EasyChair Preprint

№ 10353

Multistream Convolutional Neural Network Fusion for Pixel-wise Classification of Peatland

Fahimeh Farahnakian, Luca Zelioli, Timo Pitkänen,
Jonne Pohjankukka, Maarit Middleton, Sakari Tuominen,
Paavo Nevalainen and Jukka Heikkonen

June 7, 2023

Multistream Convolutional Neural Network Fusion for Pixel-wise Classification of Peatland

Fahimeh Farahnakian[1], Luca Zelioli[1], Timo Pitkänen[2], Jonne Pohjankukka[1],
Maarit Middleton[3], Sakari Tuominen[2], Paavo Nevalainen[1] and Jukka Heikkonen[1]
[1]Department of Computing, University of Turku, Finland
[2]Natural Resources Institute Finland (Luke)
[3] Geological Survey of Finland (GTK)
Email: {fahfar, luzeli, jjepoh, ptneva, jukhei}@utu.fi, {timo.p.pitkanen,sakari.tuominen}@luke.fi, {maarit.middleton}@gtk.fi

***Abstract*—Recently, Convolutional Neural Network (CNN) has shown higher performance than other machine learning methods for land classification. In this paper, we propose a CNN fusion architecture for peatland site type classification by combining multisource and multiresolution data. The data is acquired by optical and radar satellite remote sensing, airborne laser scanning data and multi-source forest inventory GIS datasets. Based on our data, we are dealing with the high-dimensional class-imbalanced dataset for solving pixel-wise classification of peatlands. To reduce the data dimension and find an optimal subset of inputs, we first applied the sequential feature selection method. Then, we proposed a window-based pixel classification approach based on the selected inputs. This approach can extract the spatial information around each training sample in a defined window region and produce a pixel-wise classfication map. Experiments are carried out for ecological classification of peatlands in Finland.**

***Index Terms*—convolutional neural network, land cover classification, semantic segmentation, remote sensing**

I. Introduction

Peatlands cover half of the Earth's wetlands and cover 3% of the global total land area [1]. They are very important for global climate as they can store enormous amount of carbon. Drainage for forestry, agriculture and peat extraction cause emissions of the greenhouse gasses and have turned those peatlands as net sources that are reported under the land use, land use change and forestry sector in national inventory [2]. In addition, peatlands act as reservoirs for floodwater retention, provision of wildlife habitats, and control of soil erosion [1], [3]. To avoid further loss of peatland and maintain existing peatland ecosystems, inventorying and monitoring of peatlands is important. Peatland classification can be considered as a sub-problem of land cover classification. Land cover classification is a fundamental challenge in remote sensing which aims at automatically labelling a pixel in a raster dataset to a specific category such as vegetation, soil and water.

Despite of significant improvements in the availability tools and accuracy of remote sensing data (i.e., smaller pixel size), land cover classification of complex and heterogeneous areas such as peatlands is still challenging. This is because, in highly fragmented landscapes, many peatland classes are not clearly defined but they form gradients transforming to other peatland classes which, in turn, increases the within-class variability and decreases between class separability. In addition, human activities such as drainage for forestry have profoundly changed the species structure. Thus, peatland classes have similar characteristics in commonly used optical [4], microwave [5] and ALS [6] remote sensing datasets. Fusing remote sensing data and possibly other spatial datasets may contribute to improved recognition of these complex land cover classes [7].

In recent years, Convolutional Neural Network (CNN), as one of the most well-known deep learning models has gained interest in remote sensing applications [8]–[12]. The main reason is that CNN can process the data in the form of multiple layers which is applicable to process the multi-band remote sensing data. In remote sensing data, highly complex structures and spatial patterns are generalized into raster cells. CNNs can extract them to represent global and local information. Moreover, CNN is widely-used for sensor fusion since it can explore the features from the data efficiently [11], [13]. In [14], they proposed a fully-convolutional neural network for land cover segmentation by fusing RGB with additional Synthetic Aperture Radar (SAR) and near infrared images. This work and other recent works [11], [13], [15] show that sensor fusion can improve the performance of computer vision tasks such as detection, classification and segmentation.

In this paper, we present a CNN-based fusion architecture to carry out peatland classification using multi-sensor satellite including optical and SAR images, Airborne Laser Scanning (ALS) and Multi-Source National Forest Inventory (MS-NFI). The architecture combines the multi-source data at the decision level for pixel-wise classification of peatland. Due to the large number and dimension of inputs, we proposed two techniques. (1) CNN is trained on a limited amount of field observations by proposing a window. This idea is similar the sliding window approach which is common in pixel-wise classification [16]. (2) We selected a subset of inputs that is the most relevant to the problem and reduce the number of inputs for carry out the final pixel-to-pixel segmentation. We demonstrate the efficiency of our architecture on a real dataset in Finland. The data was collected in the project called Advances in soil information-MaaTi [17]. The MaaTi project was established to develop a set of methods that would allow to produce country-wide data and information on soils quickly and efficiently utilizing the soil information in the databases of organizations.

II. RelatedWork

Deep learning techniques such as convolutional neural networks (CNNs) has made huge progress in remote sensing applications including scene classification [10], object detection [11], land cover classification [8], [18], image fusion [11], [13] and segmentation [12]. One of the main technologies that improve the performance of CNN-based segmentation methods is sensor fusion [11]. However, the multi-sensor fusion methods suffer from two main challenges: (1) the feature extraction from various types of sensory data, and (2) the selection of a suitable fusion level. Multi^{3}Net [19] is a CNN model which fuses Sentinel-1, Sentinel-2, optical and radar imagery. It uses CNN for segmentation of flooded buildings which consists of multiple deep encoder-decoder streams, each of which produces an output map based on data from a single sensor. This fusion model improved the detection accuracy compared to single sensor approach. In another work [18], a multilevel CNN architecture is targeted to land cover and crop type classification from multi-temporal multi-source satellite imagery. Their architecture outperforms the Multi-Layer Perceptron (MLP) allowing better discrimination of certain summer crop types.

Information on peatland classes is required for making decisions in forestry and environmental management and in land-use planning [20]. Because no individual wavelength range can capture the complexity of the 3D structure of peatland vegetation combining data from different sensors is required for peatland site type classification. In [20], the Random Forest (RF) algorithm was used for tropical peatland classification using multisensor satellite imagery in Greater Amanzule, Ghana. Another work conducted in [21] classifies 13 peatland vegetation types using supervised algorithms such RF, Support Vector Machine (SVM) and Logistic Regression (LR). These traditional classification algorithms have advantages and disadvantages in terms of accuracy, data type used and the ease of analysis. In [7], a pretrained AlexNet [22] is used for weatland mapping. Their results show that CNN was superior to RF for complex wetland mapping. However, none of the above mentioned used all of the input datasets tested in this paper. Based on our knowledge, there is no published research that uses CNN for peatland classification based on variety of multi-sensors and multiresolution data such as this study.

