Article

Assessment of Convolution Neural Networks for Surficial Geology Mapping in the South Rae Geological Region, Northwest Territories, Canada

Rasim Latifovic [1,*], Darren Pouliot [2] and Janet Campbell [3]

[1] Natural Resources Canada, Canadian Centre for Remote Sensing, 560 Rochester, Ottawa, ON K1A 0E4, Canada

[2] Environment and Climate Change Canada, Landscape Science and Technology, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada; Darren.Pouliot@canada.ca

[3] Natural Resources Canada, Geological Survey of Canada, 601 Booth Street, Ottawa, ON K1A 0E8, Canada; Janet.Campbell3@canada.ca

* Correspondence: Rasim.Latifovic@canada.ca; Tel.: +1-613-759-7002

Abstract: Mapping of surficial geology is an important requirement for broadening the geoscience database of northern Canada. Surficial geology maps are an integral data source for mineral and energy exploration. Moreover, they provide information such as the location of gravels and sands, which are important for infrastructure development. Currently, surficial geology maps are produced through expert interpretation of aerial photography and field data. However, interpretation is known to be subjective, labour-intensive and difficult to repeat. The expert knowledge required for interpretation can be challenging to maintain and transfer. In this research, we seek to assess the potential of deep neural networks to aid surficial geology mapping by providing an objective surficial materials initial layer that experts can modify to speed map development and improve consistency between mapped areas. Such an approach may also harness expert knowledge in a way that is transferable to unmapped areas. For this purpose, we assess the ability of convolution neural networks (CNN) to predict surficial geology classes under two sampling scenarios. In the first scenario, a CNN uses samples collected over the area to be mapped. In the second, a CNN trained over one area is then applied to locations where the available samples were not used in training the network. The latter case is important, as a collection of in situ training data can be costly. The evaluation of the CNN was carried out using aerial photos, Landsat reflectance, and high-resolution digital elevation data over five areas within the South Rae geological region of Northwest Territories, Canada. The results are encouraging, with the CNN generating average accuracy of 76% when locally trained. For independent test areas (i.e., trained over one area and applied over other), accuracy dropped to 59–70% depending on the classes selected for mapping. In the South Rae region, significant confusion was found between till veneer and till blanket as well as glaciofluvial subclasses (esker, terraced, and hummocky ice-contact). Merging these classes respectively increased accuracy for independent test area to 68% on average. Relative to the more widely used Random Forest machine learning algorithm, this represents an improvement in accuracy of 4%. Furthermore, the CNN produced better results for less frequent classes with distinct spatial structure.

Keywords: surficial materials mapping; surficial geology; deep learning; remote sensing

1. Introduction

Mapping of surficial geology is an important requirement for broadening the geoscience database of northern Canada. Surficial geology maps are an integral data source for mineral—exploration infrastructure, land-use planning, hazard assessment and other applications. However, mapping large remote regions can be a major challenge requiring significant labour and cost for timely production. With access to remotely sensed imagery, new machine learning approaches are emerging that support the surficial geological mapping of vast northern regions appropriate for regional scale mineral exploration and related land-use management. This is particularly relevant in Northern Canada, which is a huge territory that cannot easily be mapped following a systematic field sampling approach [1]. In addition, traditional mapping methods are based on interpretation of air photos requiring expert knowledge and experience. Interpretation is subjective, labour-intensive and difficult to repeat, while expert knowledge can be challenging to maintain and transfer [2]. Furthermore, fieldwork in remote regions is costly and logistically challenging.

Remote Predictive Mapping (RPM) is a relatively rapid and cost-effective way to generate an initial classification of surficial materials over large areas [3]. Machine learning applied to remote sensing data is an approach that can provide more consistent results over time and space assuming method, training samples and parameters are appropriately selected. Applying a well-trained model can reduce the subjectivity associated with expert interpretation. While it may not be as detailed or locally accurate as traditional methods, RPM products can guide fieldwork towards areas with more complex geology, serve as first order geologic maps in areas where little knowledge currently exists, and improve map production, accuracy and consistency by combining machine learning and expert interpretation.

There have been several successful examples of machine learning applied to remote sensing data for remote predictive mapping of surficial materials [4–9]. Most of the recent mapping approaches utilized Decision Trees (DT), Random Forest (RF) [10,11], Support Vector Machine (SVM) or bootstrap Maximum Likelihood (ML) algorithms [12]. In most surficial materials mapping cases, these algorithms are dominantly implemented in a manner that limit utilization of the spatial properties of the surficial material classes related to the geomorphology and class composition of base materials such as rock, gravel, sand, vegetation, and water. Often, surficial material classes are comprised of materials with very different spectral signatures. For example, till veneer is described as discontinuous sheet of diamicton, a poorly sorted sediment containing a mixture of grain-sizes from clay to boulders, overlying bedrock. Thus, to determine a particular surficial material class, a larger area needs to be sampled, such that spatial features and spectral signatures of material ensembles can be considered.

New machine learning algorithms based on deep learning provide a potential means to address some of the limitations of past efforts for surficial geology RPM. Deep learning is one of the fastest-growing trends in big data analysis and was deemed one of the 10 breakthrough technologies in the [13] MIT Technology Review of 2013. The name "deep learning neural network" reflects the inclusion of a number of hidden layers. Unlike their shallow counterparts, deep neural networks exploit feature representations learned exclusively from data, instead of hand-crafting features that are mostly designed based on domain-specific knowledge. In recent times, deep neural networks have received significant attention due to the development of graphic processing unit (GPU) technology, which has enabled large networks to be trained much more efficiently than previously. Another major advancement has been the development of deep neural networks designed specifically for image recognition tasks. These are known as convolution neural networks (CNNs). They have shown to be effective with high spatial resolution imagery where objects of interest have strong spatial structure [14–16]. CNNs take an image as input and apply convolution filters to these images to generate features that can discriminate image objects. The filters' weights are learned through stochastic gradient decent using error backpropagation and thus are optimized for the defined recognition task. The typical CNN architecture consists of a convolution filter layer followed by a down-sampling pooling layer. This sequence is repeated several times before a final dense layer that

summarizes the results and assigns the final output activation from the network [17–20]. Different architectures and methods are evolving rapidly for a range of recognition tasks.

In this research, we evaluate CNNs as means to improve surficial materials RPM for the case where a model is trained and applied in the same spatial domain and where it is trained from one area and applied in another. This CNN is compared against the more widely used RF algorithm as a benchmark to assess potential improvement.

