

## FRACTIONAL SNOW COVER MAPPING BY ARTIFICIAL NEURAL NETWORKS AND SUPPORT VECTOR MACHINES

B. B. Çiftçi<sup>a</sup>, S. Kuter<sup>b, d, \*</sup>, Z. Akyürek<sup>c, e</sup>, G.-W. Weber<sup>d, f</sup>

<sup>a</sup> Cankiri Karatekin University, Graduate School of Natural and Applied Sciences, Department of Forest Engineering, 18200, Cankiri, Turkey - boraberkeybb@hotmail.com

<sup>b</sup> Çankırı Karatekin University, Faculty of Forestry, Department of Forest Engineering, 18200, Çankırı, Turkey - semihkuter@karatekin.edu.tr

<sup>c</sup> Middle East Technical University, Faculty of Engineering, Department of Civil Engineering, 06800, Ankara, Turkey - zakyurek@metu.edu.tr

<sup>d</sup> Middle East Technical University, Institute of Applied Mathematics, 06800, Ankara, Turkey - gweber@metu.edu.tr

<sup>e</sup> Middle East Technical University, Graduate School of Natural and Applied Sciences, Department of Geodetic and Geographic Information Technologies, 06800, Ankara, Turkey

<sup>f</sup> Poznan University of Technology, Faculty of Engineering Management, Department of Marketing and Economic Engineering, 60-965, Poznan, Poland

**KEY WORDS:** Fractional Snow Cover Mapping, MODIS, Landsat ETM+, Artificial Neural Networks, Multilayer Perceptron, Support Vector Machines, Support Vector Regression.

### ABSTRACT:

Snow is an important land cover whose distribution over space and time plays a significant role in various environmental processes. Hence, snow cover mapping with high accuracy is necessary to have a real understanding for present and future climate, water cycle, and ecological changes. This study aims to investigate and compare the design and use of artificial neural networks (ANNs) and support vector machines (SVMs) algorithms for fractional snow cover (FSC) mapping from satellite data. ANN and SVM models with different model building settings are trained by using Moderate Resolution Imaging Spectroradiometer surface reflectance values of bands 1-7, normalized difference snow index and normalized difference vegetation index as predictor variables. Reference FSC maps are generated from higher spatial resolution Landsat ETM+ binary snow cover maps. Results on the independent test data set indicate that the developed ANN model with hyperbolic tangent transfer function in the output layer and the SVM model with radial basis function kernel produce high FSC mapping accuracies with the corresponding values of  $R = 0.93$  and  $R = 0.92$ , respectively.

### 1. INTRODUCTION

One challenging issue in snow mapping is the trade-off between the temporal and spatial resolution of satellite imageries. Since high spatial resolution reduces the temporal resolution, it eventually limits timely detection of the changes in snow cover. Whereas, high temporal resolution data reduces the precision of snow cover maps due to low spatial resolution.

In order to deal with this problem, various fractional snow cover (FSC) mapping approaches have been proposed and applied to low or moderate resolution images such as spectral unmixing (Painter et al., 2003) and empirical normalized difference snow index (NDSI) (Salomonson and Appel, 2004) methods. In contrast to binary classification approach where a pixel is labeled as either snow-covered or snow-free, the true class distribution can be well estimated in FSC mapping even though the precise location of class fractions within each coarse resolution pixel still remains unknown (Verbeiren et al., 2008).

Artificial neural networks (ANNs) are a machine learning algorithm that can generate an information processing model by resembling the knowledge acquisition mechanism of the brain from the environment (Haykin, 2009), and have been extensively used in various disciplines (Dongale et al., 2015; Kar, 2015; Raji and Chandra, 2016). ANNs have also gained popularity and been frequently employed in various RS applications, like change detection (Dai and Khorram, 1999), land cover classification (Paola and Schowengerdt, 1995) and

identification of clouds (Lee et al., 1990). Recently, there has also been an increasing trend in the use of ANNs for FSC mapping (Czyzowska-Wisniewski et al., 2015; Dobreva and Klein, 2011; Moosavi et al., 2014).

Many types of ANNs have been developed and multilayer perceptron (MLP), a feed-forward neural network model, is the most frequently employed one in RS applications (Kavzoglu and Mather, 2003; Mas and Flores, 2008).

Originated from Data Mining and Statistical Learning Theories, support vector machines (SVMs) (Vapnik, 1995) are a state-of-the-art supervised nonparametric classification and regression tool which has been successfully used in various applications such as seismology (Fisher et al.), energy and finance (Kumar et al., 2016; Mustafa et al., 2014; Yuan and Lee, 2015), face identification, text categorization, bioinformatics and database marketing (Campbell and Ying, 2011)..

Although SVMs were initially proposed for pattern recognition tasks, successful results were also obtained in regression and time series prediction applications (Drucker et al., 1997b; Mattera and Haykin, 1999; Müller et al., 1997). SVMs have also quickly drawn the attention of RS community, and they have many successful applications in image classification (Foody and Mathur, 2004; Huang et al., 2002) and regression (Bruzzone and Melgani, 2005; Kaheil et al., 2008; Zheng et al.,

2008) tasks in RS; however, they have not yet been employed for FSC estimation.

The primary focus of this research is to perform a comparison between ANNs and SVMs for FSC mapping applications in RS. As in all nonparametric classification and regression tools, ANNs and SVMs have also “model tuning” parameters that directly affect their generalization ability on unseen data. Thus, the secondary aim is to investigate the optimal combination of basic model building parameters of these methods for FSC mapping applications in RS. The rest of this paper is organized as follows: The study area, satellite image data set with pre-processing stages, brief overviews of both MLP networks and SVMs, and the basic experimental design including sampling strategy, training and testing of ANN and SVM models are described in Section 2. The results are represented and discussed in Section 3. Finally, conclusion and outlook for future studies are given in Section 4.