III. Study Area and Data

The Keminmaa study area belongs to the Southern *aapa* mire zone [23] which has a wide range of peatland site types from nutrient-rich to nutrient-poor types (Fig. 1). The *aapa* mires are classified as a minerotrophic vegetation–ecological peatland complex with lawn level vegetation and concave surface topography, but without clear surface patterning. Fifty percent (42800 hectares) of the land area is covered by peatlands which have a mean depth of 1.1 m [24]. Annotation data was extracted from the geological survey of Finland peatland inventory [24] and Luke National Forest Inventory (NFI) datasets [25] which contain information on peatland

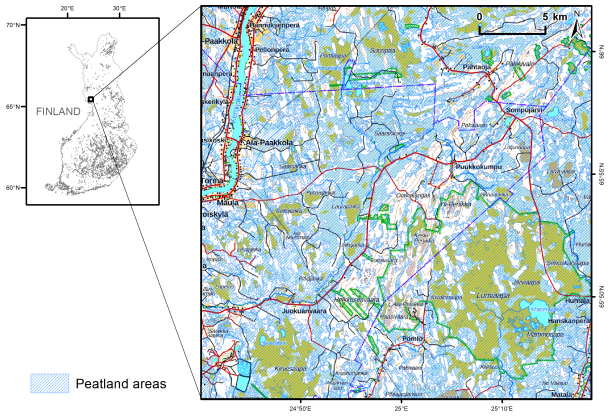

Fig. 1. Keminmaa study area. Background maps: National Land Survey of Finland general map (left) and topographic map (right), 2022.

site type observations conducted in the field. Peatland site type classes were determined according to [23], and only observations made after the year 2000 were selected from the databases to reflect the most current situation.

Table I shows the characteristics of the data used in this paper. The spatial resolution column in the table represents the resolution which is used in the paper. Totally, we used four data types as follows:

1) Synthetic Aperture Radar (SAR) satellite data: Initial Sentinel-1 (S1) data consisted of two different image types, which were downloaded from the Copernicus Open Access Hub[1] and further processed using The ESA Sentinel Applications Platform (SNAP) software.

These were intensity images (i.e., ground range detected products) and coherence images (based on the single look complex products), including both VV and VH polarizations. Intensity refers basically to backscattered signals of a single image which have been detected, multi-looked and projected, whereas coherence measures the similarity of radar reflections between two individual images. Five intensity images were included in the analysis, one per each summer month of 2017 (May-September). Regarding to coherences, three different sets of images were produced. First, coherences were calculated for each temporally consecutive pair of Sentinel-1A images at a 12-day interval between May 9 and September 30, 2017. Second, coherences were calculated for the same set of Sentinel-1A images but so that the midmost image of July 20, 2017, was always included as a "reference" image in the pair. And third, VV and VH values of multiple coherence pairs, calculated for summer 2019, were averaged as single layers in order to reduce noise. The different year of the averaged dataset resulted from its later calculation and use in other modelling purposes.

One mosaic was compiled of the TerraSAR-X SC images (DLR 2013) acquired on May 30th and June 4-5th, 2017. A seamless mosaic of the preprocessed images was made with the SAR-Mosaic processing tool. Single RADARSAT-2 SGF

[1]https://scihub.copernicus.eu/

TABLE I
DESCRIPTION OF INPUT DATA SOURCES USED IN THIS PAPER.

Data	Description	Spatial resolution(m)
S1	Microwave band C (5.5 cm)/VV, VH	10
RADARSAT-2	Microwave band C (5.6 cm/HH, HV)	10
TerraSAR-X	Microwave band X (3.1 cm/HH)	10
S2	Optical bands: B2 (490 nm), B3 (560 nm), B4 (665 nm), B5 (705 nm), B6 (740 nm), B7 (783 nm), B8 (842 nm), B8A (865 nm), B11 (1610nm), B12 (2190 nm)	10
DEM	Terrain elevation above sea level	2
CHM	Vegetation height above ground level	1
MS-NFI	Thematic map layers of forest parameters	16

images were acquired on five dates; 24th May, 17th June, 11th July, 4th August and 21 September, 2017, to test the intra-seasonal variation in separating the peatland site type. For both datasets radiometric calibration and converting output sigma0 -channels to db was performed.

2) Optical satellite: Sentinel-2 (S2) data, similarly to Sentinel-1, were downloaded from the Copernicus Open Access Hub. One image per each summer month (May-September) was selected between years 2018-2020, depending on the visually assessed image quality and lack of clouds or any haze. The study area could fit within a single Sentinel-2 frame, so mosaicing of multiple images was not needed. All the bands at an initial resolution of 10/20 m were used as input data, resampled to 10 m pixel size.

3) Airborne Laser Scanning (ALS): The open source ALS data acquired with Light Detection And Ranging (LiDAR) was provided by the National Land Survey of Finland (NLS) from approximate flight altitudes of 1800–2500 m in serious of flight campaigns conducted in 2008 – 2019 (MML 2022).

The use of ALS data in predicting peatland site types is based on using a Digital Elevation Model (DEM) of terrain [26] with the 2m grid size, and Canopy Height Model (CHM) of vegetation with 1 m resolution. The ALS points (echoes) have been classified into terrain and vegetation points, which were the basis for DEM and CHM, respectively. The ALS-based DEM and CHM were the basis for several derived features, including: local binary patterns (LBP) [27] and the steepest descent s over 6 m distance. These derivatives were used in the study to investigate the added utility of textural input features in peatland classification. Basic idea with these derived surface height features is that certain environment types contain characteristical textural properties, which can be captured or clustered properly by a suitable set of derived features. Even though the CNN method itself should be able to learn also textural characteristics itself, the derived features were also of interest for testing.

LBP has two main parameters: 1) $lbp_r = 4.8$ m, the radius of the pixel neighborhood, and 2) $lbp_{npix} = 8 \times lbp_r$, which is the number of neighboring pixels at distance lbp_r to be used for generating the binary code. The window size used was 16×16 m, which was assumed to capture most of the local peatlands properties. All these DEM derived features were later scaled to the interval $[-1, 1]$.

