4. Risk Assessment

Natural hazard risk is usually defined as the combination of hazard potential and vulnerability [72]:

$$\text{Risk} = \text{Hazard potential} \times \text{Vulnerability}$$

The derived wildfire hazard map was normalized to be from 0 to 1. Finally, a wildfire risk map was generated by multiplying hazard and vulnerability. For statistical purposes,
the wildfire risk map for QLD was represented in five categories produced by quantile method as follows: "very high", "high", "moderate", "low" and "very low" risk.

Validation

One of the important stages in susceptibility analysis and evaluating its quality is validation [73]. Especially in prediction modeling, it is critical to validate predicted outcomes so that outputs provide a meaningful interpretation concerning wildfire susceptibility [74]. The area under the curve (AUC) is a well-known method used to evaluate the correctness and reliability of wildfire susceptibility maps. AUC wildfire training and testing datasets were used to validate the results by measuring the corresponding success rates and prediction rates [75]. The success rate was measured using the training points; however, the prediction rate was that of the remaining testing wildfire points. The testing points represent independent information not used in the modeling.

The process starts by sorting the calculated wildfire hazard index values in descending order. Subsequently, those values are classified using an equal interval tool into 100 classes with 1% cumulative intervals. The classified wildfire hazard maps were crossed with the wildfire inventory maps (training and testing). Doing so, success and prediction rate curves were measured from the cross-table values and represented as continuous lines. Cumulative plots of the percentage of wildfire occurrences versus cumulative 1% equal area of ordered hazard ranks were created to measure the AUC value for each curve. To compare the results quantitatively, the areas under the curves were re-calculated with the total area equal to 100. The validation output for both methods was calculated and compared. The method with the highest AUC was selected for use in the rest of the analysis.

4. Results and Discussion

To start, multicollinearity was detected by calculating VIF and TOL for each wildfire conditioning factor. Table 3 lists the measured VIF and TOL and, according to the results, there were multicollinearity problems in the dataset. The problem was related to the factors of curvature and soil which had VIF values >5 and tolerance <0.2. Multicollinearity was corrected by omitting these factors from the dataset.

Table 3. Analysis of the multicollinearity between wildfire hazard-related factors.

<table>
<thead>
<tr>
<th>Conditioning Factor</th>
<th>VIF</th>
<th>TOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altitude</td>
<td>3.21</td>
<td>0.31</td>
</tr>
<tr>
<td>Slope</td>
<td>1.82</td>
<td>0.55</td>
</tr>
<tr>
<td>Aspect</td>
<td>2.35</td>
<td>0.42</td>
</tr>
<tr>
<td>Curvature</td>
<td>10.85</td>
<td>0.09</td>
</tr>
<tr>
<td>TWI</td>
<td>1.22</td>
<td>0.82</td>
</tr>
<tr>
<td>NDVI</td>
<td>4.11</td>
<td>0.24</td>
</tr>
<tr>
<td>Distance to road</td>
<td>1.6</td>
<td>0.62</td>
</tr>
<tr>
<td>Distance to river</td>
<td>1.04</td>
<td>0.96</td>
</tr>
<tr>
<td>Rainfall</td>
<td>3.16</td>
<td>0.31</td>
</tr>
<tr>
<td>Wind</td>
<td>1.64</td>
<td>0.61</td>
</tr>
<tr>
<td>Forest types</td>
<td>1.23</td>
<td>0.81</td>
</tr>
<tr>
<td>LULC</td>
<td>1.88</td>
<td>0.53</td>
</tr>
<tr>
<td>Geology</td>
<td>1.53</td>
<td>0.65</td>
</tr>
<tr>
<td>Soil</td>
<td>19.4</td>
<td>0.05</td>
</tr>
</tbody>
</table>

