Expert Systems With Applications 64 (2016) 269–286

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Improving remote sensing crop classification by argumentation-based conflict resolution in ensemble learning

Ştefan Conţiu, Adrian Groza*

Intelligent Systems Group, Department of Computer Science, Technical University of Cluj-Napoca, Baritiu 28, 400391, Cluj-Napoca, Romania

ARTICLE INFO

Article history:
Received 26 March 2016
Revised 16 June 2016
Accepted 25 July 2016
Available online 28 July 2016

Keywords:
Crop classification
Ensemble learning
Defeasible argumentation
Agricultural expert knowledge
Rule extraction

ABSTRACT

The acquisition of data through remote sensing has become of great importance in precision agriculture, as it covers large geographical areas faster and cheaper than ground inspections. The challenge is to develop technical solutions that can benefit from both huge amounts of raw data extracted from satellite images, but also from the robust amount of knowledge refined during centuries of agricultural practice. Aiming to accurately classify crops from satellite images, we developed a hybrid intelligent system that can exploit both agricultural expert knowledge and machine learning algorithms. As the crop raw data is characterized by heterogeneity, we drive our attention to ensemble learners, while expert knowledge is encapsulated within a rule-based system. Vote-based methods for solving conflicts between ensemble's base learners have difficulties in classifying exceptional cases correctly and also to give the rationale behind their decision. The conceptual research question is on *conflict resolution in ensemble learning*. To deal with debatable cases in ensemble learning and to increase transparency in such debatable decisions, our hypothesis is that argumentation could be more effective than voting-based methods. The main contribution is that voting system in ensemble learning is substituted by an argumentation-base conflict resolutor. Prospective decisions of base classifiers are presented to an argumentative system based on defeasible logic that performs dialectical reasoning on pros and cons against a classification decision. The system computes a recommendation considering both the rules extracted from base learners and the available expert knowledge. The investigated case study deals with crop classification into four classes: corn, soybean, cotton, and rice. The test site used for the experiment is an area of 20 square kilometers in the New Madrid County, southeast of the Missouri State, USA. The results show that our approach increases classification accuracy compared to the voting-based method for conflict resolution in an ensemble learner comprising of three base classifiers: a decision tree, a neural network, and a support vector machine algorithm. We also argue that combining ensemble learning and argumentation fits the decision patterns of human agents, who first collect various opinions and then perform dialectical reasoning on these opinions. We think that the people who can benefit from the conceptual instrumentation presented in this work are decision makers in domains characterized by high data availability, robust expert knowledge, and a need for justifying the rationale behind decisions.

1. Introduction

A reliable crop classification is essential for the analysis of agricultural land use in development and environment projects, for preventing and assessing climate events, or for monitoring and forecasting food security crisis. Remote sensing crop classification is seen by Li and Chung (2015) as a practice of precision agriculture, a field which uses information technology to aggregate data from multiple sources. Liaghat, Balasundram et al. (2010) has shown that the acquisition of data through remote sensing has become of great importance in precision agriculture. The main reason, also stressed by Cruz-Ramírez, Hervás-Martínez, Jurado-Expósito, and López-Granados (2012), is that it covers large geographically areas faster and cheaper than ground inspections. Usually, predictions of statistical machine learning classifiers are based entirely on the data they have seen during training. However, unseen data can hide unknown patterns, resulting in a limitation beyond which they can not extend. This limitation is specific to the context of supervised crops classification in remote sensing, where a huge number of samples exist from satellite images but only a small number of ground truth references are available for training. Symbolic processing and expert knowledge derived from crops

* Corresponding author.
E-mail addresses: Stefan.Contiu@isg.cs.utcluj.ro (Ş. Conţiu), Adrian.Groza@cs.utcluj.ro (A. Groza).

http://dx.doi.org/10.1016/j.eswa.2016.07.037

phenology and morphology profiles can potentially help the statistical classifiers perform better outside the training world.

Hybrid intelligent systems reviewed by Wozniak, Graña, and Corchado (2014) allow using both raw data and expert knowledge to offer innovative solutions to classification tasks characterized by complexity and data heterogeneity. Our aim is to develop such a hybrid intelligent system that can exploit both the expert knowledge and machine learning algorithms to classify accurately crops from satellite images. As heterogeneity characterizes the crop raw data, we drive our attention to ensemble learners, and we encapsulate expert knowledge within a rule-based system.

In this line, the conceptual research question is related to *conflict resolution in ensemble learning*. Our solution is to use argumentation systems on top of ensemble learning, to solve classification disputes. Argumentative reasoning has proved to be an efficient method to handle different perspectives on the same topic, with the additional benefits of providing justification of the decision taken and to introduce human knowledge during classification. From the machine learning perspective, various classifiers have been successfully utilized in the crop domain, such as decision trees by Friedl and Brodley (1997); Pal and Mather (2003), neural networks by Aitkenhead and Aalders (2008); Kavzoglu (2009); Kavzoglu and Mather (2003), and support vector machines by Huang, Davis, and Townshend (2002); Mountrakis, Im, and Ogole (2011). A possible solution to this selection problem is provided by ensemble learners, that have proved by Kuncheva (2004) to be more efficient than single models, especially when the correlation of the errors made by the base learners is low.

It has been argued that the focus on artificial intelligence (AI) has been shifted from knowledge representation two decades ago, to machine learning and statistical algorithms, up to recently. Shoham (2015) has noticed an intensification of efforts towards bringing some light to the machine learning black boxes using logic-based AI. Our hybrid intelligent system exploits both logic-based AI and statistical learning. In line with Shoham (2015), we argue that knowledge representation can bring valuable benefits to the black boxes within most of the learning algorithms or probabilistic-based computations. From the opposite viewpoint, it is hard to include new experiences, but also to handle non-linearity (Bahrammirzaee, 2010) as most machine learning algorithms do.

(Rahwan & Simari, 2009) and Bench-Capon and Dunne (2007) see argumentation as a mean to formalize common sense reasoning for supporting a decision when contradicting opinions may exist. In an argumentation system, instead of proofs, we have arguments. An argument consists of a coherent set of statements that supports a claim. An argument is accepted based on a dialectical analysis of arguments in favor and against the claim. As technical instrumentation, we rely on Defeasible Logic Programming (DeLP) formalized by García and Simari (2004) to perform argument based reasoning. DeLP has proved to be an effective knowledge representation and dialectical-based argumentation system for real-world applications: recommender systems by Bedi and Vashisth (2014); Briguez et al. (2014), ontology reasoning by Gómez, Chesñevar, and Simari (2013), safety assurance for unmanned aerial vehicles by Gómez, Goron, Groza, and Letia (2016), relational databases by Deagustini et al. (2013), or planning by Chow, Siu, Chan, and Chan (2013). From the knowledge perspective, ensemble learning method proposed by Xu, Yao, and Li (2015) integrates only the classification results of single classifiers. Based on argumentation technology, we aim to ensemble the classification knowledge encapsulated by each learner intending to develop an argumentation framework for classification, following the research direction opened by Amgoud and Serrurier (2008); Hao, Liu, Wu, and Yao (2015); Wardeh, Coenen, and Capon (2012b).

Our *hypothesis* is that argumentation can be used to deal with conflicts occurring in ensemble learning. In this paper, we propose a decision support system for crop classification based on satellite images. The conceptual contribution is that *voting system in ensemble learning is substituted by an argumentation-base conflict resolutor*. Prospective decisions of base classifiers are presented to an argumentative system that performs dialectical reasoning on pros and cons against a classification decision. The system computes a recommendation based on both the output from based learners and the available expert knowledge.

We think that the people who can benefit from the conceptual instrumentation presented in this work are decision makers in domains characterized by high data availability, robust expert knowledge, and a need for justifying the rationale behind decisions. We argue that our solution is close to the human cognitive model: Firstly, (Polikar, 2006) has explained that seeking additional opinions before making a decision is an innate behavior for human agents. Similarly, ensemble learning considers classfication decisions from different base learners. Secondly, argumentative-based decisions often occur in daily human tasks instead of various algebraic-based methods for opinion aggregation. Similarly, rule-based argumentation performs dialectical reasoning to decide on a winning argument. The above two observations suggest that combining ensemble learning and argumentation fits the decision patterns of human agents, both in terms of collecting opinions and dialectical reasoning on these opinions. Moreover, both ensemble learning and argumentative-based reasoning help us to minimize the risk of taken an obviously wrong decision. First, ensemble learning diminishes the risk to rely on a single inadequate base classifier. Second, by providing the dialectical tree, defeasible rule-based argumentation helps the human agent to identify reasoning flaws of the rationale behind the decision.

