Defence Research and Development Canada
Recherche et développement pour la défense Canada

CAN UNCLASSIFIED

DRDC | RDDC
technologysciencetechnologie

Measuring the effectiveness of different types of imagery and image-derived products in land-cover classification

An approach based on Shapley values and game theory

Sacha D. Nandlall
DRDC – Ottawa Research Centre

Koreen Millard
Environment and Climate Change Canada

Defence Research and Development Canada
Scientific Report
DRDC-RDDC-2018-R301
January 2019

CAN UNCLASSIFIED

Canada

IMPORTANT INFORMATIVE STATEMENTS

This document was reviewed for Controlled Goods by DRDC using the Schedule to the *Defence Production Act.*

Endorsement statement: This publication has been peer-reviewed and published by the Editorial Office of Defence Research and Development Canada, an agency of the Department of National Defence of Canada. Inquiries can be sent to: Publications.DRDC-RDDC@drdc-rddc.gc.ca.

Abstract

Land-cover mapping consists of determining the type and usage of particular tracts of land, and is often accomplished with remote sensing and classifiers. The maps generated with land-cover analysis are used for applications that include space-based Intelligence, Surveillance and Reconnaissance (ISR); Geospatial Intelligence (GEOINT); and Intelligence Preparation of the Operational Environment (IPOE).

Land-cover classifiers frequently employ different types of input data, such as imagery obtained from several types of sensors—including Synthetic Aperture Radar (SAR), LiDAR, and optical satellites—as well as ancillary datasets such as Digital Elevation Models (DEMs). However, it can be challenging to determine which inputs have the greatest impact on the accuracy of the classifier, as well as assess how important each input is relative to the others.

In this work, a method of quantifying the relative importance of each input is developed and demonstrated using previously developed land-cover classifiers. The proposed method employs concepts from game theory and relies on the Shapley value, which provides a quantitative assessment of each input's importance in terms of its average contribution to the accuracy of the classifier. The approach described herein thus provides a robust method of determining which types of images and image-derived products are most important in classifying terrain.

Significance for defence and security

Land-cover maps are used for applications that include space-based Intelligence, Surveillance and Reconnaissance (ISR); Geospatial Intelligence (GEOINT); and Intelligence Preparation of the Operational Environment (IPOE). Potential benefits of this work for the Canadian Armed Forces (CAF) include increased quality and lower cost of land-cover maps in these applications. This work contributes to improving the design of the classifiers that generate the maps, as well as optimizing the selection of input data for these classifiers while maintaining a sufficiently high performance.

Résumé

La cartographie de la couverture terrestre consiste à déterminer le type et l'utilisation de différentes parcelles de terrain. Elle est souvent réalisée avec la télédétection et des classificateurs. Les cartes générées avec l'analyse de la couverture terrestre sont utilisées pour des applications comprenant le renseignement, la surveillance et la reconnaissance spatiales (RSR) ; l'intelligence géospatiale (GEOINT) ; et la préparation du renseignement de l'environnement opérationnel (PREO).

Les classificateurs de couverture terrestre utilisent souvent différents types de données d'entrée, telles que des images obtenues à partir de plusieurs types de capteurs – y compris le radar à synthèse d'ouverture (RSO), le LiDAR et les satellites optiques – ainsi que des données auxiliaires tels que les modèles altimétriques numériques (*DEM*). Cependant, il est parfois difficile de déterminer quelles données d'entrée ont le plus grand impact sur l'exactitude des résultats, ou d'évaluer l'importance de tous les données d'entrée les uns par rapport aux autres.

Dans ce travail, une méthode pour quantifier l'importance relative de chaque donnée d'entrée est développée et démontrée à l'aide de classificateurs de couverture terrestre développés précédemment. La méthode proposée emploie des concepts de la théorie des jeux et utilise sur la valeur de Shapley, qui fournit une évaluation quantitative de l'importance de chaque donnée d'entrée en termes de sa contribution moyenne à l'exactitude du classeur. L'approche décrite dans ce document donne une méthode robuste pour déterminer quels types d'images et quels produits dérivés d'images sont les plus importants pour la classification du terrain.

Importance pour la défense et la sécurité

Les cartes de couverture terrestre sont utilisées pour des applications dans le renseignement, la surveillance et la reconnaissance spatiale (RSR) ; l'intelligence géospatiale (GEOINT) ; et la préparation du renseignement de l'environnement opérationnel (PREO). Les avantages potentiels de ce travail pour les forces armées canadiennes (FAC) incluent une qualité accrue et une réduction du coût des cartes d'occupation du sol dans ces applications. Ce travail contribue à l'amélioration de la conception des classeurs automatiques qui génèrent les cartes, ainsi qu'à l'optimisation de la sélection des variables pour ces applications, tout en maintenant une performance suffisamment élevée.

Table of contents

List of acronyms

CAF	Canadian Armed Forces
CDED	Canada Digital Elevation Data
CSA	Canadian Space Agency
DEM	Digital Elevation Model
GEOINT	Geospatial Intelligence
IPOE	Intelligence Preparation of the Operational Environment
ISR	Intelligence, Surveillance and Reconnaissance
MDA	Mean Decrease in Accuracy
MDA Inc.	MacDonald, Dettwiler and Associates Incorporated
MDG	Mean Decrease in Gini
SAR	Synthetic Aperture Radar
SRTM	Shuttle Radar Topography Mission

Acknowledgements

The authors would like to thank Stephanie Ahonzo (University of Ottawa), Sarah Banks (Environment and Climate Change Canada), and Katerina Biron (Defence Research and Development Canada) for their assistance. Murray Richardson (Carleton University) and the South Nation Conservation Authority provided funding for training data collection and image processing. LiDAR data was provided by the United Counties of Prescott and Russell. Synthetic Aperture Radar (SAR) data was acquired by MacDonald, Dettwiler and Associates (MDA Inc.) and provided through the Canadian Space Agency (CSA).