## 2. DATA SET AND METHODOLOGY

### 2.1 Study Area

The study area is located on the Ilgaz Mountains lying within the borders of Cankiri and Kastamonu provinces of Turkey, which also covers Kursunlu forest sub-district region (cf. Figure 1). It encompasses an area of ~8500 km<sup>2</sup> with altitude values ranging from 354 to 2326 m above mean sea level.

The vegetation of the study area mainly consists of forest, underwood and alpine flora. The low altitude parts of the northern slopes are generally covered with Oak (*Quercus* spp.), European Black Pine (*Pinus nigra* ssp. *pallasiana*) and Uludag Fir (*Abies nordmanniana* ssp. *bornmulleriana*) forests. At altitudes from 1000-1300 m, Hornbeam (*Carpinus* spp.) and Beech (*Fagus orientalis* L.) are dominantly observed with some other firewood and deciduous plants. At 1500 m and above, pure or mixed forests are established by Black Pine (*Pinus nigra* ssp. *pallasiana*) and Scots Pine (*Pinus sylvestris* L.) (Kuter, 2008).

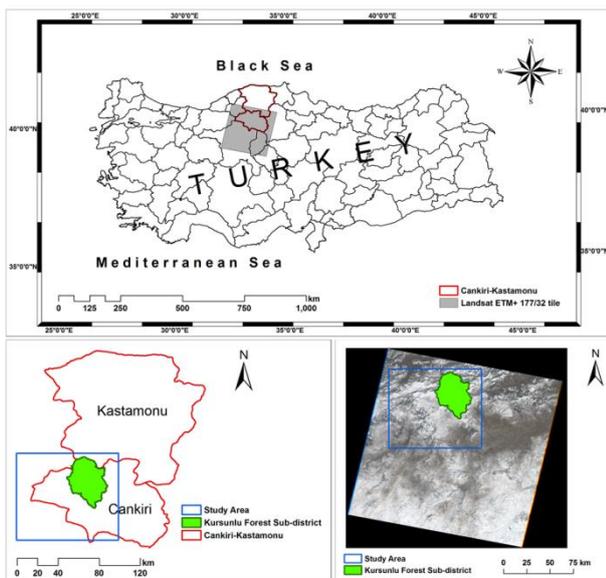


Figure 1: Study area.

Average annual temperature at the region is about 9.8 °C. The coldest month is January with an average temperature of -0.8 °C, whereas the warmest month is July with 20 °C. Average annual rainfall is 486 mm according to the recordings of meteorological station in Kastamonu. At the mountain peaks, the rainfall is about 1200 mm. Snow cover remains for about six months with thickness reaching about 1 m on slopes due to the effect of Central Anatolian climate (Aydinozu et al., 2011; Kuter and Kuter, 2016).

### 2.2 Satellite Imagery

Moderate resolution imaging spectroradiometer (MODIS) on the Terra satellite has 36 spectral bands ranging in wavelength from 0.4 to 14.4 μm at varying spatial resolutions (Salomonson et al., 2006) and has been extensively used for mapping global snow cover since its launch in 1999. The MODIS snow mapping algorithm is mainly based on NDSI (cf. Eqn. (1)), where the MODIS bands 4 and 6 are used, along with a series of threshold tests and the MODIS cloud mask, and each MODIS pixel with 500 m ground resolution is classified as either snow or non-snow (Hall et al., 2002).

$$NDSI_{MODIS} = \frac{band4 - band6}{band4 + band6} \quad (1)$$

Two MODIS and Landsat 7 ETM+ image pairs taken on December 26, 2002 and February 28, 2003 are used in the study. Spatial subset of MODIS and ETM+ pair taken in 2002 is used for training, and that of 2003 image is used for testing of ANN and SVM models. The details of the satellite image data can be found in Table 1.

#### 2.2.1 Study Area Image Pre-processing and Reference FSC Maps

All MODIS surface reflectance images are re-projected to a common UTM projection with WGS84 datum in order to match the projection of the corresponding Landsat scenes. In addition to the surface reflectance values of MODIS bands 1-7, NDSI and normalized difference vegetation index (NDVI) (cf. Eqn. (2)) are used as input variables (i.e., predictors). Cloud-covered, cloud shadow, water and bad quality pixels are identified by spatial masks obtained from the associated MODIS ancillary data and excluded from further analysis.

Training Set	
Date acquired	26.12.2002
ETM+ path/row	177/32
Total number of available training samples	8872
Total number of training samples after sampling	254
Testing Set	
Date acquired	28.02.2003
ETM+ path/row	177/32
Total number of available test samples	27338

Table 1: Training and test data used in the analysis.

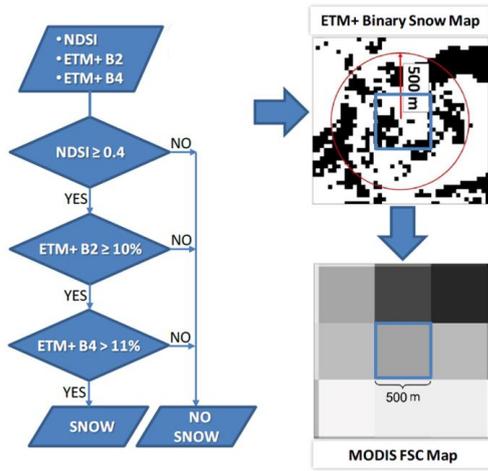


Figure 2: Derivation of reference MODIS FSC maps from binary-classified higher resolution Landsat ETM+ images.

$$NDVI_{MODIS} = \frac{band2 - band1}{band2 + band1} \quad (2)$$

Reference FSC maps are obtained by binary classification of higher spatial resolution Landsat 7 images. In this binary classification scheme, which is the equivalent of the original MODIS binary snow mapping algorithm, a pixel is labeled as snow if its NDSI  $\geq 0.4$  and Landsat ETM+ band 2 reflectance  $\geq 10\%$  and band 4 reflectance  $> 11\%$  (Klein et al., 1998). By calculating the percentage of snow-covered Landsat pixels within a circular area of 500 m radius centered at a MODIS pixel, FSC value for each MODIS pixel is obtained (cf. Figure 2). Table 2 gives detailed information on the predictor variables and the response used in the analysis.