The investigated case study deals with crop classification into four classes: corn, soybean, cotton, and rice. As farmers have an accurate estimation on small surfaces and local regions, governments have difficulties in estimating the quality and quantity of harvesting accurately for large geographical regions. The importance of accuracy of the crop classification has been shown by Cruz-Ramírez et al. (2012), because crop classification can act as a tool for the administration: (i) to estimate crop inventory, especially in the case of negative events like excessive dry or various plant diseases. (ii) to decide whether or not to continue the subsidy.

The rest of the article is structured as follows: Section 2 browses related instrumentation developed for crop classification. Section 3 introduces the technical instrumentation used to perform argumentative reasoning. Section 4 formalizes how argumentation can solve conflicts in an ensemble learner, presents the architecture of the developed system, and the running scenario. Section 5 illustrates our approach for extracting agricultural knowledge from three classifiers (neural network, decision tree, and support vector machine) and also formalizes expert knowledge. Section 6 integrates mined knowledge with expert knowledge, applies the aggregated knowledge to increase classification accuracy of four crops and shows the experimental results. Section 7 discusses related work and possible extensions of our solution. Finally, Section 8 concludes the paper.

2. Related work on crop classification

On the one hand, various learners have been proposed for crop classification: Aitkenhead and Aalders (2008); Cruz-Ramírez et al. (2012); Kavzoglu (2009); Kavzoglu and Mather (2003) have used neural networks, Friedl and Brodley (1997); Pal and Mather (2003) have relied on decision trees, while Huang et al. (2002); Mountrakis et al. (2011); Pérez-Ortiz et al. (2016) on support vector machines. On the other hand, El Hajj, Bégué, Guillaume, and Martiné (2009); Guerrero et al. (2013) have favored expert systems to distinguish between various crops.

Neural networks have been successfully used for land cover classification using remotely sensed data by Aitkenhead and Aalders (2008); Cruz-Ramírez et al. (2012); Kavzoglu (2009); Kavzoglu and Mather (2003). Optimal network structure and learning parameters for the backpropagation algorithm have been determined by using specific heuristics and validated with an experiment on two geographical test sites (Kavzoglu & Mather, 2003). Moreover, Kavzoglu (2009) proves how training data size and quality can influence the accuracy of the neural network, proposing a mechanism for eliminating outliers and mixed pixels for a better definition of class boundaries. Neural networks have also been used in a real-time multispectral imaging setup (Noh, Zhang, Shin, Han, & Feng, 2006) for determining the crop nitrogen stress level. The exercise has emphasized that neural networks can be successfully used in real-time and not only off-line environments like satellite images. The multi-objective neural network classifies olive trees, bare soil and different cover crops, using remote sensing data taken in spring and summer (Cruz-Ramírez et al., 2012). As technical instrumentation, a multi-objective evolutionary algorithm is applied to a population of neural networks. The global accuracy obtained was 97.8%.

Decision trees have been proposed for land cover classification in Pal and Mather (2003) due to their simplicity, interpretability, and fast computational model. The study has focused on comparisons of univariate and multivariate models with artificial neural networks and machine learning classifiers. The output of the experiments has shown that decision trees perform better when the univariate model is employed, and the dataset has a small dimension. Hybrid decision tree (Friedl & Brodley, 1997) have been constructed on top of different classifiers, outperforming the accuracy of other machine learning techniques.

Support vector machine for land cover classification from satellite images have been assessed by Huang et al. (2002) for different kernel configurations, with results outperforming decision trees and neural networks. It has been remarked by Mountrakis et al. (2011) that support vector machines are appropriate for remote sensing classification applications due to the small nature of training sets, on which it can give a good generalization. A machine learning system for weed mapping in sunflower and maize crops has been proposed in Pérez-Ortiz et al. (2016). Images to classify sunflower and maize crops are captured by unmanned aerial vehicles. A 95.5% classification accuracy has been obtained based on SVM classifiers. The instances are represented by statistical features (i.e., mean, deviation), texture features (energy, contrast, correlation, and homogeneity), geometric features (i.e. maximum width in pixels), or spatial features (i.e., excess green).

Time series of satellite images have been used as a related approach for classification or monitoring of agricultural crops. In El Hajj et al. (2009) a fuzzy framework has been built for the detection of sugarcane harvesting by combining expert knowledge and the sugarcane growth model. Time series have been used for detecting changes in land cover (Yang & Lo, 2002) by using unsupervised classification from multiple satellite sources on which radiometric normalization was performed. The automatic expert system in Guerrero et al. (2013) uses image processing for crop detection in maize fields. Expert knowledge in Guerrero et al. (2013) is used to separate green plants (crops and weeds) from the background (soil or stones).

3. Technical instrumentation

This section briefly introduces ensemble learning and then describes the argumentation machinery developed on top of an ensemble learner. The argumentation technology is based on Defeasible Logic Programming (DeLP), a formalism for knowledge representation and non-monotonic reasoning. In our approach, DeLP is responsible for handling situations when the classification decision is not clear-cut. That is when the base learners in an ensemble have contradictory opinions about the class of an individual.

3.1. Ensemble learning

By combining base classifiers, ensemble learning aims at a more accurate classification decision at the expense of increased complexity. Three types of reasons support why an ensemble learner might be better than a single classifier: statistical, computational, and representational (Dietterich, 2000). From a statistical perspective, an ensemble learner, even if it will not be better than the best classifier, it diminishes the risk of using an inadequate base classifier from the classifier space. From the computational perspective, different base classifiers may lead to different local optima. Hence, by aggregating them, the obtained ensemble has more chances to compute a better solution. From the representational viewpoint, the classifier space might not contain the optimal classifier for the given problem. In this case, the ensemble can better approximate the decision boundary, by aggregating the available sub-optimal classifiers.

A decision is required when the learning algorithms do not agree on how to classify particular instances. Various methods of blending the outputs of base classifiers (Kuncheva, 2004) have been developed: an algebraic combination of outputs, voting based techniques (majority vote, hierarchical majority voting, weighted majority vote) or behavior knowledge space (Raudys & Roli, 2003).

Let $\mathcal{D}(h) = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ be a classified dataset by the learner h, where $x_i = \langle x_{i,1}, \ldots, x_{i,m}\rangle$ is the vector of input features of the i^{th} instance and $y_i \in \{1, \ldots, K\}$ a discrete value corresponding to its class. Given a set of training instances $\mathcal{T} \subseteq \mathcal{D}$, a classifier $h(x_i)$ is a hypothesis about a function f, that aims $y_i = f(x_i)$, $\forall (x_i, y_i) \in \mathcal{T}$. An ensemble of classifiers $\mathcal{H} = \{h_1(x_i), \ldots, h_L(x_i)\}$ combines the predicted values of each classifier over a dataset instance i into a single decision y_i.

In voting, each single classifier has a vote. The pixel x is labeled with the class y that has obtained the most votes. That is, $y = max_{k=\{1..K\}} V_x(k)$ where y is the class of pixel x and V_x is the number of votes for class k. In the case of ties, the uncertainty of the single classifiers may be used. In probabilistic fusion $y = max_{k=\{1..K\}} \sum_{h=1}^{H} \frac{p_h^k(x)}{S_h(x)}$, where $p_h^k(x)$ is the probabilistic value of pixel x for class k with learner h and $S_h(x)$ is the classification uncertainty of instance x with learner h. For more details on ensemble learning and combination rules, the reader is referred to Kuncheva (2004). The majority voting approach lacks the interpretability for decision makers. How does one could assess the significance in real situations that two learners vote for a class and the third learner for a different class? By using argumentation instead of majority voting, we aim to rely on justified decisions for each contradictory instance.