4) Multi-Source National Forest Inventory (MS-NFI): MS-NFI of Finland produces information of forests in the form of thematic maps of forest variables [28]–[30]. The MS-NFI methodology employs satellite images, digital maps and national forest inventory (NFI) sample plot measurements as input data sources. The MS-NFI map data comprises thematic raster-form map layers of approx. 40 forest variables. The current resolution of the MS-NFI map grid is 16 m. The variables measured from NFI sample plots are used as the field reference data, and the field variables are predicted for a uniform grid of pixels using satellite image features, coarse resolution maps of tree species proportions (based on NFI field data) and digital map data based on topographic map database of Finland. Digital map data are used for extracting forestry land from other land cover classes, and for separating mineral lands from organic (peat-) lands within forestry land. This results in two distinctive silvicultural land strata for which forest characteristics are predicted separately using reference NFI plots of the respective stratum. Paludified forests with peat layer of less than 30 cm are included in the mineral land stratum. MS-NFI forms a long time series of data. Two time series were extracted from the consecutive MS-NFI data sets, 2011-2019 and 2000-2009, which were then used as input layers for peatland site predictions.

A. Dataset Description

We used two different versions of data: V1 and V2. Table II summarizes the characteristics and all differences between V1 and V2. The size of the dataset represents the number of labeled samples. Each sample belongs to a peatland site type. Table III summarizes the initial peatland site type used in this paper. The number of inputs represents whole input sources which are extracted from optical, SAR, ALS and MS-NFI data.

In the V1 dataset, all data were combined to produce a model of the pristine site types and peatly heatland site types, the latter ones being a product of artificial drainage for forestry. In addition, two landuse classes namely agricultural field parcels on organic soils and abandoned agricultural lands were included in the classification scheme.

In the V2 dataset, the site type information remained the same as in V1 but it was divided into 'drained' and 'undrained' datasets in order to diminish misclassification between the pristine areas and the once artificially drained for forestry.

Totally, we have a few number of labeled data and an imbalanced organization of class instances as it is illustrated in Fig. 2. For instance, there is one sample for some classes such as "EHWSF" and "CSPF". The maximum number of sample belongs to the class "HRTHP I" which is 335 and 311 for V1 and V2 drained. In addition, the number of inputs are high as we fuse different multisource and multiresoulation data. For this reason, we applied an input selection method to reduce the input dimension and select most relevant inputs to our classification problem.

TABLE II
CHARACTERISTICS OF THE DATASETS.

Name of dataset	No. of classes	No. of inputs	Size of dataset
Keminmaa V1	39	171	2065
Keminmaa V2 (drained)	37	162	1359
Keminmaa V2 (undrained)	37	162	706

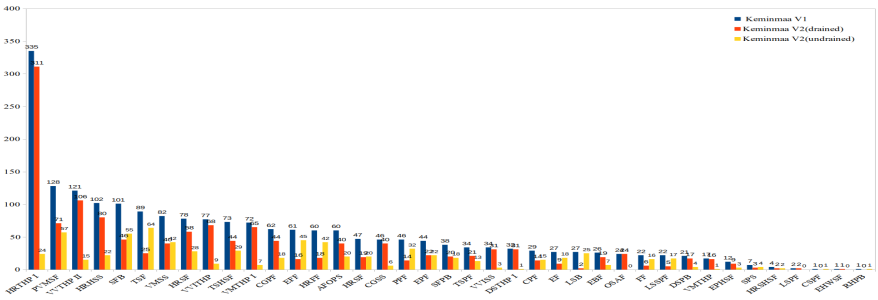

Fig. 2. Distribution of the peatland classes.

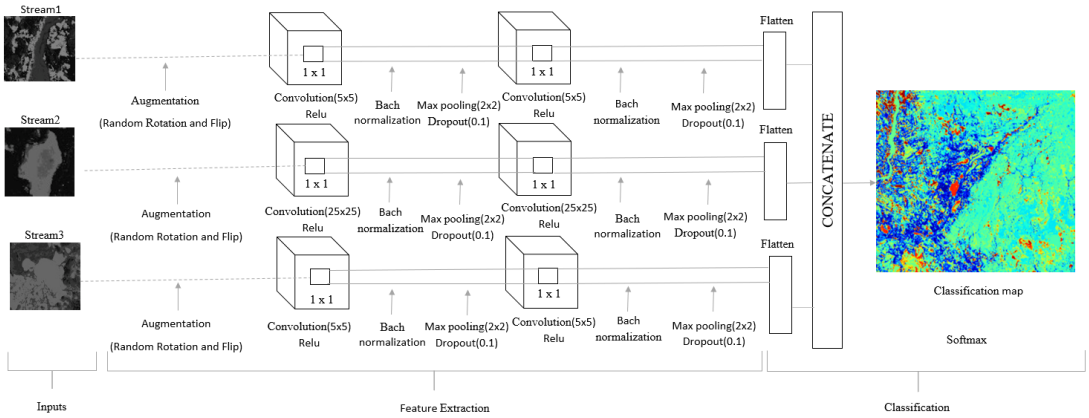

Fig. 3. The proposed multistream CNN fusion architecture for pixel-wise classification of peatland.

TABLE III
PEATLAND CLASSES AND THEIR ABBREVIATIONS.

Eutropthic paludified hardwood-spruce forest	EPHSF
Herb-rich hardwood-spruce swamp	HRHSS
Abandoned field on peat soils	AFOPS
Organic soil agricultural fields	OSAF
Herb-rich sedge hardwood-spruce fen	HRSHSF
Herb-rich type heathy peatland I	HRTHP I
Eutrophic hard-wood-spruce fen	EHWSF
Eutrophic birch fen	EBF
Paludified Vacc. myrtillus spruce forest	PVMSF
Vacc. myrtillus spruce swamp	VMSS
Vacc. myrtillus type heathy peatland I	VMTHP I
Herb-rich sedge fen	HRSF
Eutrophic fen	EF
Eutrophic flark fen	EFF
Herb-ricj flark fen	HRFF
Herb-rich sedge brich-pin fen	HRSBPF
Eutrophic pine fen	EPF
Tall-sedge hardwood-spruce fen	TSHSF
Vacc. myrtillus type heathy peatland II	VMTHP
Vacc. vitis-idaea spruce swamp	VVISS
Spruce-pine swamp	SPS
Paludified pine forest	PPF
C. globularis pine fen	CGPF
C. globularis spruce swamp	CGSS
Vacc. vitis-idaea type heathy peatland I	VVITHP
Flark fen	FF
Cottongrass-sedge pine fen	CSPF
Tall-sedge fen	TSF
Vacc. vitis-idaea type heathy peatland II	VVTHP II
Dwarf scrub pine bog	DSPB
Dwarf shrub type heathy peatland I	DSTHP I
Low-sedge S. papillosum fen	LSSPF
Cottongrass pine fen	CPF
Low-sedge pine fen	LSPF
S. fuscum pine bog	SFPB
Cladina type heathy peatland I	CTHP I
S. fuscum bog	SFB
Ridge-hollow pine bog	RHPB
Low-sedge bog	LSB