2. Materials and Methods

2.1. Study Area Location and Physiography

The area of interest is located in the southeast corner of the Northwest Territories, Canada and defined by National Topographic System map sheet 75B, Abitau Lake (Figure 1). This region is underlain by the high-grade metamorphic rocks of the South Rae geological province of the Canadian Shield [21]. Most of the region exceeds 500 m elevation, ranging from ~420 m to 575 m above sea level. Ridged to hummocky crystalline rocks form broad sloping uplands and lowlands. Bedrock exposure varies from 0 to 40%. Swaths of streamlined terrain (till), till veneers and blankets, and minor hummocky moraine comprised of sandy, silty till of varying composition and thickness dominate the upland cover. Lowlands are primarily filled with lakes and wetlands (organic deposits). Large esker systems cross the area, dominantly oriented northeast to southwest (esker-glacial landform is a long winding ridge composed of stratified sand and gravel, which was deposited by a subglacial or englacial meltwater stream). This region lies within the Selwyn Lake Upland ecoregion that extends northwest from the Churchill River in Manitoba to the East Arm Hills at the eastern end of Great Slave Lake [22] (Ecoregions Working Group, 1989). It is classified as having a low subarctic climate and is part of the boreal forest–tundra transition zone extending from Labrador to Alaska, with tree cover decreasing north–northwest. The characteristic vegetation consists of stuntd black spruce, dwarf birch and Labrador Tea, with a ground cover of lichen and moss. Poorly drained wetlands are dominated by bog–fen sequences of black spruce, ericaceous shrubs and mosses. Much of the area has been burned by forest fires over the past several decades and regrowth has been slow. Permafrost is extensive and discontinuous with low to medium ice content throughout the ecoregion. More detailed descriptions of the region's geology and landscape can be found in [23] Campbell and Eagles [24], Pehrsson et al. [25], and Campbell et al. In this study, five subareas, within National Topographic System Abitau Lake map sheet 75B (Figure 1) were used.
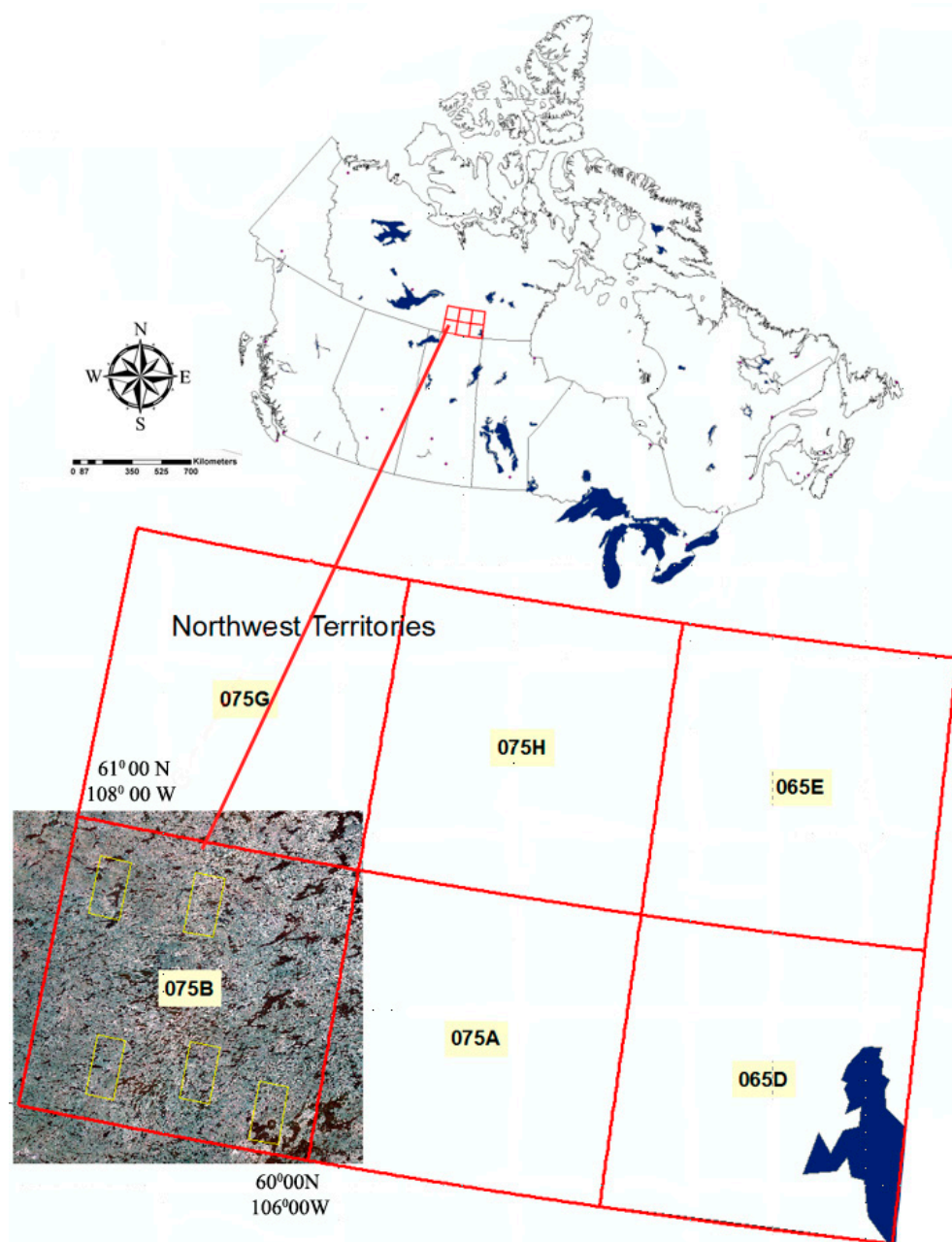
Figure 1. Location of the study area.

2.2. Data

2.2.1. Landsat Time Series

Landsat L1G TM/ETM+ scenes used were acquired by the U.S. Geological Survey between July and August, 1984–2012. To create a reflectance time series, the following processing steps were applied: reprojection, calibration, and cloud/cloud shadow detection using the processing methodology described in Latifovic et al. (2015) [26]. The Landsat L1G data are in the UTM map projection. However, the UTM is not an appropriate projection for large areas that cross several UTM zones. The Lambert Conformal Conic (LCC) projection is typically used in Canada for large area spatial datasets, as it does not require separate zones, yet keeps distortion to an acceptable level. The

LCC projection was specified with two standard parallels at 49°N and 77°N. The central meridian was 95°W and the latitude of origin was 0°.

Remote mapping of surficial materials in part relies on the association between vegetation and the underlying geological conditions. Fires can be problematic because vegetation is more indicative of the post-fire succession than an expression of surficial geology. However, fires expose bedrock potentially improving its detection. For this research, it was clear that removing fires would be beneficial as much of the area is tree covered. Ultimately, we are seeking to retrieve the sensor observations between 1985 and 2012 that represent vegetation in its latest successional stage.

Analysis of the historical fire database [27] suggests that approximately 80% of the burned area in the region occurred after 1984 and thus the observations can be replaced. To accomplish this, and develop high quality data for analysis, a data synthesis approach was developed and applied to the Landsat time series, using the normalized difference water index (NDWI) defined in Gao (1996) [28] as a selection criterion rather than normalized difference vegetation index (NDVI). The NDWI is a measure of liquid water molecules in vegetation canopies that interact with the incoming solar radiation [26] while NDVI is measure of chlorophyll absorption. In this study, NDWI is considered a better index for selection because it is significantly less affected by atmospheric scattering and absorption, it is much more sensitive to vegetation structure and it does not saturate as quickly as NDVI. In the first processing step, yearly average reflectance from clear-sky pixel observations were computed for each year and then a temporal smoothing filter was applied using a window size of three years. The maximum NDWI values from the smoothed time series were determined and, in the second step, used to find all observations in the time series within ±10% range. These values were averaged to get a robust reflectance measure for each Landsat band. Figure 2 shows that most of the area affected by fires were replaced with observations obtained before fire occurred, although the burn-scare of some pre-1984 fires are still evident in the image.

Another advantage of this processing approach was that it greatly improves image quality by averaging observations over multiple years. The blue band is usually too noisy to be used for analysis due to its strong sensitivity to atmosphere conditions. Nevertheless, Figure 3 shows an example of the blue band reflectance for both the best available measurement from the 2009–2011 composite, and the long-term average composite developed for use in this study. Much of the noise in the shorter-term composite has been removed using the described processing steps. Time series data processing provides a potential advantage as the blue band reflectance makes available specific information that improves classification of bedrock lithology and surficial materials. In addition to the band top of atmosphere reflectance, the Tasseled Cap Transformation (TCT) [29] was applied to produce Brightness, Greenness and Wetness orthogonal components, which allow for reduction of data dimensionality. The TCT was generated using coefficients provided by Huang et al. [30] for ETM and by Crist [31] for TM sensors.
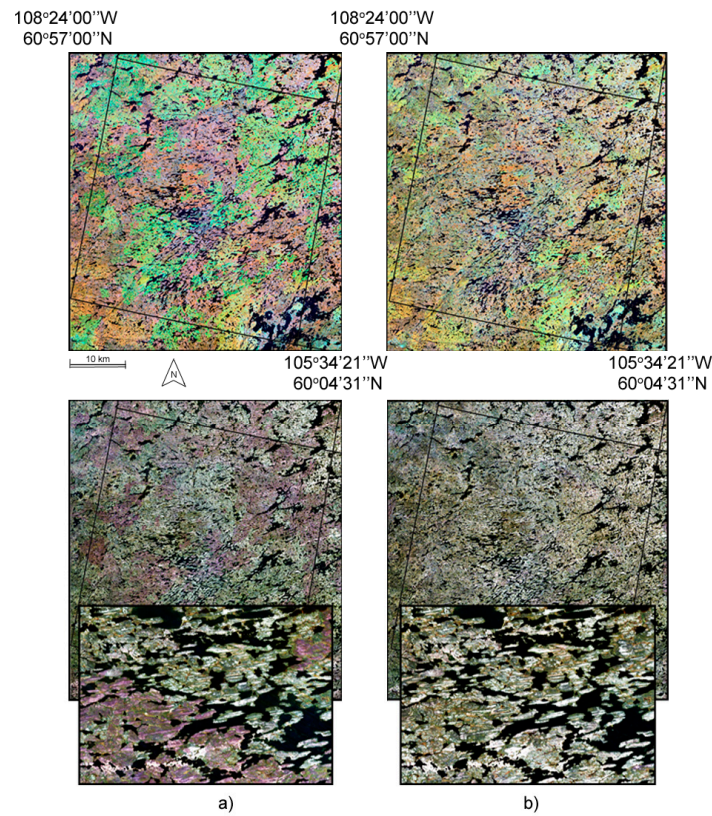
Figure 2. Landsat mosaics of National Topographic System map 75B: The image (**a**) shows the long-term average. The image (**b**) shows the long-term average with fire scars removed. The top image is RGB composite with near infrared band in red, shortwave infrared band in green and blue band in red. The bottom mosaics and the zoomed-in examples are true colour images composite as red band in red, green band in green, and blue band in blue.
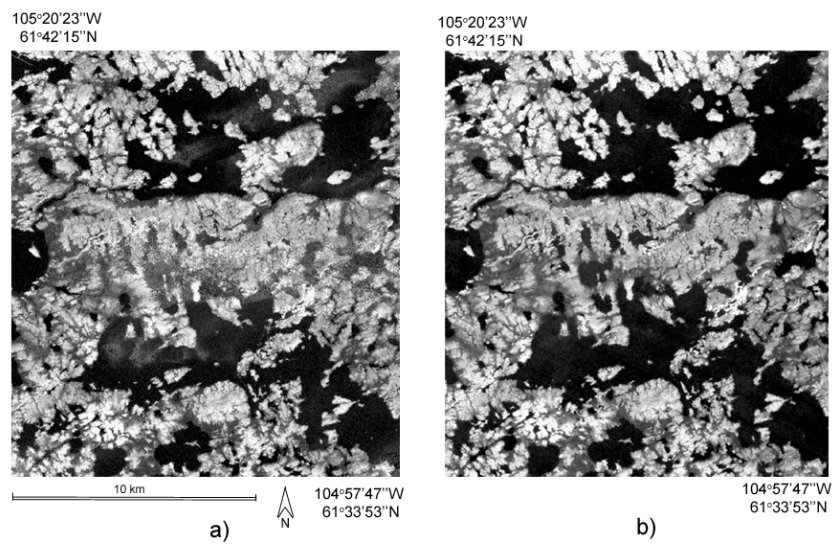