3.2. Fundamentals of defeasible logic programming

Defeasible reasoning allows that a conclusion supporting by a rule to be defeated in the case of new contradictory information (Pollock, 1995). A defeasible logic program $\mathcal{P} = (\Delta, \delta)$ includes a set Δ of strict rules $c \leftarrow p_1, \ldots, p_n$, and a set δ of defeasible rules $c \prec p_1, \ldots, p_n$, where c and p_i are literals that can be positive or negative (i.e. classically negated with $\sim$). In this paper, we restrict to propositional DeLP, where all literals of the program $\mathcal{P}$ are propositional variables. Deriving literals in DeLP results in the construction of arguments. Facts are encapsulated as strict rules without premises. Hence, facts are included in the set Δ. An argument $\mathcal{A}$ is a set of ground defeasible rules that together with the set Δ provides a logical proof for a given claim c, satisfying the additional requirements of non-contradiction and minimality Gómez et al. (2016).

We note by $\mathcal{A}[c]$ the claim c supported by argument $\mathcal{A}$.

Definition 1. An argument $\mathcal{A}$ is non-contradictory with the set Δ of strict rules if $\mathcal{A} \cup \Delta$ does not entail two complementary literals c and $\sim c$.

Definition 2. An argument $\mathcal{A}$ is minimal if there is no argument $\mathcal{A}' \subset \mathcal{A}$ supporting the same claim $\mathcal{A}[c] = \mathcal{A}'[c]$ for which there exists a defeasible derivation from $\mathcal{A}' \cup \Delta$.

Definition 3. Given a DeLP program $\mathcal{P} = (\Delta, \delta)$, an argument $\mathcal{A}$ for a claim c is a subset of ground instances of the defeasible rules δ in $\mathcal{P}$, such that: (i) there exists a defeasible derivation for c from $\Delta \cup \mathcal{A}$; (ii) $\Pi \cup \mathcal{A}$ is non-contradictory, and (iii) $\mathcal{A}'$ is minimal.

An argument $\mathcal{A}_1[c_1]$ is a sub-argument of another argument $\mathcal{A}_1[c_2]$ if $\mathcal{A}_1 \subseteq \mathcal{A}_2$. Counter-arguments are used to capture the notion of contradiction among base learners in the ensemble. Let $\mathcal{P} = (\Delta, \delta)$ a DeLP program with $\mathcal{A}_1, \mathcal{A}_2 \subseteq \delta$.

Definition 4. An argument $\mathcal{A}_1[c_1]$ is a counterargument for an argument $\mathcal{A}_2[c_1]$ iff there is a subargument $\mathcal{A}_1[c]$ of $\mathcal{A}_2[c]$ such that the set $\Pi \cup \{c_1, c\}$ is contradictory.

Let $Args(\mathcal{P})$ the set of arguments that can be generated from $\mathcal{P}$. Assuming a preference $\preceq$ on conflicting arguments defined as a partial order $\preceq \subseteq Args(\mathcal{P}) \times Args(\mathcal{P})$, we can formalize the definition of defeaters:

Definition 5. An argument $\mathcal{A}_1[c]$ is a defeater for an argument $\mathcal{A}_2[c]$ if $\mathcal{A}_1[c]$ counterargues $\mathcal{A}_2[c]$ and $\mathcal{A}_1[c]$ is preferred over $\mathcal{A}_2[c]$. The argument $\mathcal{A}_1[c]$ is a proper defeater $\mathcal{A}_2[c]$ iff $\mathcal{A}_1[c]$ is strictly preferred over $\mathcal{A}_{\in}c$ w.r.t. $\preceq$; The argument $\mathcal{A}_1[c]$ is a blocking defeater for $\mathcal{A}_2[c]$ if $\mathcal{A}_1[c]$ and $\mathcal{A}[c]$ are unrelated to each other.

The preference criteria can be explicitly specified within the defeasible rules in δ or various conflict resolution strategies can be used. An example of such strategy is *specificity* (Simari & Loui, 1992) that favorizes arguments which are more specific (i.e. those who rely on more premises).

To determine the status of an argument $\mathcal{A}$, a dialectical process recursively takes into considerations defeaters of $\mathcal{A}$ defeaters of defeaters of $\mathcal{A}$ and so on.

Definition 6. An argumentation line for argument $\mathcal{A}_{\prime}$ and query q_0 is a chain of tuples $[\langle A_0, q_0\rangle, \langle A_1, q_2\rangle, \ldots, \langle A_n, q_n\rangle]$, where the pairs $\langle A_{2k}, q_{2k}\rangle$ are conveyed by the proponent of the argument $\mathcal{A}_{\prime}$, while the pairs $\langle A_{2k+1}, q_{2k+1}\rangle$ are conveyed by the opponent of the argument $\mathcal{A}_{\prime}$.

In our approach, the proponents and the opponents are represented by base classifiers in the ensemble or the domain expert.

Definition 7. A dialectical tree $\mathcal{T}_{\langle \mathcal{A}_0, q_0\rangle}$ represents all possible argumentation lines starting with $\langle A_0, q_0\rangle$ based on a given DeLP knowledge base.

Nodes in a dialectical tree $\mathcal{T}_{\langle \mathcal{A}_0, q_0\rangle}$ can be *undefeated* (U) or *defeated* (D). The process of labeling a dialectical tree starts from the leaves, which are U-nodes as they have no defeaters. A inner node is labeled D iff it has at least one U-node among its children nodes.

Definition 8. An argument $\langle \mathcal{A}_0, q_0\rangle$ is valid w.r.t. a DeLP program $\mathcal{P}$ iff the root of its dialectical tree $\mathcal{T}_{\langle \mathcal{A}_0, Q_0\rangle}$ is labeled as *U-node*. $\mathcal{A}_0$ is also called the warrant of q_0, or inversely q_0 is warranted by $\mathcal{A}_0$.

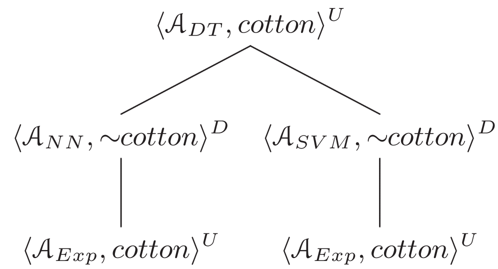


Fig. 1. Dialectical tree for the query *cotton*.

Fig. 1 exemplifies a dialectical tree in which statistical classifiers and expert knowledge argument over the literal query *cotton*. Argument $\langle \mathcal{A}_{DT}, cotton\rangle^U$ proposed by the decision tree classifier is valid as it is undefeated. There are two argumentation lines. The first counterargument $\langle \mathcal{A}_{NN}, \sim cotton\rangle^D$ derived from the neural network classifier is defeated by the valid argument $\langle \mathcal{A}_{Exp}, cotton\rangle^U$ proposed by the expert knowledge. The second counterargument $\langle \mathcal{A}_{SVM}, \sim cotton\rangle^D$ derived from the support vector machine is defeated by the same valid argument $\langle \mathcal{A}_{Exp}, cotton\rangle^U$ proposed by the expert.

4. Performing argumentation for conflict resolution in ensemble learning

This section starts by presenting the test site used for experiments. Then we formalize how argumentation can solve conflicts in an ensemble learner. Finally, the top level architecture of the developed system is described.

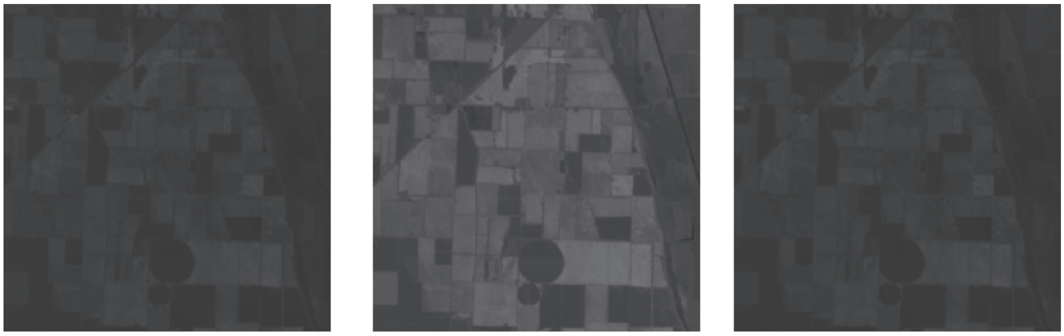
4.1. Test site and data-set

The test site used for the experiment is an area of 20 square kilometers in the New Madrid County, southeast of the Missouri State, USA. This area is characterized by a humid subtropical climate and favorable agricultural activities, with an average of 1,087 acres per farm land of which 96.5% is used as cropland (United States Department of Agriculture, 2012). Our classification experiment aims for discriminating between four types of crops: corn, soybean, cotton and rice.