IV. METHODOLOGY

This section introduces the components of our proposed methodology to high-resolution image segmentation in remote sensing. There are two main components, namely the input selection and the pixel-wise classification. The input selection responsible for selecting appropriate number of the inputs in order to reduce the number of dimension. The pixel-wise classification classifies the peatland based on the selected inputs in the first module. More details about two proposed phases will be provided in the following subsections.

A. Architecture

For both components, we use a multi-stream CNN fusion architecture as shown in Fig. 3. We assume three streams according to the different spatial resolution of inputs. The inputs of *Stream1* including S1, S2, RADARSAT-2 and TerraSAR-X with spatial resolution $10 \times 10m$. The inputs of *Stream2* are for DEM and CHM. As the spatial resolution of CHM is $1 \times 1m$, the corresponding DEM resampled values are $1 \times 1m$ pixels. *Stream3* has only one input including MS-NFI layers with resolution $16 \times 16m$. The leading idea of the three-stream architecture is to preserve all possible information of the different resolution inputs, instead of interpolating the inputs to the same resolution while losing potentially useful information. Each stream is comprised of two convolution layers, two pooling layers, and one fully connected layers. This architecture was found as a good compromise between the model complexity and number of training samples and classes in our problem. Convolutional layers are constructed with various convolution kernels with various sizes for each stream

depending on the window size. In *Stream1* and *Stream3*, the kernel matrix (e.g., 5×5) sweeps through the input image transforming the data. In *Stream2*, the kernel size is 25×25. Each of the convolutional layers is followed by a ReLU non-linearity, batch normalization [31] and max-pooling layer. Max-pooling was performed on each feature map using a 2×2 kernel with a stride of 1, thereby avoiding down-sampling and allowing for a high spatial resolution. After several convolution and pooling steps, the final feature maps are flattened in the form of a 1-D vector for each stream. The output of three stream are concatenated and the probabilities of land cover classes were extracted as output using a softmax function, which is commonly used as the activation function in the multiclass classification. To update network weights, we use Adam as an optimization algorithm [32]. In order to reduce overfitting, we also added dropout layer with a 50% drop probability after each convolutional layer.

B. Input Selection

To reduce the number of input dimensions and find most relevant inputs, we utilize Sequential Forward Selection (SFS) [33]. SFS is widely used for its simplicity and speed [33], [34]. SFS reduces an initial d-dimensional input space F to a k-dimensional input subspace S where $k < d$. As SFS is a bottom-up search procedure, it starts from an empty input set S and gradually adds inputs selected by an evaluation function, that maximize the accuracy. In the first iteration, the proposed CNN model is trained with all possible inputs one by one. The input I_i to be included in the input set S, which can achieve the maximum classification accuracy. In the second iteration, SFS takes the best input I_i from the first iteration and evaluates the classifier performance with the combination of I_i and the remaining available inputs. This is repeated until the CNN's performance is evaluated with all input combinations and finally the best input set S is selected.

The proposed CNN model is trained on small regions (windows), which are assumed to contain the spatial information around each training pixel. The window-based approach is commonly used in pixel classification [16]. In addition, this approach is well suited in this study since the annotated peatland site type is determined in the field around the training point location within a circle of 40 m in diameter. For this purpose, each training pixel is centered at a window with size $n \times n$. The parameter n depends on the spatial resolution of an input. For instance, we supposed $n = 5$ for inputs which have 10 and 16 m pixel size. We also assumed $n = 25$ for DEM and CHM which have 2 m pixel size.

For each window, we also used 5-fold Stratified Cross Validation (SCV) [35] technique to divide the dataset into training and validation datasets. SCV as one of the standard methods to evaluate classifier's generalization accuracy. Compared with the standard CV, SCV ensures that each fold of dataset has the same distribution of the classes in each fold to address class imbalance problem. We also applied data augmentation to generate additional training data by random rotation, vertical and horizontal flips. Finally, the proposed CNN is trained on whole data with the selected inputs. This trained model is used for performing the pixel-wise classification.

C. Pixel-wise Classification

This component labels each pixel according to the defined peatland site type classes. Therefore the output of this component is a pixel-wise classification map. Each pixel is classified based on the trained CNN model with the selected inputs. For this task, we proposed a sliding window approach. The sliding window plays the main role in many object detection and classification methods [36] that refers to a rectangle region with a defined width and height that moves over the image. In fact, each pixel is labelled one-by-one, with some amount of surrounding pixels as a spatial pattern to help to classify it. The pixel always is centered of the window. The classification starts from the upper left pixel. Then whole window's features are extracted separately from each stream and after that the class predictions from each individual stream are combined by first concatenating them and then applying additional convolutions. After classifying the upper left pixel, the window is moved into the next pixel. When the window arrives at the last pixel in the first row, it comes down one pixel and then starts again from the left pixel. This process iterates until the window slide over all pixels.

V. Experimental Results

We first applied SFS method to decrease the input dimension to prevent overfitting and decrease the running time for the proposed CNN architecture. The selected inputs by SFS for each version of data is illustrated in Table IV. The result show that the inputs such as S2, DEM, CHM, TerraSAR-X and MS-NFI are always selected in the two data versions. It means these inputs retain most information related to peatland classes. Therefore, the CNN model produces the highest classification accuracy with these features.

The most common evaluation measurement for classification algorithms is total accuracy which used in this paper. Table V shows that the overall classification accuracy is 32.4%, 33.6% and 31.8% for Keminmaa V1, Keminmaa V2 (drained) and Keminmaa V2 (undrained) based on the selected inputs. It demonstrates that the highest classification accuracy is gained with the V2 drained dataset. We also calculated other metrics such as Producer Accuracy (PA), User Accuracy (UA) and Kappa value. The PA represents how well a specific area can be mapped, whereas the UA is an indicator of how well the produced map represents what really exists in the study area. From the results, the highest value for PA, UA and Kappa values belongs to Keminmaa V1. Furthermore, the confusion matrix of two versions of data is shown in Table VI. In this matrix, the upper cell shows the predicted class name and the bottom cell shows the number of predicted samples. The columns "other" represents the total number of observation to classes that they have only one sample. In the last column in these tables, you can see the total and correct observations in upper and bottom cells, respectively. Therefore, the matrix can present the obtained results with more details for each

TABLE IV
THE SELECTED INPUTS FOR EACH VERSION OF DATA.