Figure 3. Blue band comparison: the image (**a**) shows the best available measurement from 2009–2011. Image (**b**) shows the long-term average from 1984–2011 generated using the temporal processing.

2.2.2. Air Photos

Historical air photos at approximately 1:50,000 scale are commonly used by the Geological Survey of Canada (GSC) for mapping surficial materials and geomorphology by expert interpretation. These easily accessible and affordable photos [32] can be scanned to high spatial resolution (typically <2 m), and cover large extents of Canada. However, they can often be difficult and time-intensive to geolocate, radiometrically balance, and mosaic into an accurate and seamless product suitable for large area interpretation and mapping applications. An automation of the required processes and methodology, for processing ~400 images covering a large portion of the South Rae region, was developed in this study. Landsat data were used as a spatial reference, as the intent was to use the resulting air photo mosaic for data fusion, to achieve a high-resolution multi-spectral image for surficial materials and/or geological mapping. Pre-processing, processing diagnostics, iterative corrections, radiometric normalization, and fusion were automated to generate an accurate and seamless product. This methodology can be applied anywhere in Canada to provide enhanced data for interpretation or mapping of geological attributes.

2.2.3. Digital Elevation Model

Two digital elevation model (DEM) data sets were used for terrain characterization (Figure 4). The first was the 1:50,000 scale Canadian Digital Elevation Data (CDED) rasterized to 30 m resolution. The second data set was from the Arctic DEM project, which is a National Geospatial-Intelligence Agency-National Science Foundation public-private initiative to automatically produce a high-resolution, high quality, digital elevation model of the Arctic using optical stereo imagery. For the current research, the 8 m spatial resolution product was acquired and used [33].
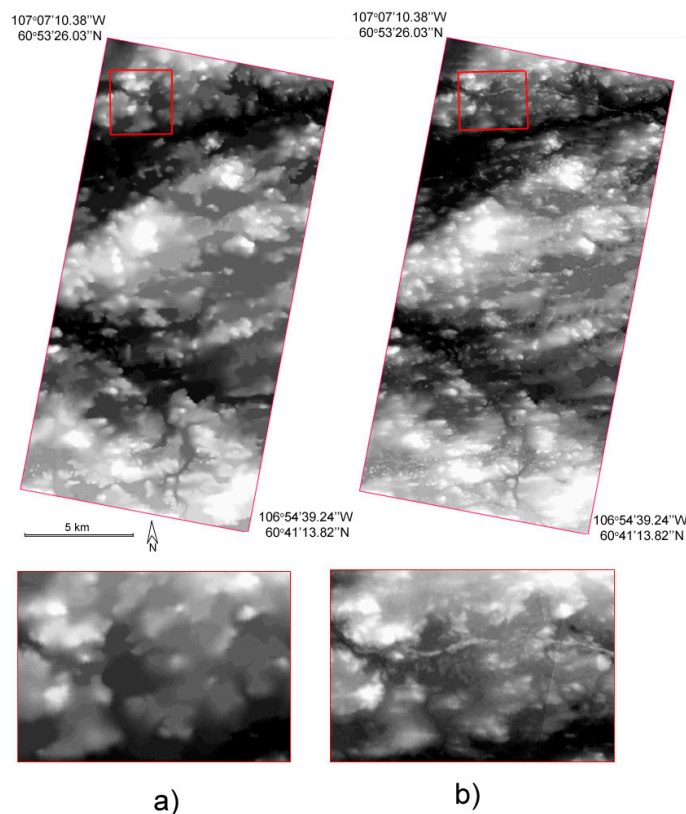

Figure 4. Examples of the two-digital elevation datasets used. The image (**a**) shows the Canadian Digital Elevation Data 1:50,000 scale rasterized to 30 m and image (**b**) shows the 8 m Arctic digital elevation model.

2.2.4. Surficial Materials Reference Data

The surficial materials reference data used includes ground truth observations acquired during field campaigns and a surficial geology map generated by expert interpretation. The surficial material labels for ground truth samples (Figure 5a) were defined using in situ field records. Secondary reference samples were defined by interpretation of photographs (Figure 5b) taken from the helicopter during flights between ground truth sites. The reference surficial geology map (Figure 5c) from [34], produced by expert interpretation of air photos and fine resolution WordView2 satellite images, remote observations and ground truth data, was used as a main source of reference data for the majority of training and testing samples.
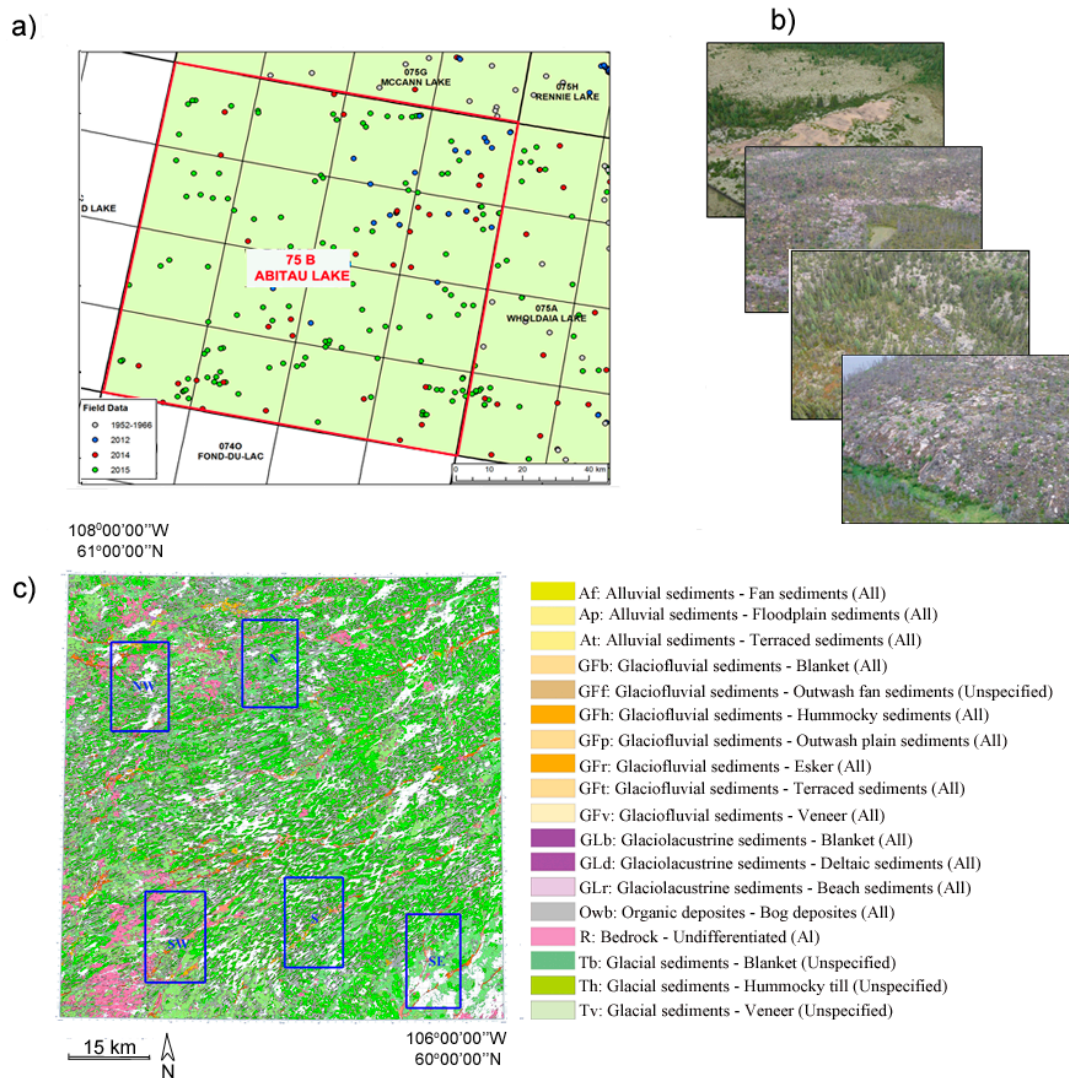