The Landsat image was acquired on July 5th, 2014 and exported into GeoTIFF format by using the USGS online system [1]. Landsat images are pre-processed by USGS using a cubic convolution re-sampling and a standard terrain correction by incorporating ground truth points. No additional pre-processing was performed after the image was received from USGS. No noise correction was required since the image had no degradation caused by clouds. Moreover, since the experiment relies solely on one image capture to derive the dataset there is no need for sensors calibration, earth model/projection corrections or other adjustments specific for correlating multiple satellite images.

A Landsat image consists of multiple grayscale 16-bit images, each storing a spectral band captured by the satellite. Four out of the nine OLI (Operational Land Imager) bands are used for constructing the classification data-set. The four bands are chosen based on their correlation to the vegetation discrimination process. Table 1 lists the four bands together with the extracted features. Bands 3 and 6 are used as features in their raw format, while bands 4 and 5 are combined into a new feature: Normalized Difference Vegetation Index (NDVI). NDVI is a proven indicator of land-use and cover changes (DeFries & Townshend, 1994), being calcu-

[1] http://earthexplorer.usgs.gov/.



(a) Green (Band 3) (b) Moisture (Band 6) (c) NDVI (Bands 4 and 5)

Fig. 2. Landsat gray-scale images corresponding to the features set.

Table 1
Features of the crop dataset obtained from Landsat 8.

Landsat 8 OLI Band	Feature Name	Justification
Band 3 (Green)	Green Level	Indicates peak vegetation.
Band 6 (Short-wave infrared)	Moisture Level	Indicates moisture content of both soil and vegetation.
Band 4 (Red) and Band 5 (Near infrared)	Normalized Difference Vegetation Index	Indicates photosynthetic activity.

lated from the red (*Red*) and near infrared bands (*NIR*) by the following formula:

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{1}$$

Fig. 2 displays the Landsat gray-scale images corresponding to the three features of the dataset: green, moisture and NDVI.

The ground truth reference is obtained from the United States Department of Agriculture - Statistics Service[2], using a filter for the area of interest and the timestamp of the Landsat image. Ground truth is drawn by hand on top of the Landsat RGB image using image processing software. Each pixel of ground truth is color coded by hand using one of the four colors corresponding to its class. The resulting ground truth image is depicted in Fig. 3.

Algorithm 1 presents the method of building the classification dataset from the ground truth and Landsat band images. In lines 1-2 the algorithm iterates over all the pixels of the ground truth image. If the current pixel belongs to one of the classes (line 3) then the classification dataset is augmented with a new instance (line 4). The new instance is represented by the features of the pixel extracted from the bands images and its class.

Algorithm 1: Constructing the classification dataset.

```
Input: GT, ground truth image (Fig. 3)
Input: B3, band 3 grayscale image (Fig. 2-a)
Input: B6, band 6 grayscale image (Fig. 2-b)
Input: NDVI, ndvi grayscale image (Fig. 2-c)
Output: 𝒟, classification dataset
1 foreach r : row of pixels in GT do
2     foreach c : column of pixels in GT do
3         if GT_{r,c} has a class color then
4             𝒟 = 𝒟 ∪ {(B3_{r,c}, B6_{r,c}, NDVI_{r,c}), GT_{r,c}}
5 return 𝒟
```

[2] http://nassgeodata.gmu.edu/CropScape/.

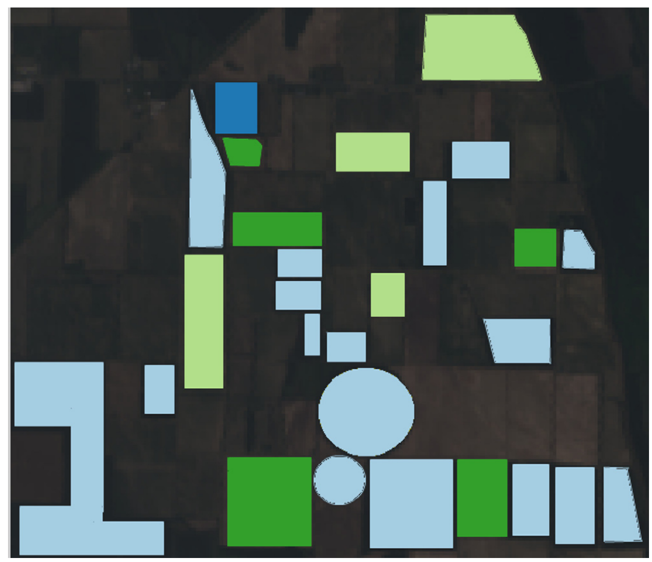


Fig. 3. Landsat RGB image with ground truth mask (highlighted pixels per crop class). The color codes for *corn, rice, cotton* and *soybean* are light blue, blue, light green and green respectively. The classification dataset contains 5,407 instances. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The ensemble base classifiers and expert knowledge have to use the same scale for the input values. Since normalization is required for at least one of the statistical classifier (i.e. the neural network), normalization is applied on the whole input dataset, by using Eq. (2):

$$x_{norm} = \frac{x - \mu}{\sigma} \tag{2}$$

where x is the real-value input while μ is the mean and σ the standard deviation of the variable.

The obtained classification dataset is split into two independent datasets: 20% used for training and validating the classification models and 80% used for testing. This split mimics the idea that the classifiers should be able to predict vast areas after being trained on a small number of plots, a characteristic of the crop classification problem, as obtaining ground truth references is often associated with the effort of inspecting the plots in person. The resulted training set contains 1,065 instances, while the test set contains 4,342 instances.

4.2. Applying argumentation machinery on inconsistent classification

In the case of inconsistent classifications by two or more learning algorithms, more analyze is required either by human intervention or by more accurate technical instrumentation. An argumentation machinery can support the decision of the human expert by providing pro and counter arguments for a debatable class. The resulted argumentation framework, complemented with human knowledge leads to justified decisions in case of class controversy among base classifiers in an ensemble.

The criteria used for deciding for which instances to accept the classification and for which to apply the argumentation machinery is when at least one base classifier outputs a different classification for the given instance. That is when there is at least a warranted argument supporting a different class.

Definition 9. An instance i belongs to the conflict set Γ iff there are at least two learners in the ensemble $\mathcal{H}$ that output different classes for that instance i. Formally:

$$i \in \Gamma \text{ iff } \exists h_l, h_{l'} \in \mathcal{H}, l \neq l',$$
$$\text{s.t. } h_l(x_i) = y_i, h_{l'}(x_i) = y'_i \text{ with } y_i \neq y'_i$$

Example 1. Consider the binary ensemble $\mathcal{H} = \{h_{dt}, h_{nn}\}$ formed by a decision tree and a neural network classifier employed for a binary classification, $y \in \{-, +\}$. The conflict set Γ is formed by all instances that the decision tree classifies "+" and the neural network "−", together with the ones that the decision tree classifies "− " and the neural network "+". Given the labeled datasets $\mathcal{D}(h_{dt}) = \{(i_1, +), (i_2, +), (i_3, -), (i_4, -)\}$ and $\mathcal{D}(h_{nn}) = \{(i_1, +), (i_2, -), (i_3, +), (i_4, -)\}$, the conflict set would be $\Gamma = \{i_2, i_3\}$. We assume that no further analysis is required for the instances outside the conflict set Γ. Here, all learners in $\mathcal{H}$ agree on the class of the instances i_1 and i_4.

Definition 10. A classification rule is an implication $condition(x) \rightarrow y$ where the condition is a conjunction of tests over the features of input x and y is the class.