Name of dataset	Feature description (data type, raster layer and date)
Keminmaa V1	Sentinel-2 optical satellite data, May 26th, 2018
	RADARSAT-2 SAR satellite data, HV/HH, May 24th, 2017
	Sentinel-1 SAR satellite data, VV/VH, coherence July 8th / July 20th, 2017
	TerraSAR-X SAR satellite data, HH, May 30th - Jun 5th, 2017
	TerrasarX, HH, 30th May-5th Jun 201
	Sentinel-1 SAR satellite data, VV/VH, intensity Sep 26th, 2017
	DEM
	CHM
	MS-NFI, mean height for years 2000-2009
	LBP derived from CHM
	Steepest descent derived from DEM
	Slope derived from DEM
Keminmaa V2 (drained)	Sentinel-2 optical satellite data, July 18th, 2018
	LBP derived from CHM
	DEM
	TerraSAR-X SAR satellite data, HH, May 30th - Jun 5th, 2017
	MS-NFI, mean diameter for years 2000-2009
	CHM
Keminmaa V2 (undrained)	Sentinel-2 optical satellite data, Jun 19th, 2020
	Sentinel-1 SAR satellite data, VV/VH, coherence averaged for Jun - Aug 2019
	Sentinel-1 SAR satellite data, VV/VH, coherence Sep 18th / Sep 30th, 2017
	RADARSAT-2 SAR satellite data, HV/HH, Sep 21th, 2017
	TerraSAR-X SAR satellite data, HH, May 30th - Jun 5th, 2017
	Sentinel-1 SAR satellite data, VV/VH, coherence Jul 20th / Sep 18th, 2017
	Sentinel-1 SAR satellite data, VV/VH, intensity Aug 3rd, 2017
	LBP derived from DEM
	DEM
	CHM
	MS-NFI, site type for years 2000-2009

peatland class. The results show that we can get the maximum accuracy 82.0% and 78.1% from class "HRTHPI" in V1 and V2 (drained) respectively. This is because this class has the maximum and sufficient training samples in data for training CNN. In V2 (undrained), we got accuracy the maximum accuracy from class "SFB" and "TSHSF". The minimum accuracy achieved for the classes that have minimum number of samples such as "CSPF" and "RHPB" which is around 16%. The entire classification maps produced from the CNN for each dataset are shown in the first row of Fig. 4 with the final pixel size 2 m. From the second row of this figure, detailed classification results for three specific subregions (1-3) with CNNs are illustrated corresponding to S2. For instance, Fig. 4 (1)(b) shows the classification results for region 1 of Keminmaa V1. Fig. 4 (1)(c) and (d) show the classification maps for the same region of Keminmaa V2 drained and undrained, respectively.

VI. CONCLUSION

In this paper, we presented a CNN fusion architecture for peatland classification based on variety of multisource and multiresolution data. We first used state-of-the-art SFS to reduce the dimension of inputs while considering the most representative features. Therefore, our proposed architecture substantially reduced the amount of time needed to produce prediction peatland site type maps. Then, the architecture performed pixel-wise classification according to the selected features and a sliding window approach. Our experimental results are collected on satellite open source optical and SAR data and airborne laser scanning and multi-source national forest inventory in northern Finland on *aapa* mires. They show that our architecture able to classify peatland types around 32% total accuracy in the study area. However, the results show that for some classes with more training samples, the architecture can classify them with accuracy above 52%. For example, we got accuracy 82% for the class "Herb-rich type heathy peatland I (HRTHP I)" in the first version of data.

TABLE V
THE CLASSIFICATION PERFROMANCE METRICS FOR EACH DATA VERSION (%).

Name of dataset	Keminmaa V1	Keminmaa V2 (drained)	Keminmaa V2 (undrained)
Classification accuracy	32.4	33.6	31.8
Average user's accuracy	25.8	15.1	16.2
Average producer's accuracy	21.0	12.9	18.8
Kappa value	29.2	21.1	27.6

ACKNOWLEDGMENT

This work is part of the MaaTi project funded by the Ministry of Agriculture and Forestry of Finland. The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

REFERENCES

[1] Y. Ghobadi, B. Pradhan, K. Kabiri, S. Pirasteh, H. Z. M. Shafri, and G. A. Sayyad. Use of multi-temporal remote sensing data and gis for wetland change monitoring and degradation. In *2012 IEEE Colloquium on Humanities, Science and Engineering (CHUSER)*, pages 103–108, 2012.

[2] Statistics Finland. Greenhouse gas emissions in finland 1990 to 2020. *National Inventory Report under the UNFCCC and the Kyoto Protocol*, 2022.

[3] Ramanathan Sugumaran, James Harken, and James Gerjevic. Using remote sensing data to study wetland dynamics in iowa. 2004.

[4] JA Dechka, Steven Franklin, M. Watmough, R. Bennett, and D. Ingstrup. Classification of wetland habitat and vegetation communities using multi-temporal ikonos imagery in southern saskatchewan. *Canadian Journal of Remote Sensing*, 28:679–685, 10 2002.

[5] Floyd M. Henderson and Anthony J. Lewis. Radar detection of wetland ecosystems: a review. *International Journal of Remote Sensing*, 29(20):5809–5835, 2008.

[6] Ilkka Korpela, Markku Koskinen, Harri Vasander, Markus Holopainen, and Kari Minkkinen. Airborne small-footprint discrete-return lidar data in the assessment of boreal mire surface patterns, vegetation and habitats. *Forest Ecology and Management*, 258(7):1549–1566, 2009.

