A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area

Dieu Tien Bui [a,*], Quang-Thanh Bui [b], Quoc-Phi Nguyen [c], Biswajeet Pradhan [d,f], Haleh Nampak [d], Phan Trong Trinh [e]

[a] Geographic Information System Group, Department of Business Administration and Computer Science, University College of Southeast Norway, Hallvard Eikas Plass, N-3800 Bø i Telemark, Norway
[b] Faculty of Geography, VNU University of Science, 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam
[c] Department of Environmental Sciences, Hanoi University of Mining and Geology, Duc Thang, Bac Tu Liem, Hanoi, Vietnam
[d] Department of Civil Engineering, Geospatial Information Science Research Center (GISRC), Faculty of Engineering, University Putra Malaysia, Serdang, Selangor Darul Ehsan 43400, Malaysia
[e] Institute of Geological Sciences, Vietnam Academy of Sciences and Technology, 84 Chua Lang Street, Dong da, Hanoi, Vietnam
[f] Department of Geoinformation Engineering, Choongmu-gwan, Sejong University, 209 Neungdong-Ro Gwangjin-Gu, Seoul 05006, Republic of Korea

ARTICLE INFO

ABSTRACT

This paper proposes and validates a novel hybrid artificial intelligent approach, named as Particle Swarm Optimized Neural Fuzzy (PSO-NF), for spatial modeling of tropical forest fire susceptibility. In the proposed approach, a Neural Fuzzy inference system (NF) was used to establish the forest fire model whereas Particle Swarm Optimization (PSO) was adopted to investigate the best values for the model parameters. Tropical forest at the province of Lam Dong (Central Highland of Vietnam) was used as a case study. For this purpose, historic forest fires and ten ignition factors (slope, aspect, elevation, land use, Normalized Difference Vegetation Index, distance to road, distance to residence area, temperature, wind speed, and rainfall) were collected from various sources to construct a GIS database, and then, the database was used to develop and validate the proposed model. The performance of the forest model was assessed using the Receiver Operating Characteristic curve, area under the curve (AUC), and several statistical measures. The results showed that the proposed model performs well, both on the training dataset (AUC = 0.932) and the validation dataset (AUC = 0.916). The usability of the proposed model was further assessed through comparisons with those derived from two benchmark state-of-the art machine learning methods, Random Forests (RF) and Support Vector Machine (SVM). Because the performance of the proposed model is better than the two benchmark models, we concluded that the PSO-NF model is a valid alternative tool that should be considered for tropical forest fire susceptibility modeling. The result in this study is useful for forest planning and management in forest fire prone areas.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Forest fire is an environmental problem that poses threats to the safety of human life, infrastructures, and the environment (Podur et al., 2003), and is considered as an important agent of pattern formation of forests, such as succession and regeneration (Huebner et al., 2012). Due to changes in climate i.e. less rainfall and/or increment of day temperature, longer dry season, and interventions of human activities (Argañaraz et al., 2015), the trend and frequency of forest fires are increased and reached an alarming rate in many areas of the world (Argañaraz et al., 2015; Arpaci et al., 2014; Littell et al., 2016; Mavsar et al., 2013; Moritz et al., 2012). Therefore, it is necessary to predict forest fires as accurately as possible. This will assist local authorities in forest management and planning, resource allocation, emergency services, and early warning systems (Eastaugh and Hasenauer, 2014).

Various approaches have been proposed for modeling of forest fire behaviors and they can be classified into three groups including physics-based method, statistical method, and machine learning method. The physics-based method simulates and predicts

* Corresponding author.
E-mail addresses: Dieu.T.Bui@hit.no, BuiTienDieu@gmail.com (D. Tien Bui).

potential fire behaviors through a set mathematical equations of fluid mechanics, combustions of canopy biomass, and heat transfer mechanisms (Pastor et al., 2003). Therefore, they are capable to model fire behaviors in both space and time. The most common used physics-based models for forest fires are EMBYR (Hargrove, 1994), FARSITE- Fire Area Simulator (Keane et al., 1998), FDS (McGrattan et al., 2000), FIRETEC (Linn et al., 2002), FireStation (Lopes et al., 2002), and LANDIS-II (Sturtevant et al., 2009). The main disadvantage of these models is that it is difficult to quantify the magnitude of the errors (Massada et al., 2011). In addition, physics-based models require detailed data, i.e. locations and dimensions of trees, fuel mass, soil moisture that are difficult to collect for large areas (Pimont et al., 2016).

Statistical method is more suitable for forest fire modeling when study areas are large, especially in combinations with geographic information system (GIS) technology (Duarte et al., 2016; Teodoro et al., 2015; Teodoro and Duarte, 2013), due to its capacity to collect and process spatial data of large regions with different scales and resolutions (Bonham-Carter, 2014; Chuvieco et al., 2010; Tien Bui et al., 2016d; Verde and Zêzere, 2010; Wittenberg and Malkinson, 2009). Consequently, various statistical methods and techniques have been adopted for forest fire modeling, such as Poisson regression (Wotton et al., 2003), generalized Pareto distribution (Bermudez et al., 2009), favorability functions (Verde and Zêzere, 2010), generalized additive model (Vilar et al., 2010), Monte Carlo simulations (Conedera et al., 2011), multiple linear regression (Oliveira et al., 2012), logistic regression (Arndt et al., 2013; Chuvieco et al., 2010; Pourghasemi, 2015), weights of evidence (Amatulli et al., 2007), and geographically weighted regression (Oliveira et al., 2014). However, forest fire regime is a typical complex and non-linear process that is difficult to assess and predict, therefore prediction accuracy of these models is not always satisfied.