This page intentionally left blank.

1 Introduction

Classification problems in remote sensing often involve using multiple kinds of imagery and image-derived products to perform categorization tasks on portions of the image. One common classification problem is land-cover mapping, which consists of determining the type and usage of particular tracts of land [1]. Land-cover maps are used for applications in space-based Intelligence, Surveillance and Reconnaissance (ISR); Geospatial Intelligence (GEOINT); and Intelligence Preparation of the Operational Environment (IPOE) [2].

To classify terrain in the context of land-cover mapping with remote sensing, a classifier might use different types of input data, such as imagery obtained from several types of sensors—including Synthetic Aperture Radar (SAR), LiDAR, and optical satellites—as well as ancillary datasets such as Digital Elevation Models (DEMs). These inputs are henceforth referred to as "variables" [1, 3, 4, 5]; in the context of classification, these are also called "features" in other works [6, 7, 8].

Classification is commonly accomplished using machine-learning algorithms (e.g., random forest, support vector machines, neural networks). While these algorithms are capable of handling high-dimensional data (i.e., with hundreds or thousands of variables), it has been demonstrated that reducing the number of variables in land-cover classifiers is necessary to reduce noise and overfitting, as well as increase computational speed [1]. Because some sensors and data processing functions may result in a more accurate classification under different classification scenarios, it is important to find the optimal set of variables in each case.

Several criteria are available to determine which variables are the most important in a particular classification scenario. In random forest classification, importance values are most commonly assessed using Mean Decrease in Accuracy (MDA) or Mean Decrease in Gini (MDG) [9]. While these metrics provide some measure of relative importance, neither the MDA nor MDG accounts for correlations between variables [10], or addresses cases in which the usefulness of one variable depends on whether or not another variable is present. Other methods such as Principal Component Analysis (PCA) are available to perform feature selection [6, 7, 8], but these do not necessarily reduce the dimensionality of the dataset or provide guidance as to why particular features are more or less important than others.

This work presents a method of quantifying the relative importance of different variables (i.e., different image types and image-derived products) in the context of land-cover mapping. The proposed method relies on a measure known as the Shapley value and employs concepts from game theory. The Shapley value provides a quantitative assessment of the importance of each variable in terms of the average contribution it makes to classification accuracy.

The remainder of this manuscript is structured as follows:

- Section 2 provides a technical description of the Shapley value, including the concepts from game theory that are needed to define it as well as the equations that are required to calculate it.

- Section 3 applies the Shapley value method to several previously published land-cover classifiers [1, 3, 4, 5], in each case quantifying the influence of different image types and image-derived products on classification accuracy.
- Section 4 contains some additional discussion on the avantages, limitations, and practical applications of the Shapley value method.
- Section 5 provides a short conclusion and summary of findings.

2 The Shapley Value Method

The approach employed in this manuscript is based on treating each variables as a "player" in a collaborative game, in which the objective is to maximize the accuracy of the classifier. The contribution of each variable to the objective is quantified using the Shapley value, a metric proposed in 1953 by mathematician Lloyd Shapley in the context of cooperative game theory [11, 12]. In this section, the definition of the Shapley value as used in this work is developed, along with the relevant concepts from game theory needed to define it.

2.1 Input profiles

Consider a cooperative game involving a number p of players, numbered 1 through p. Suppose that each player can either choose to participate in the game or not. This participation information is called an "input profile," and may be represented with a vector containing p elements, where a given element is "0" if the corresponding player does not participate in the game and is "1" if they do.

For a given input profile $\vec{x}$, the number of participating players is denoted as $|\vec{x}|$. Additionally, the following special input profiles are defined:

- The null input profile, denoted $\vec{0}$, corresponds to no players participating in the game, and is represented by a vector of zeros.
- Single-participant input profiles are denoted $\vec{e_k}$, and represent the input profile where only player k participates in the game (thus, by definition, $|\vec{e_k}| = 1$).
- The all-participant input profile, denoted $\vec{1}$, corresponds to all players participating in the game, and is represented by a vector of ones.

The universe of input profiles, denoted T, represent the set of all possible input profiles—i.e., all possible combinations of participants—for the game being played. For example, for a three-player game, the universe is

$$\begin{aligned} T \;=\; & \{\{0,0,0\};\{0,0,1\};\{0,1,0\};\{0,1,1\}; \\ & \{1,0,0\};\{1,0,1\};\{1,1,0\};\{1,1,1\}\}. \end{aligned} \tag{1}$$

In general, a game with p players will have 2^p input profiles in its universe. For the previous example, $p = 3$, and so the universe contains $2^3 = 8$ input profiles.