The relationships between wildfire occurrence and hazard-related factors have been evaluated using DT and SVM. The produced maps were reclassified into five classes by quantile for ease of visual interpretation. Figure 6 illustrates the final hazard maps. As can be seen in both figures, the north and southeast parts of the study area have been recognized as very highly hazardous zones. However, the hazardous areas around QLD detected by DT are more than SVM in terms of coverage. The correctness of these methods was evaluated through accuracy assessment.
The outcomes of the wildfire hazard mapping were evaluated by comparing them with the existing wildfire locations. The AUC method was applied and both success and prediction rates were calculated. The success rate was measured by comparing the wildfire grid cells (210 cells) in the training dataset with the DT and SVM wildfire hazard maps. The success rate shows how the wildfire hazard outcomes fit the training dataset. Since the training dataset was used for training the DT and SVM model, it may not be a precise approach to assess the prediction capacity of the methods. The prediction rate is a suitable way to test how well the wildfire models and wildfire conditioning factors predict possible future wildfire areas. The wildfire grid cells that were not used in the training process (90 cells) were used to measure the prediction rate. For both success and prediction rates, if the AUC is equal to 100 it represents perfect accuracy. The AUC values (Figure 7) obtained from two hazard maps revealed that DT showed the highest success and prediction rates (91.80% and 89.21%, respectively). The SVM success rate (87.01%) was close to the DT success rate; however, SVM’s prediction rate (83.78%) was 6% lower than that of DT. In fact, both methods provided reasonable accuracies; however, DT results were used in risk assessment due to their better performance. The proficiency of ML methods has been proven by the literature as well. Ghorbanzadeh, et al. [76], Gholamnia, et al. [77], Watson, et al. [78] are some examples of studies in which ML methods were utilized in the same field of research. Among the majority of the assessed research, ML provided the most reliable outcomes compared to other techniques. A study by Jain, Coogan, Subramanian, Crowley, Taylor and Flannigan [33] reviewed ML applications in wildfire science and management. According to their comprehensive study, the advantages of ML methods make them suitable for use in a variety of wildfire problem domains such as fire recognition and mapping, climate change and its impact on fire weather, wildfire hazard and risk, wildfire prediction, impacts of wildfire, and wildfire management. Their study is not limited to positive aspects but also discusses the limitations of various ML approaches related to the size of the dataset, analysis requirements, generalizability, and interpretability. More than forty ML algorithms were mentioned in their research. However, SVM and DT were among the most robust techniques in wildfire studies.

Although the advantages of individual ML methods are apparent, ensemble/hybrid ML techniques could enhance their strength. Mohajane, et al. [79] is one of the studies that tested hybrid DT and SVM using frequency ratio. Their outcomes showed that their developed hybrid models can increase the accuracy and performance of wildfire susceptibility studies. Therefore, for future analysis, ensemble modeling and its impact on the final outputs was considered as well.

As an additional step, classes of “High” and “Very High” derived from DT were extracted and overlaid with vulnerability feature maps to visually assess their vulnerability. Figure 8 shows that many of these features are located in high and very high susceptible zones. As can be seen visually, many of the features are located in these zones.

To create the final wildfire vulnerability map, all weighted and normalized vulnerability-related factors were overlaid, and a vulnerability map was created using Raster Calculator in a GIS environment. The vulnerability maps had classifications of “very high”, “high”, “moderate”, “low” and “very low” as illustrated in Figure 9b. As the final step, the derived hazard (Figure 9a) and vulnerability maps were multiplied, and a wildfire risk map has been created (Figure 9c).

It is apparent that high wildfire risk areas have many settlements, and in these regions the probability of wildfire occurrence is very high. Risky regions must be brought to the notice of the government and public so that they are aware of future wildfire possibilities. Doing so, lots of lives and properties could be saved. Figure 10 also illustrates the distribution of wildfire areas with respect to wildfire risk areas. It shows that 10.2% and 8.5% of QLD is located in “very high” and “high” risk zones respectively.
The success rate shows how the wildfire hazard outcomes fit the training data. As an additional step, classes of "High" and "Very High" derived from DT were extracted and overlaid with vulnerability feature maps to visually assess their vulnerability. Figure 8 shows that many of these features are located in high and very high susceptible zones. As can be seen visually, many of the features are located in these zones.

Figure 6. Wildfire hazard maps.

Figure 7. AUC.

According to the literature, the majority of people living in high wildfire-risk areas are aware of their exposure. However, this does not mean that they try to act to mitigate risks. The 2009 Black Saturday fires in Victoria are a good example in which most people understood about the risky areas, but only a few took action when the highest-level warning was issued. Therefore, all parties in society (academics, governments, and citizens) have to work together to achieve optimal mitigation actions (Haque et al., 2021 [1]). It should be taken into consideration that not only human life, but also a wide range of species, can be affected by wildfire. Especially in QLD, due to a range of threatening processes, some species are decreasing in numbers and are at risk of extinction. For instance, on 30 April 2021, there were 1020 threatened species (236 animals and 784 plants) listed as threatened under the Queensland Nature Conservation Act 1992. Therefore, it is necessary to consider all possible damage, recognize risky areas, conduct comprehensive education, and plan wisely to control this disaster.
Regarding the limitations of this study and possible improvement, the output of this study could be more precise if projected climate change was available for use in the analysis. The reason is that global warming and other climatic factors constantly change and can considerably affect future wildfire occurrences.