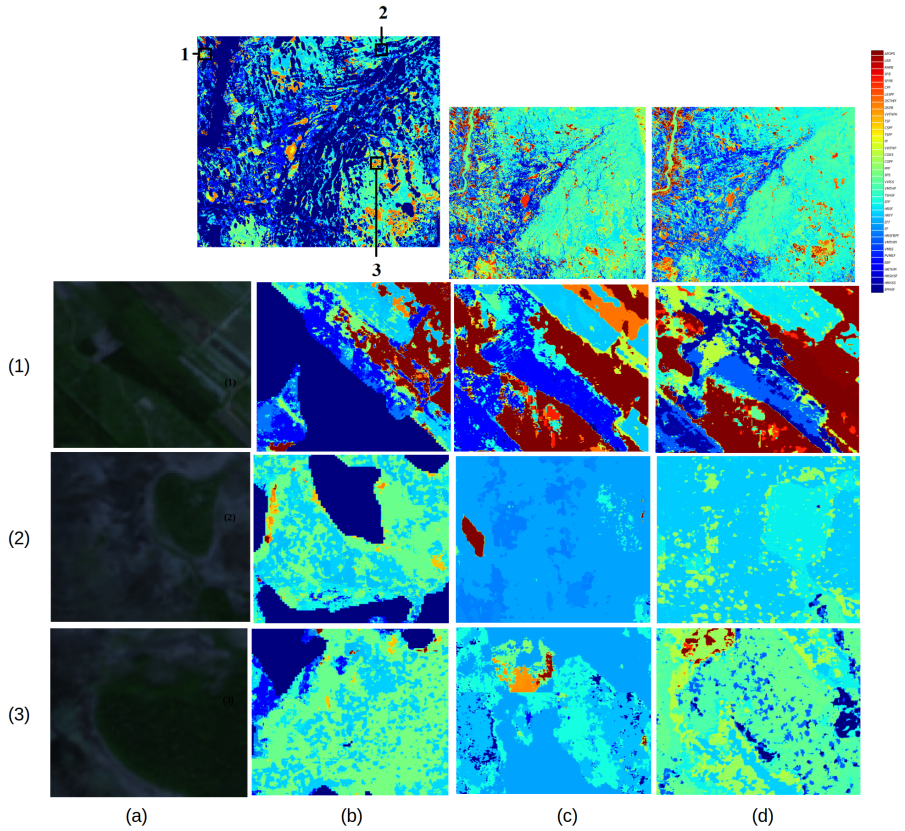

Fig. 4. First row shows the pixel-wise classification map for whole (b) Keminmaa V1, (c) Keminmaa V2 drained and (d) Keminmaa V2 undrained. From the second row, you can see more details of the classification map results for three proposed subregions (1-3) corresponding to the S2 image(a) in each version of data.

[7] Mohammad Rezaee, Masoud Mahdianpari, Yun Zhang, and Bahram Salehi. Deep convolutional neural network for complex wetland classification using optical remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(9):3030–3039, 2018.

[8] Miae Kim, Junghee Lee, Daehyeon Han, Minso Shin, Jungho Im, Junghye Lee, Lindi J. Quackenbush, and Zhu Gu. Convolutional neural network-based land cover classification using 2-d spectral reflectance curve graphs with multitemporal satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(12):4604–4617, 2018.

[9] Zhenfeng Shao and Jiajun Cai. Remote sensing image fusion with deep convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(5):1656–1669, 2018.

[10] Yuan Yuan, Jie Fang, Xiaoqiang Lu, and Yachuang Feng. Remote sensing image scene classification using rearranged local features. *IEEE Transactions on Geoscience and Remote Sensing*, 57(3):1779–1792, 2019.

[11] Fahimeh Farahnakian and Jukka Heikkonen. Deep learning based multi-modal fusion architectures for maritime vessel detection. *Remote Sensing*, 12(16), 2020.

[12] Michael Kampffmeyer, Arnt-Børre Salberg, and Robert Jenssen. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 680–688, 2016.

[13] Fahimeh Farahnakian and Jukka Heikkonen. A comparative study of deep learning-based rgb-depth fusion methods for object detection. In *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1475–1482, 2020.

[14] Usman Iqbal Ahmed, Arturo Velasco, and Bernhard Rabus. Semantic segmentation of land use / land cover (lu/lc) types using f-cnns on multi-sensor (radar-ir-optical) image data. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 4700–4703, 2021.

[15] Fahimeh Farahnakian and Jukka Heikkonen. Fusing lidar and color imagery for object detection using convolutional neural networks. In *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, pages 1–7, 2020.

[16] Pedro H. O. Pinheiro and Ronan Collobert. Recurrent convolutional neural networks for scene parsing. *CoRR*, abs/1306.2795, 2013.

[17] Natural resources institute Finland (LUKE). Advances in soil information. https://www.luke.fi/en/projects/maati, 2021.

[18] Nataliia Kussul, Mykola Lavreniuk, Sergii Skakun, and Andrii Shelestov. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5):778–782, 2017.

[19] Tim Rudner, Marc Rußwurm, Jakub Fil, Ramona Pelich, Benjamin Bischke, Veronika Kopackova-Strnadova, and Piotr Biliński. Multi3net: Segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33:702–709, 07 2019.

[20] Alex Owusu Amoakoh, Paul Aplin, Kwame T Awuah, Irene Delgado-Fernandez, Cherith Moses, and Carolina Peña Alonso. Tropical peatland classification using multi-sensor sentinel imagery and random forest algorithm in greater amanzule, ghana. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 5910–5913, 2021.

[21] Thierry Erudel, Sophie Fabre, Xavier Briottet, and Thomas Houet. Classification of peatland vegetation types using in situ hyperspectral measurements. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 5713–5716, 2017.

[22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet

TABLE VI
THE CONFUSION MATRIX FOR KEMINMAA V1 (GREEN), V2 DRAINED (RED) AND V2 UNDRAINED (BLACK).