Figure 5. Surficial materials reference data sets: (**a**) ground truth; (**b**) aerial photographs from helicopter; and (**c**) surficial geology map [34].

The percent area of each surficial material unit in the reference map, shown in Figure 5c, is given in Table 1. The main surficial materials are tills, organics, water, and exposed rock. There are several glaciofluvial sediment units covering a small percentage of the total area. Only units highlighted in bold in Table 1 were considered for mapping in this analysis. Other units did not cover a sufficiently large area to be used for training and testing models.

Table 1. Surficial geology units and distribution within an National Topographic System (NTS) map sheet 75B.

Value	Percent Area	Code	Description
1	**8.8**	**Tv**	**Glacial sediments: till veneer**
2	**18.9**	**Owb**	**Organic deposits**
3	**34.1**	**Tb**	**Glacial sediments: till blanket**
4	**4.4**	**R**	**Bedrock: undifferentiated**
5	**9.9**	**Th**	**Glacial sediments: hummocky till**
6	**1.3**	**GFh**	**Glaciofluvial sediments: hummocky**
7	**0.7**	**GFt**	**Glaciofluvial sediments: terraced**
8	**1.2**	**GFr**	**Glaciofluvial sediments: esker**
9	**0.1**	**GLr**	**Glaciolacustrine sediments: beach**
10	0.02	At	Alluvial sediments: terraced
11	0.03	Ap	Alluvial sediments: floodplain
12	0.00	Af	Alluvial sediments: fan sediments
13	0.00	GFf	Glaciofluvial sediments: outwash fan
14	0.00	GLd	Glaciolacustrine sediments: deltaic
15	0.00	GFv	Glaciofluvial sediments: veneer
16	0.01	GLb	Glaciolacustrine sediments: blanket
17	0.01	GFb	Glaciofluvial sediments: blanket
18	0.00	GFp	Glaciofluvial sediments: outwash plain
19	**20.53**	**H_2O**	**Water**

2.2.5. Data Cubes

To simplify processing and analysis, data were organized over five subareas as data cubes (Figure 6). Each data cube contained all data layers used in the analysis, including Landsat bands, DEM, TCT, mosaicked air photos, surficial material classes, vegetation indices, land cover, etc. Data cubes were resampled into a common georeferencing framework with the same spatial resolution of 2 m. Initial exploratory analysis showed that the use of air photo data with 2 m spatial resolution was one of the more critical inputs to mapping accuracy. Landsat pixels (30 m) were replicated into 15 by 15 pixels with the same value. No resampling was applied because CNN convolution and pooling operations would achieve the optimal resampling for this application. Locations of the data cubes are depicted in Figure 5. They are referred to as North (N), Northwest (NW), South (S), Southwest (SW) and Southeast (SE). Based on this data structure, software was developed for efficient data manipulation such as creating training or testing samples with different sample sizes and features, evaluating and applying RF and CNN models, comparing classified images and deriving statistics.
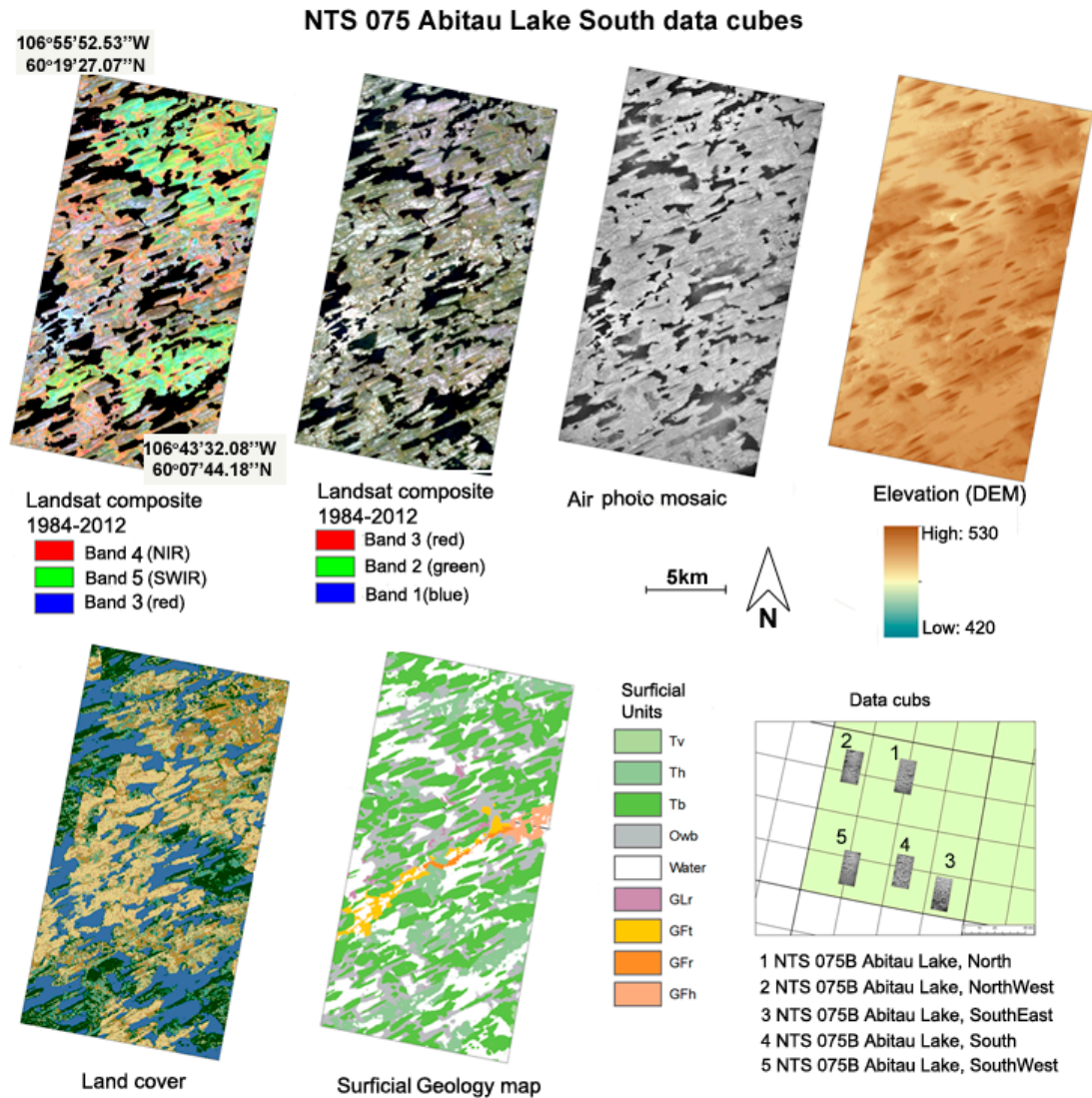
Figure 6. Example of the data cubes for National Topographic System map 075B Abitau study: these data sets are for the South subarea.