Example 2. Consider a bi-dimensional input dataset of binary values 0 and 1 that needs to be classified following the logical AND operation. The classification rule for class "+" is: $(equal(x_0, 1) \wedge equal(x_1, 1) \rightarrow \text{“}+\text{”})$, while class "−" is described by two classification rules $(equal(x_0, 0) \rightarrow \text{“}-\text{”})$ and $(equal(x_1, 0) \rightarrow \text{“}-\text{”})$.

The scope of the classification rules is to describe why a classifier $h \in \mathcal{H}$ believes that class y should be assigned to an instance x_i.

Definition 11. An ensemble knowledge base Ens_{KB} is the set of all classification rules that describe the classification for each classifier $h \in \mathcal{H}$.

An argumentation knowledge base merges two sources of knowledge: rules mined from base classifiers (Ens_{KB}) and expert knowledge (E_{KB}). As Ens_{KB} contains inconsistent rules, all the argumentation base will contain conflictual rules. The expert knowledge E_{KB} is defined by distinguishing different classes y based on the similar features of the classification dataset $\mathcal{D}$. The expert knowledge features can be augmented by deriving new features from existing ones. The two sources of knowledge are aggregated as a DeLP program that performs dialectical analysis to decide the class of the given instance. Formally:

Definition 12. An argumentation knowledge base is a tuple $\mathcal{A} = \langle Ens_{KB}, E_{KB}, \oplus \rangle$, where Ens_{KB} represents the knowledge extracted from the ensemble learner and E_{KB} is the domain expert knowledge. The aggregation strategy $\oplus$ for the set $\{Ens_{KB}, E_{KB}\}$ applies the set of conflict resolution strategies (heuristics) $\mathcal{RS}$ for computing a partial order relation between rules in $\{Ens_{KB}, E_{KB}\}$.

In the basic conflict resolution strategy of defeasible logic, strict rules are stronger than defeasible rules. We note this strategy with s_0. Two other possible conflict resolution strategies are: i) s_1: expert knowledge stronger than any classifier knowledge: $\forall r \in E_{KB}$ and $\forall s \in Ens_{KB}$, $r \succ s$ or ii) s_2: specific rules stronger than general rules: given r: $a_i \rightarrow y_1$ and s: $b_i \rightarrow y_2$ if $\{b_i\} \subset \{a_i\}$ then $r \succ s$. Hence, a possible aggregation strategy is $\oplus = [s_0, s_1, s_2]$.

Our top level approach is captured by Algorithm 2. Given the ensemble of classifiers $\mathcal{H} = \{h_1, ...h_L\}$ and an instance case x by its vector of features, the Algorithm 2 outputs the class y of the instance x. If all the classifiers h_i agree on the class of an individual, then that classification is returned (lines 1-2). In the case of conflict between classifiers in $\mathcal{H}$, the set of ensemble knowledge base Ens_{KB} is developed by unifying the extracted classification rules from all base classifiers (lines 4-7). The method ExtractRules has specific implementation for each base classifier. DeLP reasoner is asked to produce a *Undefeated* (*True*) or *Defeated* (*False*) answer for each class $y \in \{1..K\}$, by using the ensemble classification Ens_{KB} and expert rules E_{KB} as knowledge bases (lines 8-10). If there exists exactly one class that receives a *True* answer from the DeLP reasoner, then this class settles the dispute (lines 11-12). Otherwise, the classification is undecided (line 14).

Algorithm 2: Classifying a new instance case.

```
Input: H = {h_1, ..., h_L}, ensemble of classifiers h_l, l ∈ {1..L}
Input: x, feature vector of the new case
Input: E_KB, expert knowledge
Output: y, class of the new case, y ∈ {1..K}
Output: T, dialectical tree
1  if ∀h_l ∈ H, h_l(x) = y then
2      return y
3  else
4      Ens_KB ← {}
5      foreach classifier h_l ∈ H do
6          rules ← ExtractRules(h_l(x))
7          Ens_KB ← Ens_KB ∪ {rules}
8      answer ← {}
9      foreach class y ∈ {1..K} do
10         answer_y ← DelpAnswer(KnowledgeBase: Ens_KB ∪ E_KB,
           Query: y?)
11     if ∃!y ∈ {1..K} s.t. answer_y=true and ∀z ≠ y answer_z ≠ true
       then
12         return y, T
13     else
14         return undecided, T
```

4.3. System architecture

Fig. 4 presents the architecture of the developed crop classification system. The top level encapsulates data layer operations. The area of interest is extracted from the input satellite image. The features of the classification dataset are extracted from the multispectral values. The obtained dataset is normalized and split into two sets, one used for training the base classifiers and one for validating the ensemble learner.

The middle layer covers the three independent statistical classifiers: decision tree, artificial neural network and support vector machine, that compose the statistical ensemble learner $\mathcal{H} = \{h_{dt}, h_{nn}, h_{svm}\}$. The base classifiers are trained and tested by using inputs only from the training set. Each trained classifier is asked to predict the class of instances in the validation set, together with argumentation rules.

The bottom layer encloses the argumentation framework that is used in case of conflicts among the learners from the ensemble $\mathcal{H}$.

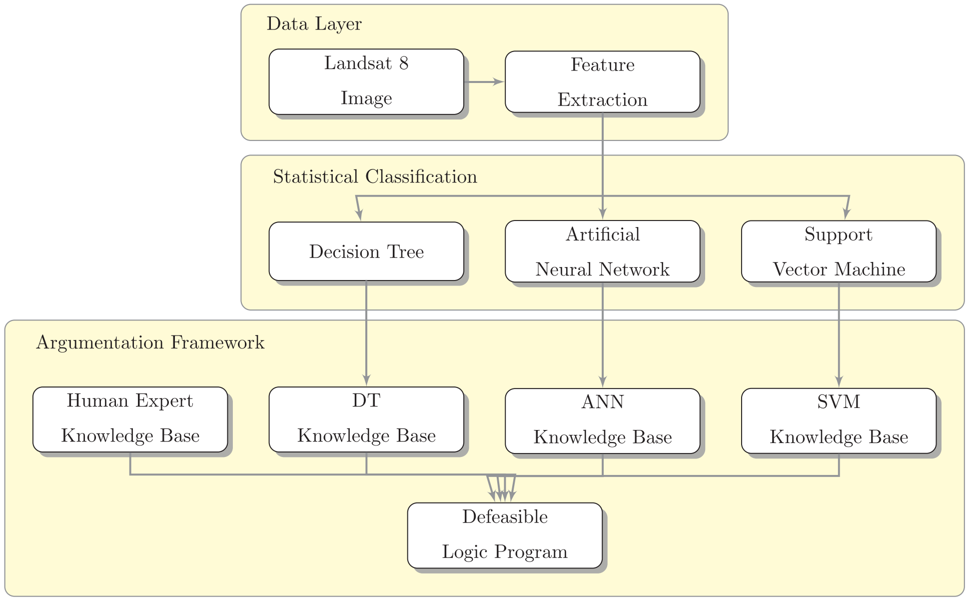


Fig. 4. Classification system architecture.

The inputs of this layer are the classification rules extracted from each classifier of the layer above. The rules are merged with expert defined knowledge and are sent to a DeLP reasoner for conflict resolution.

Regarding technological instrumentation, the top layer comprises of MATLAB and C# scripts used for image and dataset processing. The decision tree and support vector machine classifiers and the rule extraction algorithms are implemented by using the *scikit* machine learning library of Python. The neural network and the argumentation base is implemented by using C#. DeLP reasoning is performed with the REST API offered by the Tweety project (Thimm, 2014). Our complete developed system and the data set used for experiments are available at http://github.com/stefan-contiu/crops-delp.

5. Interleaving rule mining and agricultural knowledge for crop classification

This section covers the knowledge part used for the crop classification. Firstly, the section presents the methods for extracting rules from the three statistical classifiers: decision tree, neural network and support vector machine. Secondly, the strategy for building the expert knowledge is detailed, together with concrete samples of derived expert rules.

5.1. Extracting DeLP rules from base learners

This section introduces the proposed methods for extracting DeLP rules from an ensemble learner. Let $\mathcal{H} = \{h_{dt}, h_{nn}, h_{svm}\}$ and classes $y = \{corn, rice, cotton, soybean\}$.