2.2 Characteristic function

Let $g(\vec{x})$ represent the outcome of the cooperative game for an input profile $\vec{x}$. Here it is assumed for simplicity that the game outcome can always be represented as a single, non-negative number called the "score" (i.e., the score may be positive or zero), and that higher scores correspond to more desirable outcomes. The Shapley value method also assumes that the score resulting from no players participating in the game results in a score of 0, which means that the function g must satisfy $g(\vec{0}) = 0$. Under these assumptions, the function g is referred to as the "characteristic function" of the game.

2.3 Marginal contribution

For an input profile $\vec{x}$ in which a player k is marked as a non-participant (i.e., for which $x_k = 0$), the marginal contribution made by player k to the profile $\vec{x}$ is defined as

$$m(\vec{x}, k) = g(\vec{x} + \vec{e_k}) - g(\vec{x}). \tag{2}$$

That is, the marginal contribution is the difference between the score for the profile $\vec{x}$ if player k were added as a participant (accomplished by adding the single-participant profile $\vec{e_k}$ to the vector $\vec{x}$), minus the score for the actual profile $\vec{x}$ (which does not have player k participating). This may be thought of as the "value added" by player k for the specific input profile $\vec{x}$. Note that the marginal contribution of a player may be negative if adding them as a participant results in a lower score than when they do not participate.

2.4 Shapley set

For a given player k in a game with universe T, define the Shapley set $Q(T, k)$ to be the set of all input profiles in the universe of the game for which the player is marked as a non-participant. As an example, in a three-player game (the universe of which is given in Equation 1), the Shapley set for the second player would be

$$Q(T, 2) = \{\{0, 0, 0\}; \{0, 0, 1\}; \{1, 0, 0\}; \{1, 0, 1\}\}. \tag{3}$$

2.5 Shapley value

Using the aforementioned concepts and definitions, the Shapley value of a player k in a cooperative game with p players and a universe of T is defined mathematically as follows:

$$S(k) = \sum_{\vec{x} \in Q(T,k)} \frac{(|\vec{x}|)!(p - |\vec{x}| - 1)!}{p!} m(\vec{x}, k). \tag{4}$$

Note that the units of the Shapley value are the same as those of the game's characteristic function, as defined in Section 2.2. For the present application, this function would correspond to the accuracy of the classifier on its testing data.

The interpretation of Equation 4 is that the Shapley value considers all of the ways that a given player could contribute to the game by adding together all of the possible marginal contributions for that player using appropriate weighting coefficients. From the perspective of probability theory, the Shapley value can be thought of as the "expected value" of a player's contribution to the game. Scenarios in which the player reduces the score of the game by participating are accounted for in the calculation, since these correspond to negative marginal contributions.

2.6 Advantages and desirable properties of the Shapley value

The Shapley value has a number of properties that are desirable for a fair and just metric of player contribution.

- **Non-discrimination:** Players with identical contributions to the game will also have identical Shapley values (i.e., the labels or ordering of the players does not matter). In the case of land-cover classification, this property ensures that the naming or ordering of the variables in the dataset does not influence the importance attributed to them by the method.

- **Efficiency:** The Shapley value distributes the total score across all players in the game, in that the sum of the Shapley values of all players is equal to the score that occurs when all players participate:

$$\sum_{k=1}^{p} S(k) = g(\vec{1}). \tag{5}$$

  In classification, the highest accuracy is typically achieved when all available variables are used. This property therefore allows the accuracy of each variable to be expressed in terms of its contribution towards this maximum value. Note that this property does not hold for partial sums of Shapley values (i.e., the sum of the Shapley values for a subset of the players would not necessarily equal the characteristic function for the input profile containing only those players).

- **Marginality:** Players that contribute more to the game are assigned higher Shapley values, and a player that does not contribute will be assigned a Shapley value of zero. For the present application, this property ensures that variables that contribute more to overall accuracy will receive proportionally higher Shapley values, and variables that make no contribution will receive a value of zero.

It has been proven that the Shapley value is the only metric of player contribution that satisfies all three of the above properties [11, 12].

3 Application to Sensor Selection in Land-Cover Classifiers

In this section, the practical application of the Shapley value is demonstrated for three example use cases involving land-cover classification based on images acquired with different types of remote sensors (radar, optical, LiDAR, etc.) as well as image-derived products such as DEMs.

In terms of the game theoretic concepts described in Section 2, in each of these cases the objective of the "game" is to maximize land-cover classification accuracy, and the "players" of the game are the different inputs (images and image-derived products) available to the classifier. The objective of the Shapley value analysis is to assess the individual contribution of each input to the accuracy of the land-cover classifier, and thus provide a quantitative, relative measure of the importance of each input to the classification task at hand.

The three example use cases discussed in this section are described below.

- Section 3.1 examines the overall accuracy of a land-cover classifier developed by White *et al.* [3] and used over Alfred Bog. Three input variables are considered: SAR imagery, optical imagery, and a Digital Elevation Model (DEM).
- Section 3.2 studies the performance of a land-cover classifier developed using data from Millard and Richardson [1] and from Behnamian *et al.* [4], also from Alfred Bog. Three input variables are considered: SAR imagery, optical imagery, and LiDAR. Accuracy is considered both overall and for each of three land-cover classes.
- Section 3.3 examines the performance of the land-cover classifier developed by Banks *et al.* [5]. The two input variables considered are SAR imagery and optical imagery, and accuracy is assessed both overall and for each of seven land-cover classes.

In all three cases, random forest classifiers were used to perform land-cover classification.