MODIS data used as predictor variables	
Band 1	620-670 nm
Band 2	841-876 nm
Band 3	459-479 nm
Band 4	545-565 nm
Band 5	1230-1250 nm
Band 6	1628-1652 nm
Band 7	2105-2155 nm
NDSI	$(band4 - band6) / (band4 + band6)$
NDVI	$(band2 - band1) / (band2 + band1)$
Landsat ETM+ data to derive reference FSC maps used as response variable	
Band 2	520-600 nm
Band 4	770-900 nm
Band 5	1550-1750 nm

Table 2: MODIS and Landsat ETM+ data used in ANN and SVM models.

### 2.3 Multilayer Perceptron (MLP)

In this sub-section, the basics of MLP neural networks are introduced based on (Agirre-Basurko et al., 2006; Gardner and Dorling, 1998; Haykin, 2009; Mas and Flores, 2008; Rafiq et al., 2001).

An MLP is made of a set of input nodes (i.e., the input layer), one or more sets of computation neurons (i.e., the hidden

layers), and one set of computation/output nodes (i.e., the output layer) (cf. Figure 3). Connections are always made forward, on a layer-by-layer basis. A neuron on layer  $i$  on a strict feed-forward MLP always connects to a neuron on layer  $i + 1$ . Neither recurrent connections nor skipping layers are allowed.

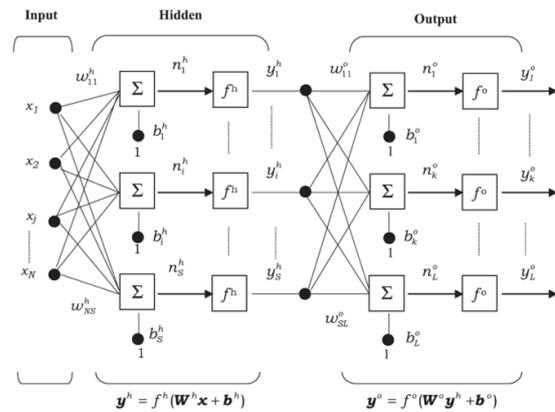


Figure 3: Basic diagram of an MLP neural network (Agirre-Basurko et al., 2006).

The training of an MLP by backpropagation algorithms involves two stages. In the first stage, also known as the forward phase, random numbers are assigned to synaptic weights of the network using a uniform random distribution and the input signal is propagated through the network, layer by layer, until it reaches the output. As a result, changes are confined to the activation potentials and outputs of the neurons in the network. Then in the second stage, namely, backpropagation, an error signal is obtained by comparing the output of the network with a desired response. The resulting error signal is transmitted through the network layer by layer, yet the propagation is performed in the backward direction this time. In this stage, successive adjustments are made to the synaptic weights of the network. The output of an MLP network is given by Eqn. (3) as follows:

$$y_k^o = f_k^o \left( b_k^o + \sum_{i=1}^S w_{ik}^o y_i^h \right) = f_k^o \left( b_k^o + \sum_{i=1}^S w_{ik}^o f_i^h \left( b_i^h + \sum_{j=1}^N w_{ji}^h x_j \right) \right) \quad (3)$$

where  $k = (1, 2, \dots, L)$ ,  $h$  and  $o$  denotes the elements of the hidden and the output layers, respectively,  $w_{ji}^h$  is the weight that connects the  $j$ th node of the input layer with the  $i$ th neuron of the hidden layer,  $w_{ik}^o$  is the weight connecting the  $i$ th neuron of the hidden layer with the  $k$ th node of the output layer,  $f_i^h$  is the transfer function of the  $i$ th neuron of the hidden layer,  $f_k^o$  is the transfer function of the  $k$ th node of the output layer,  $b_i^h$  and  $b_k^o$  are the biases of the  $i$ th neuron in the hidden layer and the  $k$ th node in the output layer, respectively. Furthermore,  $n_i^h$  is the excitation level of  $i$ th neuron of the hidden layer (cf. Eqn. (4)),  $n_k^o$  is the excitation level of  $k$ th node of the output layer (cf. Eqn. (5)),  $y_i^h$  and  $y_k^o$  are the outputs of the  $i$ th neuron of the hidden layer and the  $k$ th node of the output layer,

respectively,  $x_j$  is the  $j$ th node of the input layer,  $N$  and  $L$  indicate the number of nodes in the input and the output layers, respectively, and finally,  $S$  is the number of neurons in the hidden layer:

$$n_i^h = b_i^h + \sum_{j=1}^N w_{ji}^h x_j, \quad (4)$$

$$n_k^o = b_k^o + \sum_{i=1}^S w_{ik}^o y_i^h. \quad (5)$$

The most widely employed nonlinear transfer functions are hyperbolic tangent (tansig) and logarithmic sigmoid (logsig) and functions:

$$\text{tansig}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad (6)$$

$$\text{logsig}(x) = \frac{1}{1 + e^{-x}}. \quad (7)$$

Tansig and logsig are differentiable and bounded within the range from -1 to 1, and from 0 to 1, respectively (cf. Figure 4). As soon as the normalization of the inputs and the outputs is completed, the learning process starts. This stage can be considered as an optimization process, in which the error function  $E$  has to be minimized:

$$E = \frac{1}{L} \sum_{k=1}^L (t_k - y_k^o)^2, \quad (8)$$

where  $(t_1, t_2, \dots, t_L)^T$  is the target vector of the MLP.

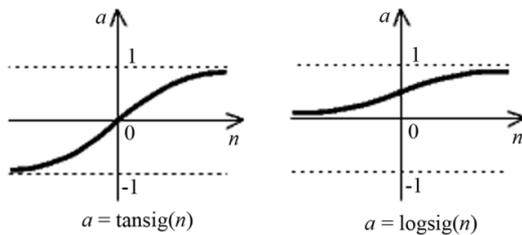


Figure 4: Tansig and logsig transfer functions.