EPHSF	HRTHP I 7,7,0	PVMSF 3,0,0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Other 2,2,3
HRHSS	HRHSS 3,8,0	HRTHP I 67,46,0	EBF 0,2,0	PVMSF 0,3,2	VMSS 0,3,2	VMTHP I 0,3,0	TSHSF 0,0,11	VVTHP II 0,3,0	AFOPS 0,3,2	AFOPS 0,0,2	-	-	-	-	-	-	-	-	-	Other 7,3,3
HRSHSF	HRTHP I 3,0,0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Other 1,3,0
HRTHP I	HRHSS 7,8,0	HRTHP I 275,248,2	EBF 2,8,0	PVMSF 4,10,3	VMSS 4,0,4	VMTHP I 5,4,0	HRSBPF 4,0,0	VMTHP 2,0,0	EPF 0,3,0	PPF 2,5,0	CGPF 3,2,0	VVIVHP 2,0,0	VVTHP 2,2,0	VVTHP II 0,15,0	SPB 4,3,0	AFOPS 7,3,2	TSHSF 0,0,9			Other 1,4,4
EHWSF	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Other 1,2,0
EBF	HRTHP I 8,5,0	EBF 3,4,0	CGPF 2,0,0	VVTHP II 10,7,0	PMSF 0,0,3	TSHSF 0,0,2	-	-	-	-	-	-	-	-	-	-	-	-	-	Other 3,3,2
PVMSF	HRHSS 6,3,0	HRTHP I 45,29,0	PVMSF 46,10,29	VMSS 3,0,4	VMTHP I 4,0,0	TSHSF 4,2,10	CGPF 5,4,2	CGSS 4,4,0	VVTHP 3,4, 0	TSF 2	EBF 0,2,0	HRSBPF 0,3,0	VVTHP II 0,4,0	EPF 0,0,2	PPF 0,0,3	AFOPS 0,0,2				Other 6,6,4
VMSS	HRHSS 11,8,2	HRTHP I 32,19,3	PVMSF 5,0,14	VMSS 18,2,7	TSHSF 2,0,8	PPF 4,0,3	CGSS 2,2,10	VVTHP 3,3,0	AFOPS 2,0,0	SPB 0,2,0	TSF 0,0,3	-	-	-	-	-	-	-	-	Other 3,1,3
VVMTHP I	HRHSS 3,3,0	HRTHP I 49,40,0	VMSS 5,0,0	VMTHP I 6	HRSBPF 3,2,0	PPF 2,0,0	PVMSF 0,3,4	VMTHP 0,2,0	EPF 0,0,0	VVTHP II 0,0,0	TSHSF 0,0,2	-	-	-	-	-	-	-	-	Other 6,4,1
HRSF	HRTHP I 2,2,0	HRSF 7,0,4	EPF 7,0,2	HRFF 3,0,2	HRSBPF 3,5,0	TSF 18	SPB 3,3,0	LSB 2,0,0	VVTHP II 0,4,0	EPF 0,0,3	EPF 0,0,3	-	-	-	-	-	-	-	-	Other 8,2,6
EF	HRSF 2,0,0	EF 4,0,8	EPF 4,0,2	HRFF 2,0,0	HRFBPF 2,0,3	CGPF 5,0,0	TSF 3	VVTHP II 0,0,0	TSHSF 0,0,2	EPF 0,0,2	-	-	-	-	-	-	-	-	-	Other 5,0,1
EFF	HRSF 3,0,0	EPF 0,2,18	HRFF 13,0,7	HRSBPF 2,0,0	EPF 5,6,6	TSF 4,0,9	EF 0,0,5	CGSS 0,0,2	-	-	-	-	FF 0,0,2	-	-	-	-	-	-	Other 6,8,6
HRFF	HRSF 2,8,0	EPF 9,3,8	HRFF 30,0,16	HRFBPF 2,0,0	FF 2,0,4	TSF 6,0,0	SPB 2,2, 0	OSAF 3,0,0	EF 0,0,3	EPF 0,0,4	LSB 0,0,2	-	-	-	-	-	-	-	-	Other 4,4,3
HRSBPF	HRTHP I 14,12,0	PVMSF 2,2,5	HRSBPF 26,11,0	EPF 5,14,6	TSHSF 3,3,6	PPF 5,0,0	CGPF 7,3,0	VVTHP 2,4,0	TSF 2,0,0	VMTHP I 0,2,0	VMTHP 0,6,0	VVTHP II 0, 3, 0	EF 0,0,2	-	-	-	-	-	-	Other 11,1,1
EPF	EFF 8,0,0	HRSBPF 11,0,0	EPF 8,8,16	VMTHP 4	VVITHP 2,2,0	TSPF 2,2,2	EBF 0,2,0	CGPF 0,4,0	EF 0,0,2	-	-	-	-	-	-	-	-	-	-	Other 9,4,2
TSHSF	HRTHP I 40,27,0	PVMSF 4,0,2	VMTHP I 2,0,0	HRSBPF 6,0,0	TSHSF 5,23,0	VVITHP 2,0,0	VVTHP II 6,2,0	EBF 0,8, 0	AFOPS 0,2,0	-	-	-	-	-	-	-	-	-	-	Other 8,4,4
VMTHP	HRTHP I 4,2,0	HRSBPF 5,2,0	EPF 2,8,0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Other 6,4,2
VVISS	HRHSS 4,0,0	HRTHP I 23,2,0	VMSS 2,0,0	CGSS 0,2,0	TSHSF 8,8,3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Other 4,9,3
SPS	HRTHP I 3,3,0	CGPF 2,0,2	PVMS 0,0,2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Other 0,3,0
PPF	HRTHP 6,0,0	PVMSF 9,0,0	VMTHP I 2,0,0	HRFF 2,0,0	HRSBPF 10,2,0	TSHSF 2,2,0	PPF 7,0,0	TSPF 3,0,0	HRTHP I 0,2,0	VMTHP 0,2,0	VVITHP 0,2,0	-	-	-	-	-	-	-	-	Other 5,4,2
CGPF	HRTHP I 8,3,0	PVMSF 5,0,5	HRSBPF 10,0,0	EPF 5,5,2	TSHSF 2,0,3	CGPF 11,11,3	VVTHP II 6,17,0	DSPB 3,0,0	SPB 8,0,0	EBF 4,4,0	-	-	-	-	-	-	-	-	-	Other 6,4,5
CGSS	HRHSS 3	HRTHP I 10,10,0	CGPF 6,2,0	CGSS 6,10,2	VVTHP II 9,17,0	SPB 3,2,0	VVTHP 0,2,0	PVMSF 0,0,2	-	-	-	-	-	-	-	-	-	-	-	Other 4,2,2
VVTHP	HRHSS 3	HRTHP I 50,38,0	PVMSF 4,4,0	HRSBPF 5	VMTHP 4	PPF 2,2,0	EPF 4,4,2	TSHSF 0,2,5	CGSS 0,3,0	VVTHP 0,6,0	VVTHP II 0,5,0	-	-	-	-	-	-	-	-	Other 7,4,2
FF	EFF 2,0,0	HRFF 2,0,2	FF 10,0,8	TSPF 2,0,0	TSF 3,0,0	LSSPF 0,0,2	-	-	-	-	-	-	-	-	-	-	-	-	-	Other 3,6,4
TSPF	HRTHP I 6,4,0	HRSF 2,0,0	HRSBPF 6,2,0	EPF 4,8,7	PPF 4,0,0	TSPF 5,0,0	TSF 2,0,0	CGPF 0,2,0	VVTHP 0,2,0	TSHSF 0,0,3	-	-	-	-	-	-	-	-	-	Other 5,3,3
CSPF	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Other 1,0,1
TSF	HRTHP I 3,3,0	HRSF 7,0,0	EFF 10,2,0	HRFF 5,0,0	HRFBPF 2,4,0	CGPF 4	TSPF 3,2,2	TSF 38,0,16	LSSPF 3,0,0	SPB 5,2,8	LSB 2,0,3	VVTHP II 0,4,0	EF 0,0,9	EFF 0,0,13	EPF 0,0,5	FF 0,0,2	-	-	-	Other 1,6,6
VVTHP II	HRTHP I 38,22,0	VMTHP I 4,0,0	HRFF 2,0,0	CGPF 3,6,0	CGSS 2	VVTHP II 57	SPB 3	AFOPS 10	EBF 0,7,0	HRSBPF 0,2,0	EPF 0,4,0	PVMSF 0,0,2	TSHSF 0,0,3	-	-	-	-	-	-	Other 7,5,6
DSPB	HRTHP I 6,3,0	HRPBPF 4,0,0	CGPF 3,2,0	PVMSF 0,2,0	VVTHP II 0,7,0	TSHSF 0,0,2	-	-	-	-	-	-	-	-	-	-	-	-	-	Other 8,3,2
DSTHP I	HRTHP I 17,11,0	HRSBPF 3,2,0	VVTHP II 2,5,0	DSTHP I 2,0,0	SFPB 2,0,0	EPF 0,2,0	VVITHP 0,4,0	-	-	-	-	-	-	-	-	-	-	-	-	Other 6,31,1
LSSPF	HRFF 3,0,0	TSF 6,0,2	LSSPF 6,0,0	SFB 3,2,7	LSB 2,0,2	EFF 0,0,2	-	-	-	-	-	-	-	-	-	-	-	-	-	Other 2,3,3
CPF	HRTHP I 5,5,0	EPF 2,2,0	CGPF 2,0,0	TSF 2,0,0	CPF 3,0,0	SFPB 7,0,0	SPB 4,0,0	LSB 2,0,0	VVTHP II 0,4,0	-	-	-	-	-	-	-	-	-	-	Other 2,14,0
LSSPF	EFF 0,0,2	TSF 0,0,2	SFB 0,0,7	LSB 0,0,2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Other 0,0,1
CPF	EPF 0,0,3	TSHSF 0,0,3	CGPF 2,0,7	SFB 0,0,3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Other 0,0,4
LSPF	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Other 2,2,0
SFPB	HRTHP I 4,0,0	EPF 2,2,2	TSPF 5,0,0	TSF 2,0,0	VVTHP II 4,4,0	DSPB 2,0,0	SFPB 6,0,0	SFB 9,0,4	HRFF 0,0,2	TSHSF 0,0,3	-	-	-	-	-	-	-	-	-	Other 4,0,7
SFB	HRTHP I 6,5,0	CGPF 2	CGSS 26,0	TSF 2,0,2	VVTHP II 7,17,0	SFPB 5	SPB 67,6,48	LSB 5,0,4	EPF 0,3,0	OSAF 0,4,0	AFOPS 0,2,0	-	-	-	-	-	-	-	-	Other 3,3,1
RHPB	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Other 1,0,1
LSB	TSF 7,0,2	LSSPF 2,0,0	SFB 7,0,13	LSB 8,0,4	EFF 0,0,2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Other 3,2,2
OSAF	OSAF 21,12,0	AFOPS 3,6,0	VVTHP 0,3,0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Other 0,3,0
AFOPS	HRTHP I 18,13,0	VVTHP II 16,6,0	OSAF 11,2,0	AFOPS 13,11,11	EBF 0,2,0	SFB 0,2,0	TSHSF 0,0,6	-	-	-	-	-	-	-	-	-	-	-	-	Other 2,4,3

classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[23] Vasander H. Hotanen J-P. Nousiainen H. Saarinen M. Penttilä T Laine, J. *Key to Finnish mire types*. 2018.

[24] Geological Survey of Finland (GTK). Suot ja turvemaat. https://gtkdata.gtk.fi/Turvevarojen_tilinpito/index.html, 2022.

[25] Kai Mäkisara, Matti Katila, Jouni Peräsaari, and Erkki Tomppo. The multi-source national forest inventory of finland – methods and results 2013. natural resources institute finland (luke). https://jukuri.luke.fi/handle/10024/532147, 2016.

[26] E.P Baltsavias. Airborne laser scanning: basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2):199–214, 1999.

[27] Matti Pietikäinen, Abdenour Hadid, Guoying Zhao, and Timo Ahonen. *Computer Vision Using Local Binary Patterns*. Springer Publishing Company, Incorporated, 1st edition, 2011.

[28] Erkki Tomppo and Matti Katila. Satellite image-based national forest inventory of finland for publication in the igarss'91 digest. *[Proceedings] IGARSS'91 Remote Sensing: Global Monitoring for Earth Management*, 3:1141–1144, 1991.

[29] Erkki Tomppo. The finnish multisource national forest inventory: small-area estimation and map production. Technical report, 2009.

[30] Erkki Tomppo, Markus Haakana, Matti Katila, and Jouni Peräsaari. *Multi-Source National Forest Inventory – Methods and Applications*, volume 18. 01 2008.

[31] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015.

[32] Fahimeh Farahnakian, Mohammad-Hashem Haghbayan, Jonne Poikonen, Markus Laurinen, Paavo Nevalainen, and Jukka Heikkonen. Object detection based on multi-sensor proposal fusion in maritime environment. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 971–976, 2018.

[33] A. Marcano-Cedeño, J. Quintanilla-Domínguez, M. G. Cortina-Januchs, and D. Andina. Feature selection using sequential forward selection and classification applying artificial metaplasticity neural network. In *IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society*, pages 2845–2850, 2010.

[34] Mineichi Kudo and Jack Sklansky. Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*, 33(1):25–41, 2000.

[35] Xinchuan Zeng and Tony Martinez. Distribution-balanced stratified cross-validation for accuracy estimation. *Journal of Experimental & Theoretical Artificial Intelligence*, 12, 04 2001.

[36] George Papandreou, Iasonas Kokkinos, and Pierre-André Savalle. Untangling local and global deformations in deep convolutional networks for image classification and sliding window detection. *CoRR*, abs/1412.0296, 2014.