3.1 Alfred Bog with SAR, optical imagery, and DEM data (White *et al.*, 2017)

In this first example use case, the individual contributions of SAR imagery, optical imagery, and DEM data to land-cover classification accuracy are quantified in the context of a study performed by White *et al.* [3]. The area studied is Alfred Bog, a boreal peatland complex with an area of over 10,000 acres (40 km^2) located near Alfred, ON, Canada (see Figure 1 of White *et al.* [3]).

Classifier design and accuracy

White *et al.* employed a supervised random forest classifier, with 330 training points and 1000 trees generated for each model. The land-cover classes employed by White *et al.* were taken from a previous study by Millard and Richardson [1]. In total, five different land-cover classes were distinguished (agriculture, forest, fen, open bog, and treed bog); however, for the purposes of this example, only the overall accuracy of the classifier across all five terrain classes is considered.

Table 1 reproduces the overall accuracy achieved by the White *et al.* land-cover classifier on its testing data for all combinations of the following three variables:

- A Synthetic Aperture Radar (SAR) image of the area from RADARSAT-2 collected on May 2, 2014 in the Wide Fine Quad 1 (FQ1W) mode;
- An optical image of the area from Landsat-8 collected on April 24, 2014 and ordered as a Landsat surface reflectance product; and
- DEM data obtained from the Shuttle Radar Topography Mission (SRTM).

Since there are three inputs in this case, a total of eight combinations are possible. As stated in Section 2.2, the accuracy achieved when none of the inputs are used (i.e., for the null input profile) is taken to be zero. This corresponds to assuming that most of the variation in the output classes is explained by the inputs, and that random chance does not play a significant role in the performance of the classifier (see, however, the discussion in Section 4 for ways to compensate for the effect of random chance when this is not the case). The overall accuracy values for the remaining seven combinations—i.e., those that include at least one input—are listed in Table 1.

Table 1: *Overall accuracy achieved by the land-cover classifier reported in Table A1 of White et al. [3] on the testing dataset for different combinations of SAR, optical, and DEM input data. The case where none of the inputs are used is assumed to yield an accuracy of zero per Section 2.2.*

Input variables used	Accuracy (%)
DEM only	78
Optical only	50
Optical + DEM	80
SAR only	49
SAR + DEM	71
SAR + Optical	59
SAR + Optical + DEM	82

A flowchart illustrating the Shapley method as applied to this example is presented in Figure 1.

Figure 1: *Flowchart illustrating the Shapley method as applied to the example use case of Section 3.1. Here, three input variables ($N = 3$) are available to perform land-cover classification: SAR imagery, optical imagery, and a DEM. A classifier is then trained, validated, and tested for each possible combination of these inputs, except for the case of no inputs; hence, a total of $2^N - 1 = 7$ classifiers are developed. Finally, the overall accuracy values for the classifiers are used to compute the Shapley value of each input variable with Equation 4.*

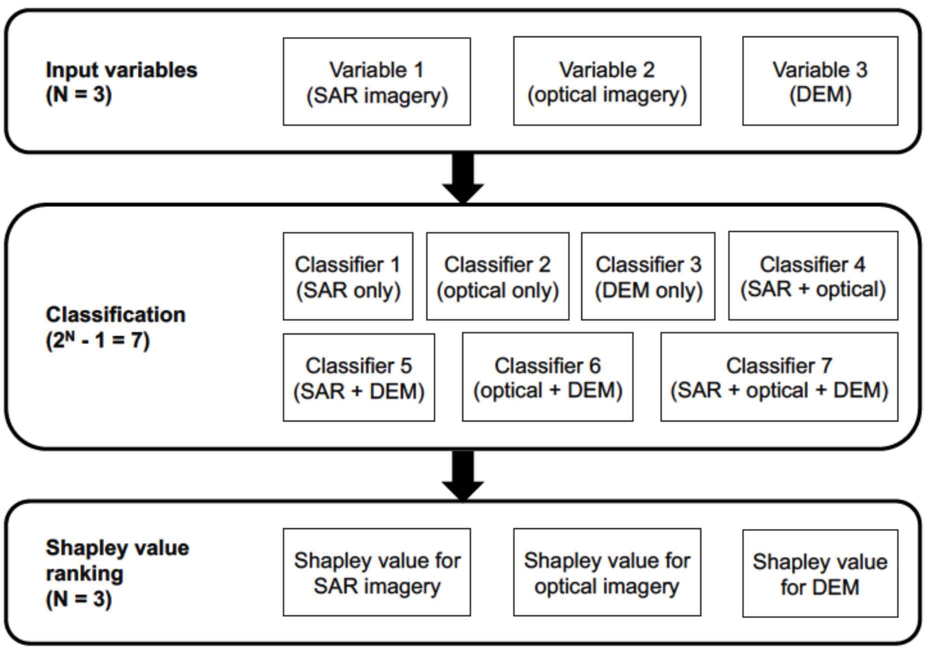

Table 2: *Shapley values (contribution to overall accuracy) for SAR imagery, optical imagery, and the DEM, based on the values reported in Table 1.*

Input variable	Shapley value (%)
SAR imagery	10.4
Optical imagery	35.4
DEM	36.3

Shapley value analysis and interpretation

Using the data in Table 1, the Shapley values for each of the three inputs (SAR, optical, and DEM) were computed using Equation 4. The results are reported in Table 2.