In order to prevent the MLP from overtraining, the early stopping approach is often employed, where the available data is divided into three subsets as the training, the validation and the test set. To update the network weights and biases, the training set is used. During the training, the validation set is used to guarantee the generalization capability of the model; therefore, training should stop before the error on the validation set begins to rise. Finally, the test set is used to check the generalization ability of the MLP.

## 2.4 Support Vector Machines (SVMs)

In this sub-section, a basic overview of SVMs are given based on (Drucker et al., 1997a; Haykin, 2009; Mountrakis et al., 2011; Richards and Jia, 2006; Smola and Schölkopf, 2004; Vapnik, 1995).

When the most basic case is considered, SVM is a linear binary classifier. If a test sample is given subsequent to training, it assigns a class from one of the two possible labels. For a

classification in RS, the test data to be labeled is generally an individual pixel with a set of numerical measurements (i.e., reflectance, radiance or raw DN values) for each band of the multi- or hyper-spectral image.

Let us assume that a set of input vectors for the training data is given by  $\mathbf{x}_i$  with a number of relevant features (i.e., predictors). Each vector is associated with the corresponding label (i.e., class) denoted by  $y_i$  such that  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_i, y_i)$ , where  $\mathbf{x} \in \mathbf{R}^N$  is an  $N$ -dimensional space and  $i$  is the number of available samples. For the simplest binary case,  $y_i$  can take only two possible labels, +1 and -1. The ultimate aim of the learning task in binary SVM is to decide such an hyperplane that (1) the data points on one side would be labeled as  $y_i = +1$  and those on the other side as  $y_i = -1$ , and (2) the distance of the hyperplane from the labeled points of two classes would be maximized. The position of the optimal separation hyperplane with maximum margin that minimizes misclassifications is defined by two hyperplanes (i.e., support vectors) parallel to the separating hyperplane bordering the nearest training points from the two classes as illustrated in Figure 5. Two hyperplanes parallel to the optimal hyperplane are expressed as:

$$\begin{aligned} \mathbf{w} \cdot \mathbf{x}_i + b &\geq +1 & \text{for } y_i = +1, \\ \mathbf{w} \cdot \mathbf{x}_i + b &\geq -1 & \text{for } y_i = -1, \end{aligned} \quad (9)$$

$(i = 1, 2, \dots, k)$ ,

where “ $\cdot$ ” is dot or scalar product,  $b$  denotes the offset of the hyperplane from the origin,  $\mathbf{x}_i$  is the vector of predictors, and  $\mathbf{w} = (w_1, \dots, w_N)^T$  is the weight vector of  $N$  elements that determines the direction of the optimal separating hyperplane.

Two inequalities in Eqn. (9) can be written in the following single inequality:

$$y_i [\mathbf{w} \cdot \mathbf{x}_i + b] \geq 1 \quad (i=1, 2, \dots, k). \quad (10)$$

The optimal separating hyperplane with maximum margin can be obtained by minimizing the norm of  $\mathbf{w}$  as shown in the following equation with the inequality constrained given in Eqn. (10):

$$F(\mathbf{w}) = \frac{1}{2} (\mathbf{w} \cdot \mathbf{w}). \quad (11)$$

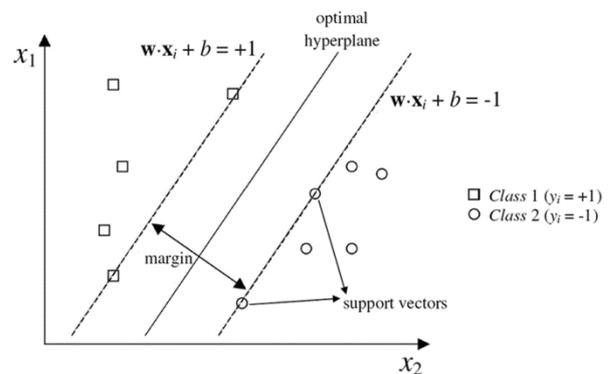


Figure 5: Optimal separating hyperplane and support vectors (Richards and Jia, 2006).

This is a constrained optimization problem in which the objective function in Eqn. (11) is minimized subject to the constrained in Eqn. (10) by using Lagrange multipliers. After some rearrangements and substitutions, the decision rule that defines the optimal separating hyperplane is given as:

$$g(\mathbf{x}) = \text{sign} \left( \sum_i^r y_i \alpha_i^0 K(\mathbf{x}, \mathbf{x}_i) - b^0 \right), \quad (12)$$

where for each of  $r$  training cases there is a vector  $\mathbf{x}_i$  that represent the response of the case together with a definition of class membership  $y_i$ ,  $\alpha_i \geq 0$ , ( $i=1,2,\dots,r$ ) are Lagrange multipliers, and finally,  $K(\mathbf{x}, \mathbf{x}_i)$  is the kernel function which enables the data points to spread in such a way that a linear hyper plane can be fitted.

Linear SVM implementation assumes that the feature data is linearly separable which is a hardly met condition in multi- or hyper-spectral RS images. Additionally, spatial and spectral overlapping of classes in RS images make it difficult for linear decision boundaries to separate classes with high accuracy. Soft margin and kernel methods have been introduced to overcome these issues. Radial basis function, linear, quadratic and polynomial kernels are frequently used to transform the raw data into a high dimensional feature space (Mountrakis et al., 2011; Pal and Mather, 2005b).