2.3. Methods

To evaluate both CNN and RF methods, we followed the procedure presented in Figure 7. The upper three rows present steps regarding data and data preparation. These steps were described in the previous sections. The bottom part of the Figure 7 shows steps involved in assessment of the RF and CNN for two scenarios, first when models were trained and applied over the same area, second when models were trained over one area and evaluated in a different area. The latter is referred to as independent test area. The following sections provide other relevant elements of the analysis including features used, RF parametrization, CNN architecture and the sampling and training approach.
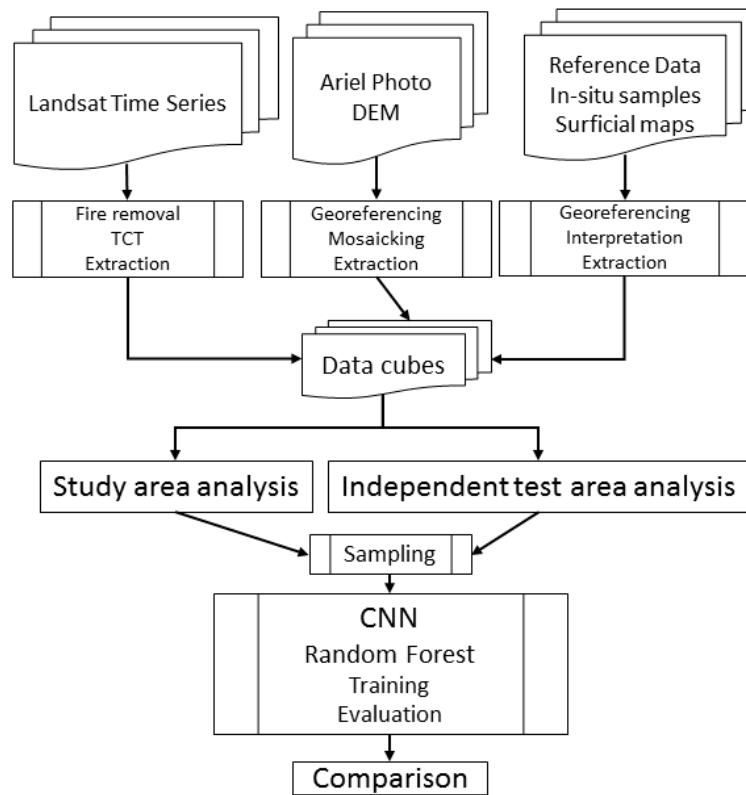
Figure 7. Assessment of Random Forest and Convolution Neural Networks for mapping surficial materials.

2.3.1. Generation of Training and Testing Data

An initial analysis and evaluation revealed that accuracy of the classification with the full thematic legend was not satisfactory. Thus, we developed a reduced legend with more generalized surficial material classes by merging some of the classes for the legend given in Table 1. In order to get a better spatial depiction of eskers and glaciolacustrine beaches, these two classes where added. Table 2 shows the generalized legend with 8 classes used in the assessment.

Table 2. Merged legend referred to as Legend 2.

Value	Class
1	Tills (merged Tv and Tb)
2	Organic deposits (Owb)
3	Bedrock (R)
4	Hummocky till (Th)
5	Undifferentiated glaciofluvial sediments (GFr, GFh, GFt): sand and gravel
6	Eskers (interpreted from high resolution imagery): sand and gravel
7	Glaciolacustrine beaches (interpreted from high resolution imagery): sand
8	Water (H_2O)

For each class, we used all available in situ samples, samples generated by interpretation of aerial photographs and samples extracted from the reference surficial geology map. The majority of samples came from the reference map; they were selected as randomly stratified samples from each subarea. The surficial geology map (Figure 5c) was used for sampling where a 20-pixel erosion operator was applied to avoid including samples at the edges of classes. Samples for eskers and

glaciolacustrine beaches were manually digitized from the high-resolution aerial photos. Thus, the sample set size for each subarea was 8 classes × 5000 samples = 40,000 samples. For training and testing models for each subarea, the sample was split to 90% for training and 10% for testing. Due to differences of the input data requirements for the RF and CNN, training and test data features were prepared differently. A number of different features combinations, spatial resolution, sampling unit size, and number of samples were assessed for each method. Final selection of the sampling design and model parametrization is described in the following Sections 2.3.2 and 2.3.3.

2.3.2. Convolution Neural Network

A number of different CNN configurations were tested ranging from simple two-layer networks to more complex configurations based on Alexnet [14] and residual convolution networks (RESNETs), respectively [15]. The best results were achieved with a 34 layer RESNET, which is consistent with published research. It consists of an initial convolution layer followed by blocks structured for residual learning and a final average pooling, overall 2.1 M parameters. Residual networks have been shown to be more robust to overfitting than other architectures. The sample unit size i.e., input image size was 96 by 96 pixels, and the selected number of input layers was seven including the air photo mosaic, Landsat TCT components brightness, greenness, and wetness, 8 m DEM, slope, and elevation variance. We tested input image windows of various sizes and found the selected size 96 × 96 the best suited to capture the main spatial properties of the desired classes. Using a smaller input window size improves delineation between class boundaries, but, if too small, the distinct spatial patterns can be missed. In exploratory analysis, we tried using Landsat pixel size 30 m as base resolution by resampling air photo and DM, but the achieved accuracy was very low when compared to the case with 2 m spatial resolution. We trained a new model using only data from this study because we used nine non-standard input features, and most widely available pretrained models are based on three band true color images. To train the model, we used 100 epochs run on a Tesla K20c GPU with batch size of 50. The training rate for the first 20 epochs was set at 0.01 and remaining 50 at 0.001

2.3.3. Random Forest

For Random Forest classification, we used the Open Source Computer Vision Library [35]. The following values of the hyper-parameters were selected: the maximum possible depth of the tree (maxDepth = 30), the number of samples in a node to be split (MinSampleCount = 0.1%), the size of randomly selected features at each tree node that were used to find the best split(s) (ActiveVarCount = 3) and the number of trees (TC = 100). The values for maxDepth, TC and MinSampleCount were defined by varying one parameter at a time while keeping other two fixed and comparing the model performance with accuracy. We used the same input layers sets as in CNN, but the window size was reduced to 15 × 15 to be consistent with past RPM research using machine learning approaches in [9,12]. In these studies, entropy measures within 7 × 7 pixel windows were used derived from Landsat. However, more recent work only used the entropy from a 30 m DEM. Thus, for each data feature, the mean was computed within the 15 × 15 window, which is equivalent to a Landsat pixel. However, for the air photos, we also computed the standard deviation to provide a texture measure. DEM entropy would be captured by the mean of the DEM variance calculated at the down sampled 2 m resolution. In initial trials, we tested different window sizes and number of trees and found little overall differences. We could have used the CNNs to identify features and fed those into the RF, but this would be little different from the CNN itself. Thus, we elected to compare with current common practice.

3. Results

3.1. Assessment of CCN (RESNET) vs. RF Methods for Surficial Materials Mapping

To evaluate both CNN and RF methods for surficial materials prediction over the study area, we extracted samples from all subareas (Figure 5c), trained models, applied these to generate maps and

compared these to holdout samples over each subarea. The overall accuracies for this scenario are shown in Table 3. To evaluate independent test sets for each subarea, the models were trained on four subareas and applied to the held-out fifth subarea. Table 3 shows the accuracies for this analysis. Accuracy with training samples inside the mapped area is generally the same between the methods with an accuracy of ~76%. However, the results over independent test sets were higher with RESNET.

Table 3. Results of accuracy assessment for different scenario and legends.

Method	Subarea	Sample from All Subareas	Independent Test Sets	Independent Test Sets (with Merged Tv and Tb Classes)	Independent Test Sets (with Merged Tv, Tb, GFr, GFh, GFt Classes)	Independent Test Sets with Legend Shown in Table 2
RESNET 18	N	0.76	0.61	0.64	0.65	0.66
	NW	0.77	0.49	0.51	0.55	0.59
	S	0.75	0.49	0.61	0.63	0.7
	SE	0.74	0.64	0.66	0.68	0.63
	SW	0.83	0.52	0.66	0.68	0.66
Average		0.77	0.55	0.62	0.64	0.64
RF	N	0.77	0.58	0.60	0.61	0.60
	NW	0.78	0.48	0.50	0.54	0.52
	S	0.76	0.50	0.62	0.62	0.63
	SE	0.76	0.60	0.62	0.63	0.61
	SW	0.78	0.50	0.63	0.63	0.64
Average		0.77	0.53	0.59	0.61	0.60