As mentioned in Section 2.5, the Shapley values represent the average contribution of each input variable to the accuracy of the classifier, when all possible combinations of variables are considered. Hence, the first row of Table 2 shows that using the SAR imagery in the classifier increases the overall accuracy by 10.4%; similarly, incorporating the optical imagery adds 35.3% to overall accuracy, while adding the DEM contributes 36.3% to overall accuracy. Additionally, as expected per the efficiency property of the Shapley value stated in Equation 5, the sum of all three Shapley values in Table 2 is 82.1%, which (within rounding error) matches the overall accuracy of 82% reported in Table 1 for when all three input variables are used.

The Shapley values in Table 2 show that the optical imagery from Landsat-8 as well as the SRTM DEM data each contribute over three times more to classification accuracy than the SAR image from RADARSAT-2, although the SAR image does make a positive contribution to accuracy as well. This is in agreement with the conclusions of White *et al.*, who reported that "while models with just Landsat-8 and SRTM data were able to achieve acceptable accuracies, the addition of SAR variables did increase overall accuracy" [3]. The Shapley value analysis presented here thus provides a quantifiable metric—namely, the anticipated contribution of each input variable to classifier accuracy—that supports this qualitative conclusion.

3.2 Alfred Bog with SAR, optical imagery, and LiDAR (Millard and Richardson, 2015)

In this second example use case, the individual contributions of SAR imagery, optical imagery, and LiDAR data to land-cover classification accuracy are evaluated both overall and by land-cover class. The location studied is Alfred Bog, the same area analyzed in the previous example (Section 3.1).

Classifier design and accuracy

The training data used herein were the same as the data used by Millard and Richardson [1] and by Behnamian *et al.* [4]. However, for the purposes of simplifying the present example, the seven land-cover classes were merged into three larger classes: agriculture, forest, and wetland. A set of 500 randomly located pixels (with a minimum point spacing of 8 m) distributed throughout the study area were selected, resulting in a training dataset with low spatial autocorrelation (Moran's I of 0.11) [1]. In this work, all 500 points were used in training, and the accuracy after bootstrap aggregation is reported (i.e., out-of-bag accuracy). Full details on training data extraction are described in Section 2.2 of Millard and Richardson [1]. A random forest classifier was trained on the aforementioned dataset; for all of the classifications, the randomForest package in R 3.5.1 software was used with 10,000 trees and *mtry* set to the square root of the number of predictor variables.

Table 3 shows the accuracy achieved by the Millard and Richardson land-cover classifier for all seven combinations of the following three inputs (excluding the null input profile):

- A Synthetic Aperture Radar (SAR) image of the area from RADARSAT-2 image collected on May 2, 2014 in the Fine Quad 1 (FQ1) mode;
- An optical image of the area from Landsat-8 collected on April 24, 2014 and ordered as a Landsat surface reflectance product; and
- A LiDAR dataset collected on May 14, 2014 using point cloud upscaled to 8 m, with a total of 28 LiDAR terrain and vegetation derivatives computed (full details on the LiDAR dataset collection, processing, and derivatives are given in Millard and Richardson [1]).

Table 3: *Accuracy achieved by the land-cover classifier on the testing dataset used by Millard and Richardson [1] and by Behnamian et al. [4] for different combinations of SAR, optical, and LiDAR input data.*

Land-cover type	Accuracy type	Accuracy (%) by input data used (S = SAR, O = optical, L = LiDAR)						
		L	O	L+O	S	S+L	S+O	S+O+L
Agriculture	User's	83.6	48.5	86.2	92.0	97.1	92.8	97.1
	Producer's	88.2	65.1	87.1	91.3	94.8	90.4	95.1
Forest	User's	70.1	46.3	69.4	55.2	78.4	55.2	75.4
	Producer's	76.4	58.5	79.5	66.7	84.7	67.9	83.5
Wetland	User's	87.5	77.6	87.1	94.9	94.7	94.5	95.1
	Producer's	82.2	61.9	83.5	91.1	94.5	91.6	94.1
All	Overall	83.7	62.4	84.4	88.5	93.4	88.6	93.2

In additional to overall accuracy, Table 3 also provides both user's and producer's accuracy for each class of land-cover. The user's accuracy refers to how accurate the classifier is when it determines that the land-cover is of a particular type, and thus provides an indication of what accuracy a user might expect when the classifier is given new input data and categorizes it as being of that type; in other applications, this is called the precision or the "positive predictive value" of the classifier. On the other hand, the producer's accuracy indicates how well the classifier performed for a particular land-cover type in the test data set; in other applications, this is called the sensitivity or the "true positive rate" of the classifier. The relevance of each accuracy type depends on the intended usage of the classifier: if the goal is to classify the testing dataset as accurately as possible, the producer's accuracy is most relevant; on the other hand, if the goal is to classify images that are similar to (but not part of) the testing set, then the user's accuracy is of concern.

Shapley value analysis and interpretation

Using the data in Table 3, the Shapley values for each of the three inputs (SAR, optical, and LiDAR) were computed using Equation 4. The results are reported in Table 4.