SVMs have also exhibited quite satisfying performance in regression even though they were initially developed for pattern recognition tasks. In SVM regression, the main goal is to find a function  $f(\mathbf{x})$  that has at most  $\varepsilon$  deviation from the actually obtained targets  $y_i$  for all the training data, and at the same time is as flat as possible. Errors less than  $\varepsilon$  are considered as negligible, yet any deviation larger than this is not acceptable. Then,  $f(\mathbf{x})$  takes the following form:

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x}_i + b \text{ with } \mathbf{w} \in \mathbf{X}, b \in \mathbf{R}, \quad (13)$$

where  $\mathbf{X}$  denotes the space of input patterns. According to Eqn. (13), “flatness” can be expressed in terms of small  $\mathbf{w}$ , and one way to ensure this is to minimize  $\|\mathbf{w}\|^2 = (\mathbf{w} \cdot \mathbf{w})$ , which can be re-written as a convex optimization problem:

$$\begin{aligned} &\text{minimize } \frac{1}{2} \|\mathbf{w}\|^2, \\ &\text{subject to } \begin{cases} y_i - \mathbf{w} \cdot \mathbf{x}_i - b \leq \varepsilon, \\ \mathbf{w} \cdot \mathbf{x}_i + b - y_i \leq \varepsilon, \end{cases} \quad (14) \\ &(i = 1, 2, \dots, N). \end{aligned}$$

Eqn. (14) tacitly assumes that such a function  $f(\mathbf{x})$  exists and it approximates all pairs  $(\mathbf{x}_i, y_i)$  with  $\varepsilon$  precision, i.e., the convex optimization problem is “feasible”. In the case of infeasible scenario where some errors may be allowed, slack variables  $\xi_i, \xi_i^*$  can be used to deal with otherwise infeasible constraints of the optimization problem in Eqn. (14), which is also known as “soft margin” method. Consequently, the following formulation given in Vapnik (1995) is obtained:

$$\begin{aligned} &\text{minimize } \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*), \\ &\text{subject to } \begin{cases} y_i - \mathbf{w} \cdot \mathbf{x}_i - b \leq \varepsilon + \xi_i, \\ \mathbf{w} \cdot \mathbf{x}_i + b - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0. \end{cases} \quad (15) \end{aligned}$$

The constant  $C > 0$  controls the trade-off between the flatness of  $f(\mathbf{x})$  and the amount up to which deviations larger than  $\varepsilon$  are tolerated. This leads to a so called  $\varepsilon$ -insensitive loss function  $|\xi|_\varepsilon$  which is graphically illustrated in Figure 6 and can be expressed as:

$$|\xi|_\varepsilon := \begin{cases} 0, & \text{if } |\xi| \leq \varepsilon, \\ |\xi| - \varepsilon, & \text{otherwise.} \end{cases} \quad (16)$$

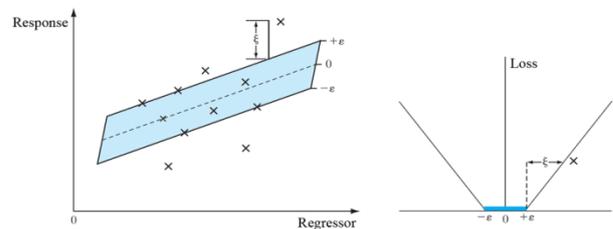


Figure 6: A basic illustration of the soft margin loss setting for a linear SVM (Haykin, 2009).

## 2.5 Experimental Design

After the removal of water, cloud cover and bad-quality pixels from the MODIS images, 8872 and 27338 pixels are available in the training and testing data sets, respectively. In order to train ANN and SVM regression models, 3% of 8872 (i.e., 254) pixels in the training data set is selected by stratified random sampling (cf. Table 1). Stratification is carried out with respect to snow cover fraction from 0.0 to 1.0 with 0.1 intervals in order to prevent ANN and SVM regression models from being biased towards a certain snow cover fraction. Surface reflectance values of MODIS bands 1-7, NDSI and NDVI are used as predictors, and associated FSC values derived from higher spatial resolution ETM+ images are used as response variable.

The chosen neural network structure for FSC mapping is a feed-forward network with one hidden layer trained via backpropagation learning rule with 9 nodes in the input layer and 1 node in the output layer. Since there is no unique theory to determine the optimal values of an ANN's internal variables (Moosavi et al., 2014), the number of nodes in the input and output layers are set equal to the number of predictor (i.e., input) and response (i.e., output or target) variables. The gradient-based Levenberg-Marquardt backpropagation is used during ANN training. The tansig function is employed in the hidden layer. For the output layer, logsig and tansig functions are alternatively assigned in order to investigate their effect on the estimated FSC values.

Various methods have been proposed to determine the optimum number of neurons in the hidden layer such as  $2n + 1$ ,  $2n$  and  $n$ , where  $n$  is the number of nodes in the input layer (Moosavi et al., 2014); however, trial-and-error approach is an appropriate way as indicated by Mishra and Desai (2006), and Shirmohammadi et al. (2013). Therefore, 4-22 nodes with

increment of 3 are tested for the hidden layer. The ANN training data is split into three subsets by random sampling: 70% for training, 15% for validation, and 15% for testing.

A successful SVM application requires the optimal selection of two SVM key parameters; kernel function and regularization parameter  $C$ . However, there is no unique and universally accepted heuristics for the selection of these parameters which often leads to a basic trial-and-error approach (Mountrakis et al., 2011; Yuan and Lee, 2015).

Kernel functions employed in SVM applications can be grouped under four main categories; linear, polynomial, radial basis function and sigmoid kernels. In RS literature, radial basis function (RBF) (i.e.,  $e^{-\gamma\|x-x_i\|^2}$ ) and polynomial kernel (i.e.,  $((\mathbf{x} \cdot \mathbf{y}) + 1)^d$ ) are frequently preferred (Huang et al., 2002; Pal and Mather, 2005a); therefore, these two kernels are used in the study. For RBF kernel, the width of kernel  $\gamma$  and  $C$  should be defined. On the other hand, the degree of polynomial  $d$  and again  $C$  should be determined for polynomial kernel. In order to select optimal values for these parameters, grid search method with cross validation is used as proposed by Kavzoglu and Colkesen (2009). In this approach, a coarser grid with exponentially growing sequence of  $(C, \gamma)$  and  $(C, d)$  is applied, and once the optimal region on the grid is identified, a finer grid search is performed. Finally, the pair of parameters that gives the highest cross validation accuracy is used for final training process. For RBF kernel  $C$  takes values in  $\{2^1, 2^2, \dots, 2^{15}\}$ , whereas  $\gamma$  is varied from  $2^{-5}$  to  $2^3$ . For polynomial kernel, the degree  $d$  takes values from the set  $\{2, 3, 4\}$ .