Figure 8. Overlaid vulnerability-related factors with “high” and “very high” wildfire susceptibility classes.
To create the final wildfire vulnerability map, all weighted and normalized vulnerability-related factors were overlaid, and a vulnerability map was created using Raster Calculator in a GIS environment. The vulnerability maps had classifications of “very high”, “high”, “moderate”, “low” and “very low” as illustrated in Figure 9b. As the final step, the derived hazard (Figure 9a) and vulnerability maps were multiplied, and a wildfire risk map has been created (Figure 9c).

Figure 9. Wildfire risk map.
5. Conclusions

Wildfire risk analysis can help scientists and managers better understand the timing, location, and potential effects of wildfires on economic and ecological systems. It can address forest management issues and disclose tradeoffs that other analysis techniques may not account for. This study focused on areas with high wildfire risk in QLD, Australia, to provide proper knowledge for scientists and citizens. Two datasets of wildfire hazard-related factors and vulnerability-related factors were prepared and used in the analysis. DT and SVM were applied to assess the hazard and extract the correlation between wildfire conditioning factors and wildfire inventory. Using AUC, optimal accuracy was achieved from DT with 89.21% prediction accuracy. Therefore, the output from DT was used in the rest of the analysis. Vulnerability-related factors were ranked and prepared as an index from 0 to 1. The ranking factors were then overlaid to extract vulnerability areas. Finally, both the DT-derived hazard map and vulnerability map combined to create the risk map. Results revealed that most of the urban regions in QLD are at high risk of wildfire. These areas need proper mitigation action to avoid or minimize future wildfire damage. In conclusion, regardless of the surrounding ecosystem conditions, all communities can better coexist with fire by participating in risk awareness and continuing education.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References
5. Miller, C.; Ager, A.A. A review of recent advances in risk analysis for wildfire management. Int. J. Wildland Fire 2012, 22, 1–14. [CrossRef]


34. Chapelle, O.; Vapnik, V.; Bousquet, O.; Mukherjee, S. Choosing multiple parameters for support vector machines. *Mach. Learn.* 2002, 46, 131–159. [CrossRef]


43. Eskandari, S.; Amir, M.; Sadhasivam, N.; Pourghasemi, H.R. Comparison of new individual and hybrid machine learning algorithms for modeling and mapping fire hazard: A supplementary analysis of fire hazard in different counties of Golestan Province in Iran. *Nat. Hazards* 2020, 104, 305–327. [CrossRef]

44. Pourtaghi, Z.S.; Pourghasemi, H.R.; Aretano, R.; Semeraro, T. Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. *Ecol. Indic.* 2016, 64, 72–84. [CrossRef]


52. Lambin, E.F.; Geist, H.J.; Lepers, E. Dynamics of land-use and land-cover change in tropical regions. *Annu. Rev. Environ. Resour.* 2003, 28, 205–241. [CrossRef]


64. Kim, J.H. Multicollinearity and misleading statistical results. Korean J. Anesthesiol. 2019, 72, 558−569. [CrossRef]
70. Noble, W.S. What is a support vector machine? Nat. Biotechnol. 2006, 24, 1565−1567. [CrossRef]
76. Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Aryal, J. Forest fire susceptibility and risk mapping using social/infrastructural vulnerability and environmental variables. Fire 2019, 2, 50. [CrossRef]
77. Gholamnia, K.; Gudiyangada Nachappa, T.; Ghorbanzadeh, O.; Blaschke, T. Comparisons of diverse machine learning approaches for wildfire susceptibility mapping. Symmetry 2020, 12, 604. [CrossRef]
79. Mohajane, M.; Costache, R.; Karimi, F.; Pham, Q.B.; Essahlaoui, A.; Nguyen, H.; Laneve, G.; Oudija, F. Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area. Ecol. Indic. 2021, 129, 107869. [CrossRef]