Table 4: *Shapley values (contribution to accuracy) for SAR, optical, and LiDAR input data, based on the values reported in Table 3. The highest Shapley value in each row is highlighted in bold.*

Land-cover type	Accuracy type	Shapley value (%)		
		SAR	Optical	LiDAR
Agriculture	User's	18.7	**40.1**	38.3
	Producer's	22.9	35.6	**36.6**
Forest	User's	9.6	32.7	**33.1**
	Producer's	15.5	**35.0**	32.9
Wetland	User's	28.5	33.1	**33.5**
	Producer's	23.3	**35.8**	34.9
All	Overall	22.2	**35.7**	35.3

The Shapley values in Table 4 show that optical imagery and LiDAR data contribute more to accuracy than SAR imagery in all cases. However, the input data source with the highest Shapley value varies according to the terrain class as well as the intended usage of the classifier. For example, if the classifier is to be used for classifying agricultural land-cover in new images as accurately as possible, optical imagery is ranked slightly higher than LiDAR, since it contributes 1.8% more to user's accuracy. On the other hand, if the goal is to classify agricultural land as accurately as possible within the existing dataset and only one input can be chosen, LiDAR contributes 1% more to producer's accuracy than the optical imagery. Note that these differences are relatively small, and may not be statistically significant; however, performing an assessment of statistical significance is beyond the scope of the present work.

3.3 Arctic terrain with SAR and optical imagery (Banks *et al.*, 2017)

As a final example use case for Shapley values in the context of land-cover classification, the individual contributions of SAR imagery and optical imagery to classification accuracy are assessed in the context of a study performed by Banks *et al.* [5]. The area considered in this study spans over 40,000 km^2 and is located in the Canadian Arctic over the Northwest Passage, within the Kitikmeot region of Nunavut (see Figure 1 of Banks *et al.* [5]).

Classifier design and accuracy

Banks *et al.* employed a supervised random forest classifier, with 1000 trees generated for each model. A total of seven land-cover classes were studied (see Table 5 or 6 for the list of classes), and 250 sites per class spaced at least 100 m apart were selected for training and validation.

Table 5 reproduces the overall accuracy achieved by the land-cover classifier for all three combinations of the following two inputs (excluding the null input profile):

- Synthetic Aperture Radar (SAR) images of the area from RADARSAT-2 collected during August and September of 2014 in Wide Fine Quad Polarization mode (only one image per area was used); and
- Optical images of the area from Landsat-5 collected during August of 2009, 2010, and 2011, with atmospheric correction applied (only one image per area was used).

Shapley value analysis and interpretation

Using the data in Table 5, the Shapley values for each of the two inputs (SAR and optical) were computed using Equation 4. The results are reported in Table 6.

The Shapley values in the last row of Table 6 show that, from an overall perspective, SAR imagery from RADARSAT-2 and optical imagery from Landsat-5 make similar contributions to the overall accuracy of the classifier (45.4% and 47.2%, respectively). However, examining the Shapley values associated with the different classes in Table 6 reveals that the two types of imagery serve complementary purposes in several cases: optical imagery is the most important contributor for classifying water, bedrock, and tundra, while SAR is the most important contributor for sand/mud as well as pebble/boulder (as mentioned previously in Section 3.2, assessing the statistical significance of these differences is beyond the scope of the present work). This is in agreement with the findings of Banks *et al.*, who concluded that "optical and SAR data provide relevant and complementary information" [5]. Section 5.1 of their report provide a more detailed physical explanation of why RADARSAT-2 performs better than Landsat-5 for some classes and vice-versa for others, based on factors such as surface roughness, incidence angle, polarization, and image resolution [5].

Table 5: *Accuracy achieved by the land-cover classifier reported in Table 10 of Banks et al. [5] on the testing dataset for different combinations of SAR and optical input data.*

Land-cover type	Accuracy type	Accuracy (%) by input data used		
		SAR only	Optical only	SAR + Optical
Water	User's	97.6	94.0	98.8
	Producer's	96.4	84.8	96.5
Sand/Mud	User's	74.7	80.7	94.0
	Producer's	69.7	77.9	94.0
Mixed Sediment	User's	78.3	61.4	91.6
	Producer's	69.9	70.8	86.4
Pebble/Boulder	User's	63.9	88.0	88.0
	Producer's	79.1	88.0	93.6
Bedrock	User's	88.0	72.3	95.2
	Producer's	90.1	85.7	95.2
Wetland	User's	83.1	83.1	89.2
	Producer's	80.2	80.2	89.2
Tundra	User's	84.3	77.1	91.6
	Producer's	86.4	69.6	93.8
All	Overall	79.5	81.4	92.6

Table 6: *Shapley values (contribution to accuracy) for SAR and optical input data, based on the values reported in Table 5. The highest Shapley value in each row is highlighted in bold.*

Land-cover type	Accuracy type	Shapley value (%)	
		SAR	Optical
Water	User's	47.6	**51.2**
	Producer's	42.4	**54.1**
Sand/Mud	User's	**50.0**	44.0
	Producer's	**51.1**	42.9
Mixed Sediment	User's	37.3	**54.2**
	Producer's	**43.7**	42.7
Pebble/Boulder	User's	**56.0**	31.9
	Producer's	**51.2**	42.4
Bedrock	User's	39.8	**55.4**
	Producer's	45.4	**49.8**
Wetland	User's	**44.6**	**44.6**
	Producer's	**44.6**	**44.6**
Tundra	User's	42.2	**49.4**
	Producer's	38.5	**55.3**
All	Overall	45.4	**47.2**

4 Discussion

In the land-cover classification examples described in the previous section, the Shapley value provides a quantitative means of comparing the importances of different input variables by providing the average contribution of each variable to classification accuracy on the test data. Moreover, the Shapley value is expressed in the same units as the quantity to be maximized—in this case, land-cover classification accuracy on a scale of 0 to 100%—and thus lends itself to a relatively straightforward interpretation. The Shapley value also accounts for cases where the usefulness of one input variable depends on the presence or another variable.