Due to critical accuracy of forest fire models, machine learning methods have been explored and investigated such as Decision Tree learning (Camp et al., 1997), Kernel methods (Gonzalez-Olabarria et al., 2012), Random Forests (Arpaci et al., 2014; Oliveira et al., 2012), Kernel logistic regression (Tien Bui et al., 2016b), Maximum Entropy (Arpaci et al., 2014), artificial neural networks (ANN) (Cheng and Wang, 2008; Satir et al., 2015; Vasconcelos et al., 2001). In general, performance of machine leaning models is better when compared with the statistical models (Massada et al., 2013). Nevertheless, prediction of forest fire at regional scales is still difficult due to multiple and complex interactions of ignition factors. In addition, forest fire modeling requires collecting ignition factors with different resolutions i.e. climate and weather conditions (temperature, humidity, rainfall, and wind), topography (elevation, slope, aspect), and land use (Ganteaume et al., 2013). Therefore, it is difficult to eliminate uncertainties as well as imprecisions (Tien Bui et al., 2016c). As a result, these above models have difficulties in processing fire ignition maps with inaccuracies of GIS information, such as spatial errors due to age of data, different scales and resolutions (Benz et al., 2004; Short, 2014; Zald et al., 2014). To address this issue, fuzzy reasoning has been proposed (Loboda and Csiszar, 2007). However, the subjective determination of fuzzy membership values prevents this method from producing high accuracy results. Therefore, developments of new prediction models for forest fires that lead to deal with uncertainties and imprecisions and improve fire prediction capabilities are highly necessary.

This paper attempts to partially fill this gap in literature by proposing a novel hybrid artificial intelligent model, named as Particle Swarm Optimized Neural Fuzzy (PSO-NF) model, for spatial prediction of tropical forest fire susceptibility with a case study at the province of Lam Dong in the Central Highland region (Vietnam), a typical province that has seriously affected by forest fires during the last ten years (Dan et al., 2014). In the proposed approach,

the forest fire model is established based on a neural fuzzy inference system that combined fuzzy logic and neural networks (Jang, 1993). Development of the neural fuzzy model has a challenge of finding the best premise and consequent parameters that strongly influence the model performance, therefore, this study proposes Particle Swarm Optimization (PSO)(Eberhart and Kennedy, 1995) for optimizing the neural fuzzy model. PSO is considered a powerful algorithm that has been widely used in soft computing for model optimization, such as in rainfall–runoff modeling (Taormina and Chau, 2015), flood susceptibility modeling (Tien Bui et al., 2016e), dam behavior modeling (Bui et al., 2016), and slope instability analysis (Gordan et al., 2016). Consequently, the proposed model is a powerful inference system that has capabilities to process uncertainty data with high accuracy. Finally, the usability of the proposed model is confirmed through comparisons with benchmarks, Random Forests (RF) and Support Vector Machine (SVM), and concluding remarks are provided.

2. The study area and data used

2.1. Description of the study area

Lam Dong province (Fig. 1) is located on the southern part of the Central Highland region of Vietnam, between latitudes 11°12′00″N and 12°15′00″N, and between longitudes 107°15′00″E and 108°45′00″E. The province occupies 9805.4 km^2 with complex topographical variations. The altitude varies between 120 m and 2280 m above sea level, with the mean of 907.6 m and the standard deviation of 392.1 m.

Climate in the province is influenced by tropical monsoon with moderate temperatures and high humidity. The local climate varies with altitude and can be divided into two distinct seasons, rainy and dry seasons. The rainy season extends from May to November, whereas the dry season is from December to April. The temperature varies significantly across the region, ranging from 18°C to 25°C with mild and cool weathers. The rainfall in the rainy season is up to 90% of the total annual rainfall, with values between 1600 and 2700 mm per year. The average relative humidity of the year ranges from 85% to 87%.

The forest coverage of the province is approximately 60% of the total study area (Vu et al., 2013) whereas agricultural land and populated areas covers around 28% and 6%, respectively. The remaining is for other types of land covers. Tree species are dominated by *Suzygium, Dipterocarpus, Anisopkera cochinchinensia, and Schima superba Gardner & Champ* (between 1000 and 1300 m), *Pinus merkusii* (600 and 1000 m), *Pinus khasya* (>1000 m), *Dipterocarpus obtusifolius and Shorea obtusa* (1300 m).

According to Vu et al. (2013), deforestation and forest degradation continue to be key threats to the forest cover of the province during the period 1990–2010. Although the plantation forest was increased 81.7% (5.1% of the total study area), however the total forest area was reduced by 118067.9 ha (12.04% of the total study area). Broadleaf forest, bamboo forest, and coniferous forest were reduced by 30.5%, 37.1%, and 28.2% respectively. These degradations are equivalent to 17% of the total study area and causes of forest loss can be named as fire, illegal logging, land conversion and inefficient management (Vu et al., 2013). In fact, there is a significant pressure of population growth on forest resources due to increased demands for residential lands and lands for productions.

2.2. Historical forest fires and ignition factors

2.2.1. Historical forest fires

Because prediction models for forest fire susceptibility are developed based on relationship analysis of historical forest fires