The performances of ANN and SVM regression models during both training and testing stages are assessed by root mean square error (RMSE) and correlation coefficient ( $R$ ) values:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}}, \quad (17)$$

$$R = \frac{N \sum_{i=1}^N y_i \hat{y}_i - \sum_{i=1}^N y_i \sum_{i=1}^N \hat{y}_i}{\sqrt{N \sum_{i=1}^N y_i^2 - \left(\sum_{i=1}^N y_i\right)^2} \sqrt{N \sum_{i=1}^N \hat{y}_i^2 - \left(\sum_{i=1}^N \hat{y}_i\right)^2}}, \quad (18)$$

where  $N$  is the total number of observations,  $y_i$  is the  $i$ th reference value, and  $\hat{y}_i$  is the  $i$ th predicted value.

### 3. RESULTS

Both ANN and SVM regression models give high accuracy in FSC mapping on the training data set. The high  $R$  ( $\geq 0.91$ ) and low RMSE ( $\leq 0.4144$ ) values not only show good agreement with reference Landsat-derived FSC data, but also indicate correct adjustment of model parameters during the training process (cf. Table 3).

	Training		Testing	
	$R$	RMSE	$R$	RMSE
MLP-logsig	0.91	0.4144	0.84	0.4481
MLP-tansig	0.98	0.0772	0.93	0.1505
SVM-RBF	0.95	0.1167	0.92	0.1638

SVM-Poly <sup>2</sup>	0.94	0.1741	0.82	0.2012
SVM-Poly <sup>3</sup>	0.94	0.1996	0.85	0.2279
SVM-Poly <sup>4</sup>	0.92	0.2971	0.78	0.3204

Table 2:  $R$  and RMSE values of ANN and SVM models on training and test data.

Table 4 shows RMSE values obtained during the training of ANN models with logsig and tansig transfer functions in the output layer for different number of neurons in the hidden layer. The optimal number of neurons in the hidden layer that gives the lowest RMSE values are 10 and 7 for MLP-logsig and MLP-tansig networks, respectively. During the training of the MLP networks, MLP-tansig gives the best training performance with  $R = 0.98$  and  $RMSE = 0.0772$ , whereas  $R$  and RMSE of MLP-logsig are 0.91 and 0.4144, respectively.

Network type	Number of neurons in the hidden layer						
	4	7	10	13	16	19	22
MLP logsig	0.4391	0.4663	<b>0.4144</b>	0.4267	0.4266	0.4647	0.4284
MLP tansig	0.0823	<b>0.0772</b>	0.0850	0.0839	0.0868	0.0861	0.0920

Table 4. Results of MLP networks with two different transfer functions in the output layer with respect to different number of neurons in the hidden layer during training. The boldface figures indicate the minimum RMSE values obtained on the test subset of the training data.

For each kernel function, different SVM regression models are trained with various settings as explained in the previous subsection. Next, the performance of each model is evaluated on the training data by using  $R$ . It can be seen from Table 5, all of the kernel functions gives quite satisfactory results on the training data set; however, RBF kernel with  $C = 270$  and  $\gamma = 3$  provides the highest  $R$ . The second best performance belongs to both 2nd and 3rd order polynomial kernel functions with the same  $R$  value of 0.94, and with  $C$  values of 245 and 280, respectively. SVM with 4th order polynomial kernel exhibits relatively poor performance on the training data with  $R = 0.92$  and  $C = 310$ .

Kernel Function	$C$ parameter	$\gamma$	$R$
RBF	270	3	0.95
2nd order Polynomial	245	N/A	0.94
3rd order Polynomial	280	N/A	0.94
4th order Polynomial	310	N/A	0.92

Table 5. Optimal training settings for SVM kernel functions.

The independent test data provides a more rigorous and realistic assessment of ANN and SVM regression models since it is not included in the training. Not surprisingly, accuracy as measured by  $R$  and RMSE values are lower for the test than training. According to the results given in Table 3, MLP-tansig network gives the best performance on the test data set with  $R = 0.93$ . The MLP network with logsig transfer function in the output layer has lower accuracy as compared to the MLP-tansig network with  $R = 0.84$  and  $RMSE = 0.4481$ . However, when the scatter plot diagrams of modelled FSC vs. reference FSC values given in Figure 7 are analysed, an interesting behaviour of MLP-logsig network is revealed. It cuts off the lower half of the FSC values (i.e.,  $FSC < 0.5$ ) and only gives the upper half (i.e.,  $FSC \geq 0.5$ ). This trend is also observed during the training stage for all settings of number of neurons in the hidden layer, which shows that MLP network with logsig transfer function in the output layer is not suitable for FSC estimation.

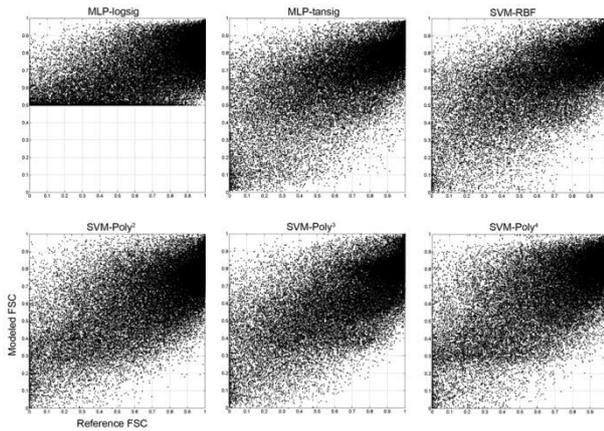


Figure 7: Landsat-derived reference FSC values vs. estimated FSC values by ANN and SVM models.