Generating defeasible rules from decision tree classifier. The decision tree classifier has the advantage of simplicity and easy interpretation due to its white box model. The classifier can explain its predictions by producing a set of if-then-else decision rules, usually visualized as a tree. Decision trees have been assessed as acceptable good for crop classification from multispectral data (Pal & Mather, 2003) when using a univariate model and not on high-dimensional datasets.

The chosen decision tree model is binary univariate. Each non-leaf tree node represents a condition of the form: $x_i < threshold$, where x_i is a feature of the dataset while each leaf node denotes a class. Optimal features and threshold values are determined by using the CART (Classification and Regression Trees) algorithm (Breiman, Friedman, Stone, & Olshen, 1984) which maximizes the information gain for each node. The 10-fold cross validation method is used for assessing the best criterion and strategy for splitting the nodes of the decision tree. All combinations of split criteria (gini or entropy) and split strategies (best or random) produced the same cross validation accuracy scores of 99.7 (+/- 0.9). Therefore, the chosen parameter values for criterion and strategies are set to gini and best split.

Translating decision tree classification rules into DeLP rules for an instance classified as y is performed by the following steps of the ExtractRules(h_{dt}) method:

Step 1. Express the branch of the tree that determined the classification as a conjunction of conditions: $\mathcal{C} = condition_1, condition_2, \ldots, condition_n$.
Step 2. Introduce one defeasible DeLP rule of the form: $y \prec \mathcal{C}$.
Step 3. Introduce DeLP rules for all the other classes $y' \neq y$ of the form: $\sim y' \prec \mathcal{C}$.

The rationale of the translated rules is that, given the conjunction of conditions $\mathcal{C}$, y is chosen as predicted class as long as nothing is posed against it. Similarly, any $y' \neq y$ is an incorrect prediction as long as $\mathcal{C}$ is not defeated.

Example 3. Consider the classification of *cotton* instances in the decision tree in Fig. 5. The cotton tree branch is expressed as a conjunction of conditions (step 1):

$$decision\_tree(g, m, ndvi) \prec m > -0.01, ndvi \in [-1.31, 0.07].$$

The DeLP rule pleading for *cotton* class is then introduced (step 2):

$$cotton \prec decision\_tree(g, m, ndvi).$$

The defeasible rules which signal that all the other classes are incorrect predictions are introduced (step 3):

$$\begin{aligned} \sim corn &\prec decision\_tree(g, m, ndvi). \\ \sim rice &\prec decision\_tree(g, m, ndvi). \\ \sim soybean &\prec decision\_tree(g, m, ndvi). \end{aligned}$$

Generating defeasible rules from neural network classifier. We adopt a feed-forward neural network model containing a single hidden

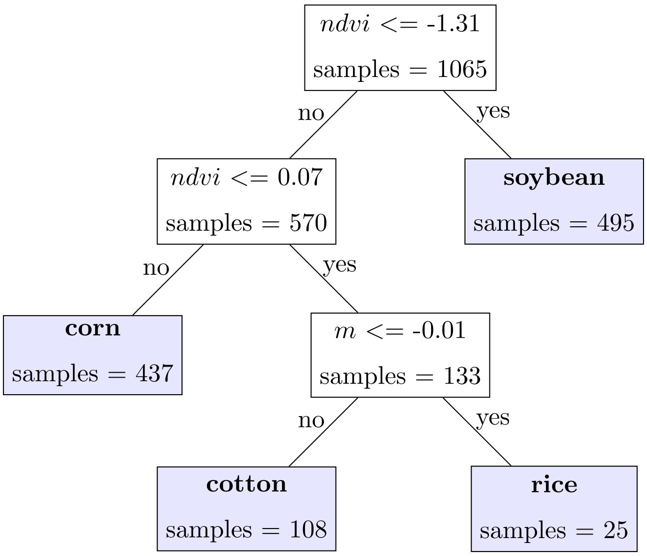


Fig. 5. Decision tree obtained with the CART algorithm for the crop training dataset.

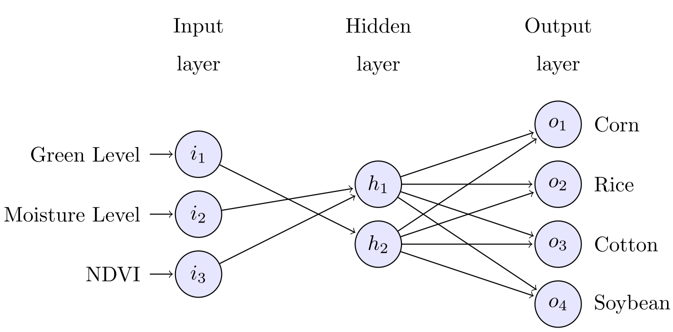


Fig. 6. The topology of the pruned artificial neural network.

layer. The input layer contains three units corresponding to the input features green level, moisture level, and NDVI. The output layer contains four units corresponding to the classes of crops we intend to discriminate: corn, rice, cotton, and soybean. The predicted class is determined by the output layer unit having the maximum value. The 10-fold cross validation method is used to determine the number of units in the hidden layer. Experiments with more than two hidden units show no growth from 98.4(+/–2.4) averaged accuracy across the ten folds, therefore the network is set with two nodes in the middle layer.

Skeletal pruned neural networks produce simpler classification rules (Setiono & Liu, 1997) thus, the redundant connections between the input and hidden layers are removed. Input layer contains three nodes, while hidden layer contains two nodes. Hence, there are six possible directed edges from the input to the hidden layer. Considering that any edge of the six can be on or off, there are $2^6 = 64$ different ways of connecting the input and the hidden layers. Each of the 64 network configurations is trained and evaluated using 10-fold cross validation. The selected optimal configuration has the least number of edges from input to hidden layers and a maximal accuracy. The network with three edges between input and hidden layers depicted in Fig. 6 has achieved a maximal accuracy of 98.4(+/–2.4) on the crops training dataset.

The hyperbolic tangent function is used instead of the standard logistic function as an activation method due of faster convergence (LeCun, Bottou, Orr, & Müller, 2012). The hyperbolic tangent function is symmetric to the origin, similarly with the training dataset, on which we perform Gaussian normalization during the features extraction phase.

The neural network is trained by using the backpropagation algorithm in conjunction with gradient descent. Learning rate and momentum are set to 0.5 and 0.01 respectively by using a trial-and-error method. The network is trained for a maximum number of 2000 epochs.

Once the network is trained to optimal accuracy, the set of classification rules is extracted. The neural network will use these rules as arguments when asked by the DeLP reasoner to explain its classification. The extraction steps are based on the NeuroLinear algorithm, used for extracting oblique decision rules from trained neural networks (Setiono & Liu, 1997). This method is complemented with a CART decision tree classifier (Breiman et al., 1984) and with a translation of decision rules to defeasible logic rules. The steps of the ExtractRules(h_{nn})) method are as follows:

Step 1. Start by training the pruned neural network depicted in Fig. 6, which consists of i_1, i_2, i_3 input units, h_1, h_2 hidden units and o_1, o_2, o_3, o_4 output units. The network has three edges between the input and hidden layers, their optimal weights being determined during training w_1 (from i_1 to h_2), w_2 (from i_2 to h_1) and w_3 (from i_2 to h_2).

Step 2. Re-pass all dataset inputs through the trained neural network and collect the values of the two hidden nodes h_1 and h_2. Hence, a new bi-dimensional dataset H is obtained.

Step 3. Train a decision tree learner by using the CART algorithm on the new bi-dimensional dataset H and obtain a set of classification rules. These rules are expressed in terms of the hidden nodes values h_1 and h_2.

Step 4. Generate rules that are expressed in terms of the input features i_1, i_2, i_3. First, hidden nodes are expressed in terms of the input nodes as $h_1 = i_2 * w_2 + i_3 * w3$ and $h_2 = i_1 * w_1$. Second, as the hidden units are the result of *tanh* function, the inverse function $tanh^{-1}$ is applied on the rules decision boundaries.

Step 5. Translate the decision rules to DeLP statements by using the same steps described for the decision tree classifier ExtractRules(h_{dt}).