The results in Table 3 do not suggest a large improvement with RESNET. Overall accuracy is rather low reflecting difficulty of separating classes. This is due to the nature of the class distribution where the majority of the area consists of classes not strongly defined by spatial properties. These include organics, water, and till making up ~84% of the area. If we examine class specific accuracies (computed as the average of the user's and producer's accuracy), we see a greater improvement in the classes that are more defined by spatial structure such as hummocky till, eskers, and beaches, which improve on RF by 16% on average for the independent test sets results (Figure 8).
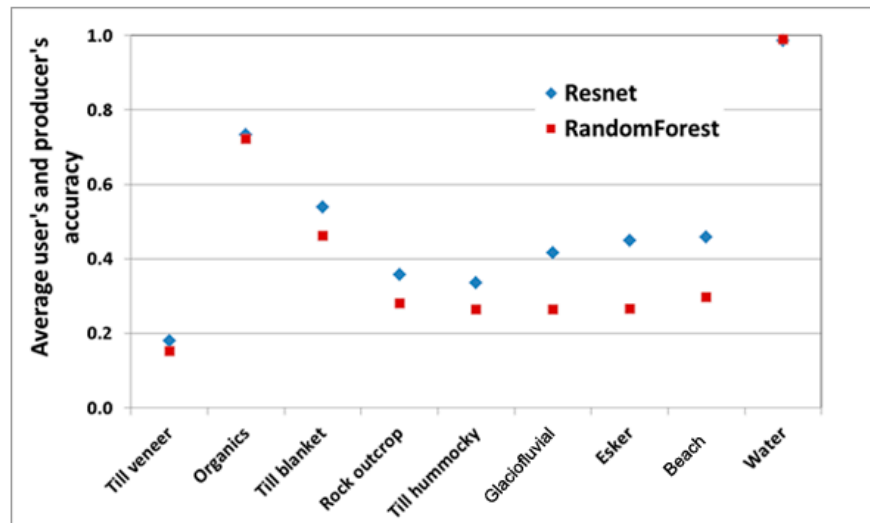

Figure 8. Accuracy of rare classes.

3.2. Assessment of RESNET Surficial Materials Mapping and Spatial Extension

Subsets of the surficial materials maps generated by RF and RESNET are shown in Figure 9. Visual examination of the RESNET results with the surficial geology reference map suggests a reasonable general agreement that is consistent with the accuracy assessment (Table 3). In general,

the spatial structure and main classes are reasonably well captured (Figure 9). The map generated by RF shows less agreement, particularly for the classes strongly defined by spatial properties.
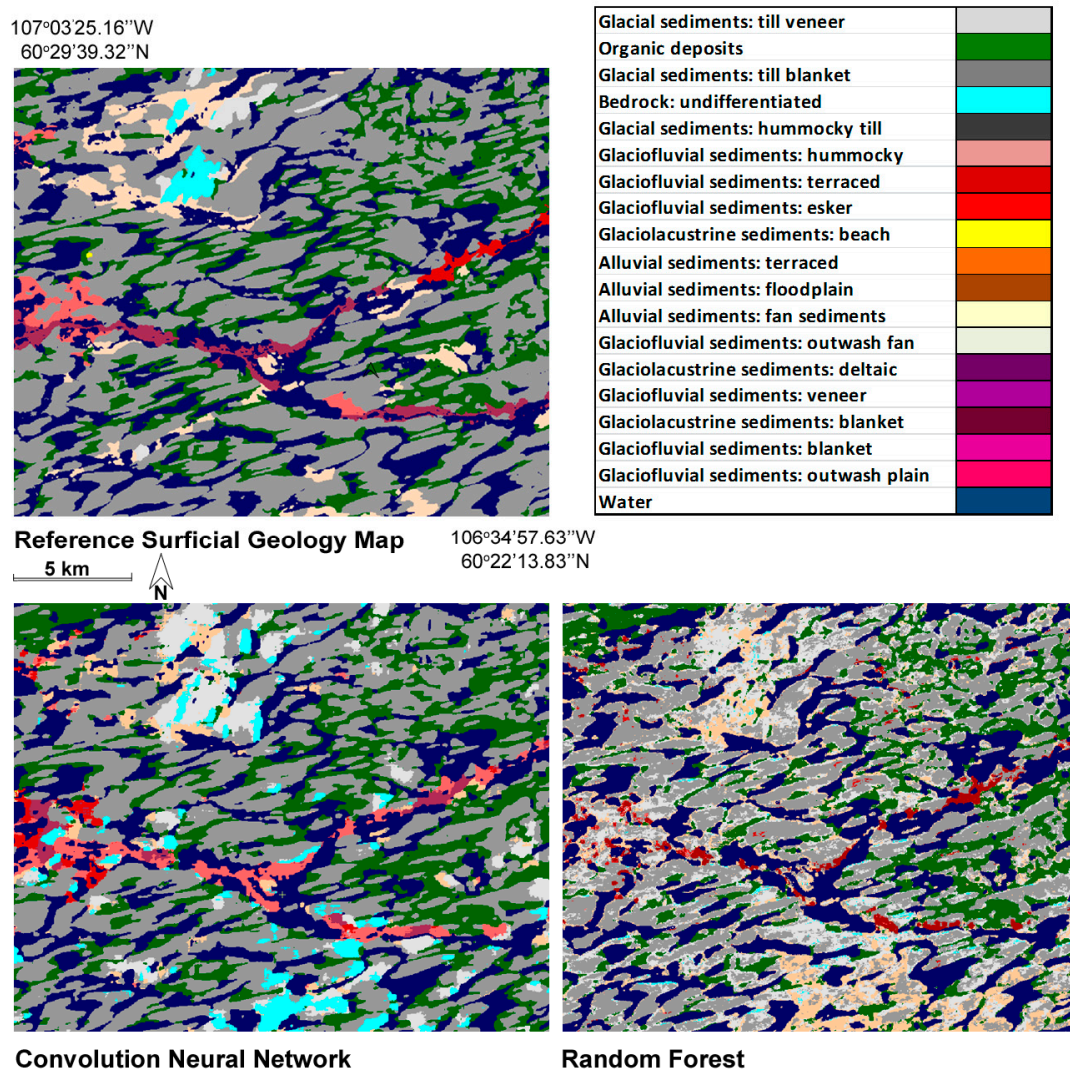

Glacial sediments: till veneer	
Organic deposits	
Glacial sediments: till blanket	
Bedrock: undifferentiated	
Glacial sediments: hummocky till	
Glaciofluvial sediments: hummocky	
Glaciofluvial sediments: terraced	
Glaciofluvial sediments: esker	
Glaciolacustrine sediments: beach	
Alluvial sediments: terraced	
Alluvial sediments: floodplain	
Alluvial sediments: fan sediments	
Glaciofluvial sediments: outwash fan	
Glaciolacustrine sediments: deltaic	
Glaciofluvial sediments: veneer	
Glaciolacustrine sediments: blanket	
Glaciofluvial sediments: blanket	
Glaciofluvial sediments: outwash plain	
Water	

Figure 9. Example area of reference surficial geology map, RF and RESNET result.