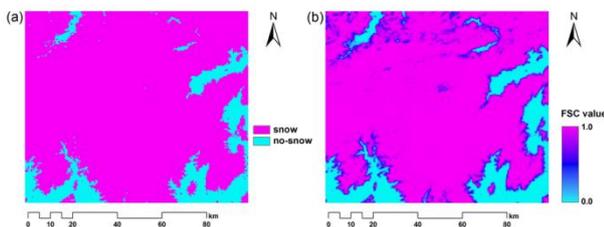


Figure 8: (a) Reference binary snow map of the test area obtained from Landsat ETM+ data, and (b) FSC map of the test area generated by the MLP-tansig model.

The performance of SVM with RBF kernel is quite comparable with the MLP-tansig network with  $R = 0.92$  and  $RMSE = 0.1638$ . SVMs with 2nd and 3rd order polynomial kernels gives similar accuracies with  $R = 0.82$  and  $R = 0.85$ , respectively; whereas, SVM with 4th order polynomial kernel has the worst overall performance with  $R = 0.78$ . The FSC image of the test area constructed by MLP-tansig model is given in Figure 8 with the reference binary snow map derived from the corresponding Landsat ETM+ data.

#### 4. CONCLUSIONS AND OUTLOOK

Continuous monitoring of snow cover is often considered as an essential task in order to observe the changes in hydrology, meteorology and climatology of earth at both local and global scales. In this study, the suitability of nonparametric ANN and SVM methods for FSC mapping is investigated. Surface reflectance values of MODIS bands 1-7, NDSI and NDVI are used as predictor variables; whereas, the reference FSC values are generated from higher resolution Landsat ETM+ data, and they are used as response variable.

The best performance is achieved by the MLP-tansig model over the independent test area. The performance of SVM-RBF model is slightly less than that of MLP-tansig. Advantage of ANNs and SVMs over traditional statistical methods lies in the fact that they make no assumptions about underlying relationship between a set of predictor variables and a response. Additionally, once the models are trained they can efficiently and quickly calculate FSC values.

The training process of SVMs depends only on two parameters; the choice of kernel function and  $C$ , therefore it can be considered as much simpler. However, the same conclusion cannot be made for ANNs, since the number of “model tuning” parameters to be optimally decided is large. Parameters such as the number and size of the hidden layers, the type of transfer function in the hidden and output layers, the range of the initial weights, the learning rate and the value of the momentum term, the number of samples employed during the training of a network have a significant impact on the predictive performance and the generalization ability of the network and should be considered during the design and implementation of an ANN. Consequently, a potential future extension of this study is to investigate the effect of these model building parameters of ANNs in a more detailed fashion by using different and larger satellite image data sets.

#### REFERENCES

- Agirre-Basurko, E., Ibarra-Berastegi, G. and Madariaga, I. (2006). Regression and multilayer perceptron-based models to forecast hourly O<sub>3</sub> and NO<sub>2</sub> levels in the Bilbao area. *Environmental Modelling & Software*, 21(4), pp. 430-446.
- Aydinozu, D., Ibret, U. and Aydin, M. (2011). *Analysis of Terrain Usage in Kastamonu-Ilgaz Mountain Natural Park*. Paper presented at the International Symposium on Environmental Protection and Planning: Geographic Information Systems (GIS) and Remote Sensing (RS) Applications (ISEPP), 28-29 June 2011, İzmir - TURKEY.
- Bruzzone, L. and Melgani, F. (2005). Robust multiple estimator systems for the analysis of biophysical parameters from remotely sensed data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(1), pp. 159-174.
- Campbell, C. and Ying, Y. (2011). *Learning with Support Vector Machines*: Morgan & Claypool Publishers.
- Czyzowska-Wisniewski, E. H., van Leeuwen, W. J. D., Hirschboeck, K. K., Marsh, S. E. and Wisniewski, W. T. (2015). Fractional snow cover estimation in complex alpine-forested environments using an artificial neural network. *Remote Sensing of Environment*, 156, pp. 403-417.
- Dai, X. and Khorrarn, S. (1999). Remotely sensed change detection based on artificial neural networks. *Photogrammetric Engineering and Remote Sensing*, 65, pp. 1187-1194.
- Dobrev, I. D. and Klein, A. G. (2011). Fractional snow cover mapping through artificial neural network analysis of MODIS surface reflectance. *Remote Sensing of Environment*, 115(12), pp. 3355-3366.
- Dongale, T., Patil, K., Vanjare, S., Chavan, A., Gaikwad, P. and Kamat, R. (2015). Modelling of nanostructured memristor device characteristics using Artificial Neural Network (ANN). *Journal of Computational Science*, 11, pp. 82-90.
- Drucker, H., Burges, C. J., Kaufman, L., Smola, A. and Vapnik, V. (1997a). Support vector regression machines. *Advances in Neural Information Processing Systems*, 9, pp. 155-161.
- Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. and Vapnik, V. (1997b). Support Vector Regression Machines. In M. C. Mozer, M. I. Jordan and T. Petsche (Eds.), *Advances in*