Example 4. Consider the process of extracting classification rules for the soybean class from the neural network in Fig. 6. After collecting all the neural network activation values (steps 1 and 2) and applying the decision tree classifier (step 3), the decision rule for soybean class is determined to be (step 4):

$$h_1 > 1.83 = tanh^{-1}(0.95)$$

where h_1 is a hidden node connected to the moisture level(m) and ndvi nodes by edges with weights 0.21 and –1.44 respectively. The classification rule can be re-written as (step 5):

$$0.21 * m - 1.44 * ndvi > 1.83$$

The DeLP rules extracted from the neural network classifier for the soybean class are:

$$\begin{array}{rcl} neural\_net(g, m, ndvi) & \prec & 0.21 * m - 1.44 * ndvi > 1.83. \\ \sim corn & \prec & neural\_net(g, m, ndvi). \\ \sim rice & \prec & neural\_net(g, m, ndvi). \\ \sim cotton & \prec & neural\_net(g, m, ndvi). \\ soybean & \prec & neural\_net(g, m, ndvi). \end{array}$$

Generating defeasible rules from support vector machine classifier.. The support vector machine (SVM) is chosen as a base classifier as it can offer a different perspective on the decision boundaries between the four classes.

Since SVM conceptually works with binary classification, a strategy needs to be employed for solving the four class (corn, rice, cotton, and soybean) task classification. "One against one" strategy is chosen, as it has been proven more suitable for practical use

Table 2
Expert knowledge used for deriving expert rules. The knowledge is specific for the test site defined in Section 4.1. Green, moisture, and NDVI values use the same scales as the normalized crops dataset.

	Corn	Rice	Cotton	Soybean
Green Margin	−1.13 to 0.05	−0.69 to -0.5	−0.03 to 2.48	0.21 to 3.15
Moist. Margin	−0.94 to 0.07	−1.33 to y0.51	0.18 to 2.08	0.36 - 2.31
NDVI Margin	−0.59 to 1.07	−1.37 to 0.18	−1.56 to 0.06	−2.22 to −0.28
Planting	Apr 20-May 25	May 1-May 25	May 5-May 20	May 15-Jul 1
Harvesting	Sep 20-Oct 30	Sep 25-Oct 25	Oct 5-Oct 30	Oct 10-Oct 30
Harvest Signif. Color Change	Yes	No	Yes	Yes

than "one-against-all" or DAGSVM methods (Hsu & Lin, 2002). In the "one against one" strategy, one SVM is built for each pair of classes. That is N(N - 1)/2 SVMs are constructed for N classes. In our case, six classifiers are constructed for a four-class classification.

A 10-fold cross validation is run on the training dataset for the following kernels: RBF, polynomial, and linear. The results 99.9(+/-0.3) for linear kernel, 99.8(+/-0.3) for RBF, and 99.3(+/-1.5) for polynomial indicate that the linear kernel is a suitable choice for the SVM model.

Rule extraction is performed by a learning-based decompositional algorithm (Setiono & Liu, 1997) complemented with a CART (Breiman et al., 1984) classifier to extract if-then-else classification rules. Decision rules are then translated to DeLP rules such that the argumentation framework can make use of them. The steps for the ExtractRules(h_{svm}) method are:

Step 1. Train a linear SVM model on the input dataset.
Step 2. Identify the dataset instances which are chosen by SVM as support vectors and add them to a set V. The three-dimensional set V is a subset of the input dataset and does not include any predicted class.
Step 3. Classify V by using the same SVM model. Therefore, V is augmented with predicted classes.
Step 4. Train a decision tree learner by using the CART algorithm on the V dataset to obtain the classification rules.
Step 5. Translate the decision rules to DeLP statements by using the same steps described for the decision tree classifier ExtractRules(h_{dt}).

Example 5. Consider the classification of cotton instances of the crops dataset by using an SVM model (step 1). There are $|V| = 25$ dataset instances chosen to form the support vectors (step 2). After reclassifying V by the same SVM model (step 3) and applying a decision tree classifier (step 4), the decision rule for soybean is determined to be (step 5):

$g > 0.001$ and $ndvi \leqslant -1.1$

The DeLP rules extracted from the SVM classifier explaining the soybean classification are:

$$\begin{aligned} svm(g, m, ndvi) &\prec g > 0.001, ndvi \leqslant -1.1 \\ \sim corn &\prec svm(g, m, ndvi). \\ \sim rice &\prec svm(g, m, ndvi). \\ \sim cotton &\prec svm(g, m, ndvi). \\ soybean &\prec svm(g, m, ndvi). \end{aligned}$$

5.2. Expert knowledge

The expert knowledge is built as a subset of some of the most important morphological and phenological characteristics of the four crops. The expert knowledge is not exhaustive. Its scope is to demonstrate the feasibility of the hybrid classification method and it is derived and valid only for the area of interest. Agriculture experts should be able to refine or adapt this knowledge to other crop classification contexts. Table 2 lists the knowledge encapsulated by the expert system.

Each of the four crops has unique morphological and phenological characteristics. Plant morphology represents the external form and structure of the plants. Plant phenology represents the occurrence of biological events in the plant life cycle. An example of a morphological feature is the plant pigmentation which accounts for the photosynthesis function, possibly telling if the crop was dry or fresh when harvested. Examples of phenological features are date observations that can be correlated to planting and harvesting dates.

Corn is a large grain plant widely cultivated throughout the area of interest. Planting of the corn is usually done between April 20th and May 25th, and the harvesting is done between September 20th and October 30th. Corn is a plant that develops slower under low temperatures and faster under high temperatures, as long as water requirements are met. Thus, the harvesting date can vary depending on the average temperatures and water requirements. Rice is a specific humid subtropical plant and the least cultivated out of the four crop types. The cultivation process includes a nursery stage before the vegetative stage. The rice is transplanted to the fields usually between May 1st and May 25th. Harvesting is usually performed during September 25th and October 25th. Cotton is also specific to the humid subtropical climate. It is usually planted between May 5th and May 20th and emerges within five to ten days after planting. The white cotton flowers start to blossom between days 45 and 65. It is usually harvested between October 5th and October 30th. The plant is in a dry state during harvesting. Soybean is a legume, being the most cultivated within the Missouri state. Planting period is the latest one among the four crops, between May 15th and July 1st. Harvesting is performed between October 10th and October 30th.

The first three rows of Table 2 display phenological margins of NDVI, green and moisture levels for each of the four crops. Margin values are determined from phenological profiles of sample points from an enlarged area of interest surrounding the test site. Eq. (2) is applied to the margin values to maintain scale consistency, with the mean and standard deviation used for the input dataset. If the margins are an indicator for their class, values outside the margins deny the class. One defeasible rule is introduced for indicating the class and three defeasible rules to negate the class. The following expert rules are derived from *corn* margins, each of the other three classes produce a similar set of rules but with specific margin values:

$$\begin{aligned} expert\_corn(g) &\leftarrow g \in [-1.13, 0.05]. \\ expert\_corn(m) &\leftarrow m \in [-0.94, 0.07]. \\ expert\_corn(ndvi) &\leftarrow ndvi \in [-0.59, 1.07]. \\ corn &\prec expert\_corn(g), expert\_corn(m), \\ & \quad expert\_corn(ndvi). \\ \sim corn &\prec \sim expert\_corn(g). \\ \sim corn &\prec \sim expert\_corn(m). \\ \sim corn &\prec \sim expert\_corn(ndvi). \end{aligned}$$

Some expert rules make use of crops planting and harvesting dates. Their reference values are displayed in the fourth and fifth rows of Table 2 and are extracted from the USDA Agricultural Handbook (NASS, 2010) for the area of interest. Because the statistical classification dataset does not make use of these features, the input dataset is augmented with values used only for the exported knowledge. Margin values are determined by an empiric method, considering that Landsat images follow a period of two weeks. The past and future images are observed, plotting the NDVI to validate the crop life time-frame and extract the approximate planting and harvesting date. Planting and harvesting rules, like margin rules, can indicate or negate a class. Examples of such expert rules derived for *corn* are:

$$\begin{aligned} expert\_corn(plant) &\leftarrow plant \in [Apr\ 20, May\ 25].\\ expert\_corn(harvest) &\leftarrow harvest \in [Sep\ 20, Oct\ 30].\\ corn &\prec expert\_corn(plant),\\ &\quad expert\_corn(harvest).\\ \sim corn &\prec \sim expert\_corn(plant).\\ \sim corn &\prec \sim expert\_corn(harvest). \end{aligned}$$

Whether there is a significant crop color change during harvesting can be empirically correlated to the dataset by observing the decreases in the green, moisture, and ndvi features. Corn, cotton, and soybean turn into a yellow or gold color at maturity while rice still preserves a component of green. The following rules are introduced by this new feature in the set of expert rules:

$$\begin{aligned} corn &\prec harvest\_color\_change\\ \sim rice &\leftarrow harvest\_color\_change\\ cotton &\prec harvest\_color\_change\\ soybean &\prec harvest\_color\_change \end{aligned}$$

A total of 42 strict and defeasible expert rules were derived for the four crops, as follows: 28 rules by using the marginal expert values for *green, moisture* and *ndvi*, 20 rules by using *plant* and *harvest* dates, and 4 rules by using the color change at the harvesting time. These rules are used for conflict resolution to improve the accuracy of our ensemble classifier.