One potential limitation of using the Shapley value approach for single-input selection is that the number of classifications that must be performed scales exponentially with the number of inputs, since for N inputs, $2^N - 1$ classifications are required. However, in most land-cover classification scenarios—including the three example cases in Section 3—only three or four types of imagery and image-derived products are considered (e.g., SAR, optical, LiDAR, DEM). For this number of input variables, the effect of the exponential scaling does not present a significant overhead.

As discussed in Section 2.2, the Shapley value method assumes that the null input profile (i.e., when no inputs are given to the classifier) makes no contribution to the accuracy of the classifier. This corresponds to assuming that most of the variation in the output classification may be adequately explained by the inputs, and that random chance does not play a significant role in the performance of the classifier. However, if the impact of random chance on classification accuracy is of concern—for example, if the baseline performance of the classifier is already relatively low and is close to what would be expected from random chance alone—it is possible to compensate for this effect in the Shapley method. Specifically, the null input profile may be assigned an accuracy of $1/n$ where n is the number of classes (i.e., the accuracy achieved by random chance alone). This baseline accuracy could then be subtracted from all of the accuracy values used in the computation. The Shapley values so obtained would provide a more conservative estimate of the importance of each variable, but may overestimate the influence of random chance on the results in cases where the classifier inputs are strongly correlated with the output classes.

The Shapley value only answers the question of which single player (in this case, a single image type or image-derived product) out of all of those available is the most important. However, once a player is chosen for use in classification (i.e., once the participation of that player is no longer optional), the Shapley values should be recomputed with this player definitively included (i.e., only considering input profiles that have a "1" for the included player). At that point, the second-most important player may no longer be the one that had the second-highest Shapley value previously. One instance in which this can occur is when the two players with the highest Shapley values are correlated, meaning that only one of them is actually needed by the classifier (for example, if two images provide reasonably similar data, or if one input is just a scaled version of another).

Because all of the inputs in Tables 2, 4, and 6 have positive Shapley values, each input makes a positive contribution to classifier accuracy. Hence, from the sole perspective of maximizing the accuracy of the classifier, using all of the available variables would achieve the best outcome. However, as discussed in Section 1, this may lead to overfitting, increased accuracy solely due to chance correlations, and/or a classifier that requires too many computational resources to be practical. Moreover, cost or other resource limitations may mean that only so many choices of input data are available (e.g., only one sensor between SAR or LiDAR can be employed, or cost constraints require choosing between a SAR image and an optical one). For these reasons, selecting the most important input data sources of those available is critical. As mentioned in Section 2.6, Shapley values provide a framework for making this selection in a way that is optimal and fair, in the game-theoretic sense of being non-discriminatory, efficient, and marginal in how gains in accuracy are attributed to the different input data sources.

5 Conclusion

This study demonstrates that Shapley values provide a framework for quantifying the relative importance of different sensor types (e.g., radar, optical, LiDAR) and image-derived products (such as Digital Elevation Models, DEMs) in land-cover classification problems. The approach described herein provides a robust method of determining which types of images and image-derived products are most important in classifying terrain, and provides a quantitative foundation for further investigations into the physical mechanisms that underlie the mapping process.

This page intentionally left blank.

References

[1] Millard, K. and Richardson, M. (2015), On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping, *Remote Sensing*, 7(7), 8489–8515.

[2] Secker, J. and Norris, J. (2003), Commercial Satellite Imagery Acquisition Planning System (CSIAPS): Concept Document, Defence R&D Canada – Ottawa, Technical Report, DRDC-OTTAWA-TR-2003-210.

[3] White, L., Millard, K., Banks, S. N., Richardson, M., Pasher, J., and Duffe, J. (2017), Moving to the RADARSAT Constellation Mission: Comparing Synthesized Compact Polarimetry and Dual Polarimetry Data with Fully Polarimetric RADARSAT-2 Data for Image Classification of Peatlands, *Remote Sensing*, 9(6), 1–29.

[4] Behnamian, A., Millard, K., Banks, S. N., White, L., Richardson, M., and Pasher, J. (2017), A Systematic Approach for Variable Selection With Random Forests: Achieving Stable Variable Importance Values, *IEEE Geoscience and Remote Sensing Letters*, 14(11), 1988–1992.

[5] Banks, S. N., Millard, K., Behnamian, A., White, L., Ullmann, T., Charbonneau, F., Chen, Z., Wang, H., Pasher, J., and Duffe, J. (2017), Contributions of Actual and Simulated Satellite SAR Data for Substrate Type Differentiation and Shoreline Mapping in the Canadian Arctic, *Remote Sensing*, 9(12), 1–27.

[6] Yitayew, T. G., Brekke, C., and Doulgeris, A. P. (2012), Multisensor data fusion and feature extraction for forestry applications, In *2012 IEEE International Geoscience and Remote Sensing Symposium*, pp. 4982–4985, IEEE.

[7] Mokdad, F., Bouchaffra, D., Zerrouki, N., and Touazi, A. (2015), Determination of an optimal feature selection method based on maximum Shapley value, In *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)*, pp. 116–121.

[8] Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., Kalogirou, S., and Wolff, E. (2018), Less is more: optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application, *GIScience & Remote Sensing*, 55(2), 221–242.