- Neural Information Processing Systems 9* (Vol. 9, pp. 155-161). USA: The MIT Press.
- Fisher, W. D., Camp, T. K. and Krzhizhanovskaya, V. V. Anomaly detection in earth dam and levee passive seismic data using support vector machines and automatic feature selection. *Journal of Computational Science*. doi: <http://dx.doi.org/10.1016/j.jocs.2016.11.016>.
- Foody, G. M. and Mathur, A. (2004). A Relative Evaluation of Multiclass Image Classification by Support Vector Machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(6), pp. 1335-1343.
- Gardner, M. W. and Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron) - A review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14–15), pp. 2627-2636.
- Hall, D. K., Riggs, G. A., Salomonson, V. V., DiGirolamo, N. E. and Bayr, K. J. (2002). MODIS snow-cover products. *Remote Sensing of Environment*, 83, pp. 181-194.
- Haykin, S. (2009). *Neural Networks and Learning Machines* (3rd ed.). New Jersey: Prentice Hall, Pearson Education.
- Huang, C., Davis, L. and Townshend, J. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4), pp. 725-749.
- Kaheil, Y. H., Rosero, E., Gill, M. K., McKee, M. and Bastidas, L. A. (2008). Downscaling and forecasting of evapotranspiration using a synthetic model of wavelets and support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 46(9), pp. 2692-2707.
- Kar, A. K. (2015). A hybrid group decision support system for supplier selection using analytic hierarchy process, fuzzy set theory and neural network. *Journal of Computational Science*, 6, pp. 23-33.
- Kavzoglu, T. and Colkesen, I. (2009). A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, 11(5), pp. 352-359.
- Kavzoglu, T. and Mather, P. M. (2003). The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing*, 24(23), pp. 4907-4938.
- Klein, A. G., Hall, D. K. and Riggs, G. A. (1998). Improving snow cover mapping in forests through the use of a canopy reflectance model. *Hydrological Processes*, 12(10), pp. 1723-1744.
- Kumar, D., Meghwani, S. S. and Thakur, M. (2016). Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets. *Journal of Computational Science*, 17, Part 1, pp. 1-13.
- Kuter, N. (2008). Evaluation of Ilgaz Mountain National Park in Terms of Forest Landscape and Aesthetics. *Turkish Journal of Forestry*, 1, pp. 36-47.
- Kuter, N. and Kuter, S. (2016, 23-26 May, 2016). *Binary Snow Mapping in Kurşunlu Forest Sub-District by Using Support Vector Machines*. Paper presented at the 4th International Geography Symposium - GEOMED 2016, Antalya, Turkey.
- Lee, J., Weger, R. C., Sengupta, S. K. and Welch, R. M. (1990). A neural network approach to cloud classification. *IEEE Transactions on Geoscience and Remote Sensing*, 28(5), pp. 846-855.
- Mas, J. and Flores, J. (2008). The application of artificial neural networks to the analysis of remotely sensed data. *International Journal of Remote Sensing*, 29(3), pp. 617-663.
- Mattera, D. and Haykin, S. (1999). Support vector machines for dynamic reconstruction of a chaotic system. In S. Bernhard, J. C. B. Christopher and J. S. Alexander (Eds.), *Advances in Kernel Methods - Support Vector Learning* (pp. 211-241). Cambridge: MIT Press.
- Mishra, A. and Desai, V. (2006). Drought forecasting using feed-forward recursive neural network. *Ecological Modelling*, 198(1), pp. 127-138.
- Moosavi, V., Malekinezhad, H. and Shirmohammadi, B. (2014). Fractional snow cover mapping from MODIS data using wavelet-artificial intelligence hybrid models. *Journal of Hydrology*, 511, pp. 160-170.
- Mountrakis, G., Im, J. and Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), pp. 247-259.
- Müller, K.-R., Smola, A. J., Rätsch, G., Schölkopf, B., Kohlmorgen, J. and Vapnik, V. (1997). Predicting time series with support vector machines. In W. Gerstner, A. Germond, M. Hasler and J.-D. Nicoud (Eds.), *Artificial Neural Networks — ICANN'97: 7th International Conference Lausanne, Switzerland, October 8–10, 1997 Proceedings* (pp. 999-1004). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Mustaffa, Z., Yusof, Y. and Kamaruddin, S. S. (2014). Enhanced artificial bee colony for training least squares support vector machines in commodity price forecasting. *Journal of Computational Science*, 5(2), pp. 196-205.
- Painter, T. H., Dozier, J., Roberts, D. A., Davis, R. E. and Green, R. O. (2003). Retrieval of subpixel snow-covered area and grain size from imaging spectrometer data. *Remote Sensing of Environment*, 85(1), pp. 64-77.
- Pal, M. and Mather, P. (2005a). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, 26(5), pp. 1007-1011.
- Pal, M. and Mather, P. M. (2005b). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, 26(5), pp. 1007-1011.
- Paola, J. D. and Schowengerdt, R. A. (1995). A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. *International Journal of Remote Sensing*, 16(16), pp. 3033-3058.

Rafiq, M. Y., Bugmann, G. and Easterbrook, D. J. (2001). Neural network design for engineering applications. *Computers & Structures*, 79(17), pp. 1541-1552.

Raji, C. and Chandra, S. V. (2016). Graft survival prediction in liver transplantation using artificial neural network models. *Journal of Computational Science*, 16, pp. 72-78.

Richards, J. A. and Jia, X. (2006). *Remote sensing digital image analysis: An introduction* (4th ed.). Germany: Springer.

Salomonson, V. V. and Appel, I. (2004). Estimating fractional snow cover from MODIS using the normalized difference snow index. *Remote Sensing of Environment*, 89, pp. 351-360.

Salomonson, V. V., Barnes, W. and Masuoka, E. J. (2006). Introduction to MODIS and an Overview of Associated Activities. In J. J. Qu, W. Gao, M. Kafatos, R. E. Murphy and V. V. Salomonson (Eds.), *Earth Science Satellite Remote Sensing. Volume 1: Science and Instruments* (pp. 12-32). Beijing: Springer.

Shirmohammadi, B., Vafakhah, M., Moosavi, V. and Moghaddamnia, A. (2013). Application of several data-driven techniques for predicting groundwater level. *Water Resources Management*, 27(2), pp. 419-432.

Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), pp. 199-222.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*: Springer Heidelberg.

Verbeiren, S., Eerens, H., Piccard, I., Bauwens, I. and Van Orshoven, J. (2008). Sub-pixel classification of SPOT-VEGETATION time series for the assessment of regional crop areas in Belgium. *International Journal of Applied Earth Observation and Geoinformation*, 10(4), pp. 486-497.

Yuan, F.-C. and Lee, C.-H. (2015). Using least square support vector regression with genetic algorithm to forecast beta systematic risk. *Journal of Computational Science*, 11, pp. 26-33.

Zheng, S., Shi, W.-z., Liu, J. and Tian, J. (2008). Remote sensing image fusion using multiscale mapped LS-SVM. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5), pp. 1313-1322.