In Figure 10 NW cube, the glaciolacustrine beach class was added and is fully predicted in the result image only. It is a source of the differences observed as only a small area in the training maps was modified to include the beach class. There is confusion between thin and thick till and glaciofluvial subclasses. However, due to the uncertainty associated with the initial mapping, these results appear reasonable. Surficial geology mapping by air photo interpretation can be subjective and somewhat generalized spatially. Comparing surficial geology maps along map boundaries generated by different interpreters often reveals the difficultly of consistently mapping detailed classes between experts. This is largely a function of the thematic detail of the map products where generalization to simpler classes reduces the problem. For example, differences can exist between rock outcrop and thin till, but confusion between rock and thick till would be rare. Another factor to
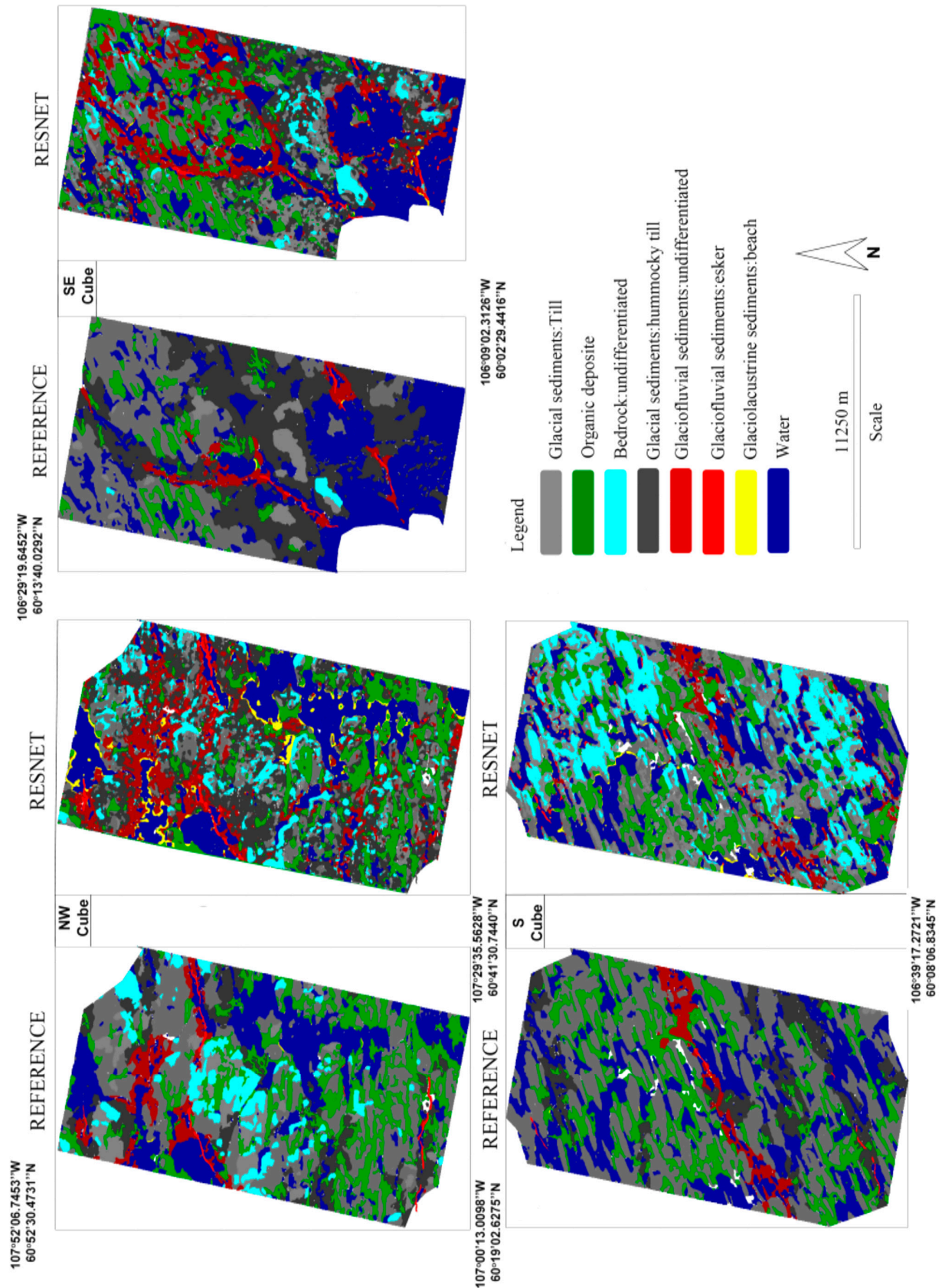
Figure 10. RESNET classification results for Northwest, South and Southeast subareas.

consider is the spatial generalization of the reference surficial geology maps. The mapping process carried out by geological experts comprise careful analysis of data and knowledge of geological processes; an expert generalizes the region by depicting polygons that mostly contain a single class. However, often patches of other classes can exist within the polygon. For example, a polygon labeled as thin till is likely to contain areas of exposed rock. The use of the expert-based surficial geology maps is challenging for training machine learning algorithms and the accuracies are likely higher then what was reported here due to these factors.

Overview results for three of the subareas are shown in Figure 10 for the RESNET results. These are independent test area results, which generally produced good agreement. They contain greater spatial variation than the reference maps, which is one of the larger differences in the comparisons. The Northwest (NW) subarea (Figure 10) has the lowest agreement with the reference map and appears to overestimate glaciofluvial and hummocky till classes due to complex terrain in the area. The South (S) subarea strongly overestimates rock outcrop (Figure 10). Converting this to thin till greatly improves the agreement between the reference and predicted results by approximately 10%. The Southeast (SE) subarea overestimated organics relative to the reference map (Figure 10). However, examination of the Landsat data suggests an underestimate of organics in the reference map.

4. Discussion

Deep learning is gaining increasing interest in remote sensing as an approach for development and enhancement of mapping applications [36]. Remote sensing presents some new challenges for deep learning because it aims at retrieving geo-physical or bio-chemical quantities rather than detecting or recognizing objects. Much of the knowledge of deep learning CNN performance is based on benchmark databases such as CIFAR-100 [37] or ImageNet [38]. These databases generally contain objects with distinct spatial structure and thus the objective is ultimately to capture and encode that structure in the network. In this research, the surficial geology classes are more variable existing over a large range of scales with weakly distinct spatial structure between some classes. In addition, the subjective nature of expert geological interpretation causes some conflicts in the training dataset that would not be seen to the same degree in benchmark databases. Thus, this analysis is unique in that it presents a significant deep learning challenge to separate classes with strong spectral and spatial confusion in the presence of training data error.

From this analysis, we can summarize some of the main considerations for the use of CNNs for remote sensing applications that are consistent with other research. Deep learning approaches require large and consistent training datasets to work well, as well as the infrastructure to train them in a reasonable time. The lack of sufficient training samples might cause severe overfitting and, therefore, greatly limit the capability of generalizing the model. It requires data with much finer spatial resolution to be able to extract high-level, hierarchical, and abstract features, which are generally more robust. Data augmentation is a common approach to train a CNN with a small sample. In this research, we did not use data augmentation as large samples were extracted. However, because of the strong dependence on sample quality, an improvement for future work would be to work with the geological expert to select and carefully quality control a small sample to be used in a data augmentation scheme. In addition, deep learning algorithms require much more experience. Setting up a network is much more tedious than using off-the-shelf classifiers such as random forests, as an in-depth knowledge of network architecture is needed. The architecture is important to achieve top performance, but, like most machine learning algorithms, the quality of the input data is generally more critical than the specific algorithm used. This was the case for the set of CNN architectures tested in this research.

Another important aspect of CNNs is that they develop features from the inputs that are directly optimized for the specified object recognition task. Thus, only features that cannot be obtained from convolution filters should be used as inputs to the CNN. This also suggests that the classifier should be more robust to input feature sets than other approaches, provided steps to avoid overtraining are

undertaken. In initial experimentation, we tested several features set modifications without major changes in the accuracy.

Apart from the potential performance advantage, the ability to use and adapt a pre-trained model is considered a potential unique advantage of CNNs. A CNN can be trained from an initial large training dataset and model weights can be updated for a specific region with a new training sample. For surficial geology mapping, we see deep learning CNNs as providing initial predictions that are refined by geologist and fed back into the model in an ongoing cycle reducing error and adapting to new or local conditions. This integrates the advanced knowledge of geological experts and ideally reduces subjectivity in the final products. It also can greatly reduce sampling requirements, as, theoretically, a smaller sample would be required to retrain an existing model to new areas. Evaluating this aspect for surficial geology mapping is planned for future work. However, in this research, the objective was to first determine that CNNs can perform as well or better than other machine learning methods currently in practice and that a model can be spatially extended over short distances. This is referred to as within landscape extension and is a key requirement for operational implementation over large regions.

5. Conclusions

CNNs are an interesting advancement in machine learning combining spectral and spatial properties, feature optimization for the specific classification task, and the ability to adapt pre-trained models to new tasks. However, understanding CNN performance for moderate resolution remote sensing classification has not been widely undertaken. In this research, we assessed CNNs for surficial geology mapping. A surficial materials classification map generated using a CNN could be considered as a first iteration in map production followed by geological expert refinement in a recursive process. Accuracies in this analysis were ~77% for holdout samples and 64% for extended i.e., model trained using samples from one area and applied on other adjacent areas. We see these results as reasonable for the more difficult case of independent testing area and the challenges associated with using existing surficial geology maps as training reference. We do believe that the presented mapping approach is worth improving by building an accurate reference database and testing different configurations. Future research will seek to improve training data to better evaluate accuracy for various network architectures. The study indicates that use of CNNs would improve remote predictive mapping as an effective tool for remote regions.

Author Contributions: Rasim Latifovic and Darren Pouliot conceived and designed the experiments; Janet Campbell provided expertise in surficial geology mapping, reference data and contributed to analysis of the results; Rasim Latifovic and Darren Pouliot prepared the manuscript with valuable input from Janet Campbell.

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

References

1. Harris, J. *Remote Predictive Mapping: An Aid for Northern Mapping* [1 DVD]; Open File 5643; Geological Survey of Canada: Ottawa, ON, Canada, 2008; pp. 1–4, doi:10.4095/226008.
2. Morris, K.P. Automatic Detection of Three-Dimensional Geological Features from Remotely Sensed Imagery and Digital Terrain Models. In *Remote Sensing: An Operational Technology for the Mining and Petroleum Industries*; Springer: Dordrecht, The Netherlands, 1990; pp. 59–74, doi:10.1007/978-94-010-9744, ISBN 978-94-010-9744-4.
3. Behnia, P.; Harris, R.J.; Rainbird, H.R.; Williamson, C.M.; Sheshpari, M. Remote predictive mapping of bedrock geology using image classification of Landsat and SPOT data, western Minto Inlier, Victoria

Island, Northwest Territories, Canada. *Int. J. Remote Sens.* **2012**, *33*, 6876–6903. doi:10.1080/01431161.2012.693219.