[9] Breiman, L. (2017), Classification and Regression Trees, London, UK: Routledge.

[10] Nicodemus, K. K., Malley, J. D., Strobl, C., and Ziegler, A. (2010), The Behaviour of Random Forest Permutation-Based Variable Importance Measures Under Predictor Correlation, *BMC Bioinformatics*, 11(1), 110.

[11] Shapley, L. S. (1953), A Value for n-person Games, *Annals of Mathematical Studies*, 28, 307–317.

[12] Roth, A. E. (1988), The Shapley Value: Essays in Honor of Lloyd S. Shapley, Cambridge, England: Cambridge University Press.

DOCUMENT CONTROL DATA

*Security markings for the title, authors, abstract and keywords must be entered when the document is sensitive

1. ORIGINATOR (Name and address of the organization preparing the document. A DRDC Centre sponsoring a contractor's report, or a tasking agency, is entered in Section 8.) DRDC – Ottawa Research Centre 3701 Carling Avenue, Ottawa ON K1A 0Z4, Canada	2a. SECURITY MARKING (Overall security marking of the document, including supplemental markings if applicable.) CAN UNCLASSIFIED 2b. CONTROLLED GOODS NON-CONTROLLED GOODS DMC A

3. TITLE (The document title and sub-title as indicated on the title page.)

Measuring the effectiveness of different types of imagery and image-derived products in land-cover classification: An approach based on Shapley values and game theory

4. AUTHORS (Last name, followed by initials – ranks, titles, etc. not to be used. Use semi-colon as delimiter)

Nandlall, S. D.; Millard, K.

5. DATE OF PUBLICATION (Month and year of publication of document.) January 2019	6a. NO. OF PAGES (Total pages, including Annexes, excluding DCD, covering and verso pages.) 26	6b. NO. OF REFS (Total cited in document.) 12

7. DOCUMENT CATEGORY (e.g., Scientific Report, Contract Report, Scientific Letter)

Scientific Report

8. SPONSORING CENTRE (The name and address of the department project or laboratory sponsoring the research and development.)

DRDC – Ottawa Research Centre
3701 Carling Avenue, Ottawa ON K1A 0Z4, Canada

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.) JFD5-05EA	9b. CONTRACT NO. (If appropriate, the applicable contract number under which the document was written.)
10a. DRDC DOCUMENT NUMBER DRDC-RDDC-2018-R301	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11a. FUTURE DISTRIBUTION WITHIN CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

11b. FUTURE DISTRIBUTION OUTSIDE CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

12. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Use semi-colon as a delimiter.)

Shapley value, game theory, remote sensing, image classification, variable selection, variable importance, random forest, land-cover mapping

13. ABSTRACT/RÉSUMÉ (When available in the document, the French version of the abstract must be included here.)

Land-cover mapping consists of determining the type and usage of particular tracts of land, and is often accomplished with remote sensing and classifiers. The maps generated with land-cover analysis are used for applications that include space-based Intelligence, Surveillance and Reconnaissance (ISR); Geospatial Intelligence (GEOINT); and Intelligence Preparation of the Operational Environment (IPOE).

Land-cover classifiers frequently employ different types of input data, such as imagery obtained from several types of sensors—including Synthetic Aperture Radar (SAR), LiDAR, and optical satellites—as well as ancillary datasets such as Digital Elevation Models (DEMs). However, it can be challenging to determine which inputs have the greatest impact on the accuracy of the classifier, as well as assess how important each input is relative to the others.

In this work, a method of quantifying the relative importance of each input is developed and demonstrated using previously developed land-cover classifiers. The proposed method employs concepts from game theory and relies on the Shapley value, which provides a quantitative assessment of each input's importance in terms of its average contribution to the accuracy of the classifier. The approach described herein thus provides a robust method of determining which types of images and image-derived products are most important in classifying terrain.

La cartographie de la couverture terrestre consiste à déterminer le type et l'utilisation de différentes parcelles de terrain. Elle est souvent réalisée avec la télédétection et des classificateurs. Les cartes générées avec l'analyse de la couverture terrestre sont utilisées pour des applications comprenant le renseignement, la surveillance et la reconnaissance spatiales (RSR); l'intelligence géospatiale (GEOINT); et la préparation du renseignement de l'environnement opérationnel (PREO).

Les classificateurs de couverture terrestre utilisent souvent différents types de données d'entrée, telles que des images obtenues à partir de plusieurs types de capteurs – y compris le radar à synthèse d'ouverture (RSO), le LiDAR et les satellites optiques – ainsi que des données auxiliaires tels que les modèles altimétriques numériques (*DEM*). Cependant, il est parfois difficile de déterminer quelles données d'entrée ont le plus grand impact sur l'exactitude des résultats, ou d'évaluer l'importance de tous les données d'entrée les uns par rapport aux autres.

Dans ce travail, une méthode pour quantifier l'importance relative de chaque donnée d'entrée est développée et démontrée à l'aide de classificateurs de couverture terrestre développés précédemment. La méthode proposée emploie des concepts de la théorie des jeux et utilise sur la valeur de Shapley, qui fournit une évaluation quantitative de l'importance de chaque donnée d'entrée en termes de sa contribution moyenne à l'exactitude du classeur. L'approche décrite dans ce document donne une méthode robuste pour déterminer quels types d'images et quels produits dérivés d'images sont les plus importants pour la classification du terrain.