4. Grunsky, E.; Harris, J.R.; McMartin, I. *Predictive Mapping of Surficial Materials, Schultz Lake Area (NTS 66A), Nunavut, Canada* [1 DVD]; Open File 5153; Geological Survey of Canada: Ottawa, ON, Canada, 2006; p. 49, doi:10.4095/221968.

5. Brown, O.; Harris, J.R.; Utting, D.; Little, E.C. *Remote Predictive Mapping of Surficial Materials on Northern Baffin Island: Developing and Testing Techniques Using Landsat TM and Digital Elevation Data*; Current Research 2007-B1; Geological Survey of Canada: Ottawa, ON, Canada, 2007.

6. LaRocque, A.; Leblon, B.; Harris, J.R.; Jefferson, C.W.; Tschirhart, V.; Shelat, Y. Surficial materials mapping in Nunavut, Canada, with multi-beam RADARSAT-2 dual-polarization C-HH and C-HV, Landsat-7 ETM+, and DEM data. *Can. J. Remote Sens.* **2012**, *38*, 281–305.

7. Byatt, J.; LaRocque, A.; Leblon, B.; Harris, J.; McMartin, I. 2017. Mapping Surficial Materials in Nunavut Using RADARSAT-2 C-HH and C-HV, Landsat-8 OLI, DEM and Slope Data;   unpublished Master of Science Forestry thesis, University of New Brunswick, Fredericton, NB, 97p. https://unbscholar.lib.unb.ca/islandora/object/unbscholar%3A8312 (accessed on 14 February 2018).

8. Stevens, C.W.; Kerr, D.E.; Wolfe, S.A.; Eagles, S. *Predictive Surficial Material and Surficial Geology Derived from LANDSAT 7, Hearne Lake, NTS 85I, Northwest Territories*; Open File 7233; Geological Survey of Canada: Ottawa, ON, Canada, 2013; p. 22, doi:10.4095/292394.

9. Olthof, I.; Kerr, D.; Wolfe, S.; Eagles, S. *Predictive Surficial Materials and Surficial Geology from LANDSAT-7, Upper Carp Lake, NTS 85-P, Northwest Territories*; Geological Survey of Open File 7601; Natural Resources Canada: Ottawa, ON, Canada, 2014; p. 21, doi:10.4095/293970.

10. Campbell, J.; Harris, J.; Huntley, D.; McMartin, I.; Wityk, U.; Dredge, L.A.; Eagles, S. *Remote Predictive Mapping of Surficial Earth Materials: Wager Bay North Area, Nunavut*; Open File 7118; Geological Survey of Canada: Ottawa, ON, Canada, 2013; 42p. doi:10.4095/293158.

11. Breimann, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. doi:10.1023/A:1010933404324.

12. Harris, J.; He, J.; Grunsky, E.; Gorodetzky, D.; Brown, N. A robust, cross validation classification method (RCM) for improved mapping accuracy and confidence metrics. *Can. J. Remote Sens.* **2012**, *38*, 69–90.

13. MIT Technology Review. 2013. Available online: https://www.technologyreview.com/lists/technologies/2013/ (accessed on 14 February 2018).

14. Krizhevsky, A.; Sutskever, I.; Hinto, G. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1106—1114.

15. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385. Available online: https://arxiv.org/abs/1512.03385 (accessed on 14 February 2018).

16. Castelluccio, M.; Poggi, G.; Sansone, C.; Verdoliva, L. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. Computer Vision and Pattern Recognition (cs.CV). *arXiv* **2015**, arXiv:1508.00092.

17. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444, doi:10.1038/nature14539.

18. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Handwritten Digit Recognition with a Back-Propagation Network. In Proceedings of the Advances in Neural Information Processing Systems (NIPS 1989), Denver, CO, USA, 27–30 November 1989; Touretzky, D., Ed.; Morgan Kaufmann: Burlington, MA, USA, 1990; Volume 2.

19. Hinton, G.; Deng, L.; Yu, D.; Dahl, E.G.; Abdel-rahman, M.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, N.T.; et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Process. Mag.* **2012**, *29*, 82–97.

20. Glorot, X.; Bordes, A.; Bengio, Y. Deep sparse rectifier neural networks. In Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, Lauderdale, FL, USA, 11–13 April 2011; pp. 315–323.

21. Pehrsson, S.J.; Currie, M.; Ashton, K.E.; Harper, C.T.; Paul, D.; Pana, D.; Berman, R.G.; Bostock, H.; Corkery, T.; Jefferson, C.W.; et al. *Bedrock Geology Compilation South Rae and Western Hearne Provinces, Churchill Province, Northwest Territories, Saskatchewan, Nunavut, Manitoba, and Alberta*; Open File 5744, Scale 1:550,000; Geological Survey of Canada: Ottawa, ON, Canada, 2014, doi:10.4095/292232.

22. Ecoregions Working Group. *Ecoclimatic Regions of Canada, First Approximation*; Ecological Land Classification Series, No. 23, Report with Map at 1:7.5 Million Scale; Sustainable Development Branch, Conservation and Protection, Environment Canada: Ottawa, ON, Canada, 1989; 199p.

23. Campbell, J.E.; Eagles, S. *Report of 2014 Activities for the Geologic and Metallogenic Framework of the South Rae Craton, Southeast Northwest Territories: Reconnaissance Surficial and Bedrock Fieldwork in the GEM 2 South Rae Project Area*; Open File 7701; Geological Survey of Canada: Ottawa, ON, Canada, 2014; 7p. doi:10.4095/295463.

24. Pehrsson, S.; Campbell, J.; Martel, E.; McCurdy, M.; Agosta-Gongora, P.; Theissen, E.; Jamieson, D.; Lauzon, G.; Buller, G.; Falck, H.; et al. *Report of 2015 Activities for the Geologic and Metallogenic Framework of the South Rae Craton, Southeast Northwest Territories: GEM 2 South Rae Quaternary and bedrock Project*; Open File 7958, Geological Survey of Canada: Ottawa, ON, Canada, 2015, 24p. doi:10.4095/297387.

25. Campbell, J.E.; Lauzon, G.; Dyke, A.; Haiblen, A. *Report of 2016 Activities for the Regional Surficial Geological Mapping of the South Rae Craton, Southeast NWT: GEM 2 South Rae Quaternary and Bedrock Project*; Open File 8143; Geological Survey of Canada: Ottawa, ON, Canada, 2016; 13p. doi:10.4095/299391.

26. Latifovic, R.; Pouliot, D.; Sun, L.; Schwarz, J.; Parkinson, W. *Moderate Resolution Time Series Data Management and Analysis: Automated Large Area Mosaicking and Quality Contro*; Geomatics Canada, Open File 6; Natural Resources Canada: Ottawa, ON, Canada, 2015; 25p. doi:10.4095/296204.

27. National Forestry Database. 2016. Available online: http://nfdp.ccfm.org (accessed on February 14, 2018).

28. Gao, B. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **1996**, *58*, 257–266.

29. Jensen, J.R. *Introductory Digital Image Processing: A Remote Sensing Perspective*, 3rd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2005; 517p.

30. Huang, C.; Wylie, B.; Yang, L.; Homer, C.; Zylstra, G. Derivation of a tasselled cap transformation based on Landsat 7 at-satellite reflectance. *Int. J. Remote Sens.* **2002**, *23*, 1741–1748.

31. Crist, E.P. A TM Tasseled Cap Equivalent Transformation for Reflectance Factor Data. *Remote Sens. Environ.* **1985**, *17*, 301–306.

32. National Air Photo Library, Natural Resources Canada. Available online: http://www.nrcan.gc.ca/earth-sciences/geomatics/satellite-imagery-air-photos/9265 (accessed on 14 February 2018).

33. Polar Geospatial Center. Available online: https://www.pgc.umn.edu/data/arcticdem/ (accessed on 14 February 2018).

34. Dyke, A.S.; Campbell, J.E. *Surficial Geology, Abitau Lake, Northwest Territories, NTS 75-B*; Canadian Geoscience Map 350 (Preliminary), Scale 1:100,000; Geological Survey of Canada: Ottawa, ON, Canada, 2018 in press.

35. Open Source Computer Vision Library Available online: http://opencv.org (accessed on 14 February 2018).

36. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. doi:10.1109/mgrs.2017.2762307.

37. Krizhevsky, A. *Learning Multiple Layers of Features from Tiny Images*; Technical Report; University of Toronto: Oronto, ON, Canada, 2009; pp. 1–60.

38. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252, doi:10.1007/s11263-015-0816-y.