6. Classification conflict resolution through argumentation

The section presents the DeLP argumentation mechanism for conflict resolution and exemplifies the argumentation analysis on a conflicting sample of the input dataset. We also show the experiments supporting the feasibility of our solution.

6.1. Resolving classification conflicts

Our method for conflict resolution makes use of an argumentative framework based on defeasible logic programming. The knowledge base of the argumentation framework is the aggregate of the statistical classifiers and expert knowledge. A defeasible logic program is constructed for each of the debatable instances. The program is asked to resolve the classification dispute and argument its decision. The following steps describe the process leading to conflict resolution:

Step 1. Add the expert generated rules to the DeLP program.

Step 2. Ask the decision tree, neural network and support vector machine for DeLP rules to provide supporting arguments for their prediction. There is no need to request the complete knowledge of the base learners since during argumentation only the reasons that led them to output conflicting classification predictions are used. Aggregate the knowledge extracted from the statistical learners with the expert knowledge into the DeLP program.

Step 3. Eliminate all mathematical formulas from the DeLP program, such that the resulted program is based solely on logic programming. All such statements are evaluated and replaced with facts.

Step 4 Query the DeLP program using each of the four crops as a query. If exactly one crop has a positive answer, then the dispute is considered settled. Otherwise, the classification is undecided.

Algorithm 3: Resolving classification using DeLP for each debatable instance.

Input: $\mathcal{H} = \{h_{dt}, h_{nn}, h_{svm}\}$, ensemble of three classifiers
Input: $\mathcal{Y} = \{corn, cotton, soybean, rice\}$, class labels
Input: Γ, conflict set of debatable instances
Input: E_{KB}, expert knowledge
Output: Y, the set of classes assigned to each instance of Γ

1 **foreach** $x_i \in \Gamma$ **do**
2 $\quad \mathcal{P} \leftarrow E_{KB}$
3 $\quad$ **foreach** *classifier* $h \in \mathcal{H}$ **do**
4 $\quad\quad \mathcal{P} \leftarrow \mathcal{P} \cup$ ExtractRules$(h(x_i))$
5 $\quad$ **foreach** *rule* $r \in \mathcal{P}$ **do**
6 $\quad\quad$ **if** *r is a mathematical formula* **then**
7 $\quad\quad\quad$ fact $\leftarrow$ evaluate rule r for input x_i
8 $\quad\quad\quad \mathcal{P} \leftarrow \mathcal{P} \setminus \{r\} \cup \{fact\}$
9 $\quad y_i \leftarrow$ DelpResolution$(\mathcal{P}, \mathcal{Y})$

The formal representation of the above four steps appears in Algorithm 3. During the first step (line 2) expert rules are added to the DeLP program. Since strict rules are introduced only from expert knowledge, we assume that the fact that they are non-contradictory can be validated beforehand. During the second step (lines 3-4), contradictory defeasible rules are extracted from the three statistical classifiers. In the third step (lines 5-8), all mathematical formulas are pre-processed and removed from the DeLP. In the fourth step (line 9), the constructed DeLP program is asked to produce the resolved class of the contradictory instance. The implementation of DelpResolution subroutine is detailed in Algorithm 4.

Algorithm 4: DelpResolution: producing the resolved crop class by DeLP

Input: $\mathcal{P}$, a DeLP program
Input: $\mathcal{Y} = \{corn, cotton, soybean, rice\}$
Output: $y \in \mathcal{Y}$, the resolved crop class
Output: $\mathcal{T}$, dialectical tree

answer $\leftarrow \langle null, null, null, null\rangle$
foreach *class* $y \in \mathcal{Y}$ **do**
$\quad \mathcal{T} \leftarrow$ Build dialectical tree to warrant y over $\mathcal{P}$
$\quad$ **if** *root*$(\mathcal{T})$ *is labeled Undefeated* **then**
$\quad\quad answer_y \leftarrow \langle True, \mathcal{T}\rangle$
$\quad$ **else**
$\quad\quad \overline{\mathcal{T}} \leftarrow$ Build dialectical tree to warrant $\sim y$ over $\mathcal{P}$
$\quad\quad$ **if** *root*$(\overline{\mathcal{T}})$ *is labeled Undefeated* **then**
$\quad\quad\quad answer_y \leftarrow \langle False, \overline{\mathcal{T}}\rangle$
$\quad\quad$ **else**
$\quad\quad\quad answer_y \leftarrow \langle False, null\rangle$
if $\exists! y \in \mathcal{Y}$ *s.t.* $answer_y$*=True and* $\forall z \neq y$ $answer_z \neq True$ **then**
$\quad$ **return** y, $\mathcal{T}$
else
$\quad$ **return** *undecided*, $\mathcal{T}$

Example 6. Consider the pre-processing phase of the expert rule $expert\_corn(plant) \leftarrow plant \in [Apr\ 20, May\ 25]$

If the disputed instance plant value is May 10, the rule will be replaced by the fact:

$expert\_corn(plant) \leftarrow true$

If the disputed instance plant value is June 20, the rule will be replaced by the fact:

$\sim expert\_corn(plant) \leftarrow true$

Algorithm 4 is a formal representation of the resolution process. The DeLP program $\mathcal{P}$ is asked to produce an argumentation for each of the four crops to resolve the classification debate. If exactly one crop argumentation is successful then this crop is resolving the classification. Otherwise, the classification is left undecided. The answers to the four queries, corresponding to the four crops, are stored in the vector:

$$answer = \langle answer_{corn}, answer_{rice}, answer_{cotton}, answer_{soybean} \rangle$$

where each element consists of a pair $\langle b, \mathcal{T} \rangle$, where b can be *True* or *False* and $\mathcal{T}$ is the dialectical tree. The three possible configurations for an $answer_y$ pair are:

- $\langle True, \mathcal{T} \rangle$, if y is warranted
- $\langle False, \mathcal{T} \rangle$, if $\sim y$ is warranted
- $\langle False, null \rangle$, if nor y neither $\sim y$ are warranted

In line with DeLP interpreter answers (García & Simari, 2004), $\langle True, \mathcal{T} \rangle$ corresponds to YES, $\langle False, \mathcal{T} \rangle$ to NO and $\langle False, null \rangle$ to UNDECIDED. The fourth possible status UNKNOWN is ignored as it can arise exclusively when y is not found in the program $\mathcal{P}$.

Within the conflict resolution Algorithm 4, the *answer* vector is initialized on line 1. The pair $answer_y$ is built for each crop by a loop (lines 2-11). First, the algorithm tries to warrant y by building a dialectical tree having an *undefeated* labeled root (lines 3-5). If it succeeds, the answer is marked as positive. Contrary, it tries to warrant the complement $\sim y$ such that a negative answer can be inferred by a dialectical tree(lines 7-9). If nor y neither $\sim y$ are warranted, the answer is marked as negative (line 11). Once all the four answers are built, we check if there is exactly one that came up positive (line 12). If there is such an answer, then its crop class is considered the true class of the crop instance (line 13). Otherwise, the classification is left undecided (line 14).

Example 7. Consider a debatable instance which produces the following answer vector within Algorithm 4: $answer = \langle answer_{corn}, answer_{rice}, answer_{cotton}, answer_{soybean} \rangle$, where:

$$\begin{aligned} answer_{corn} &= \langle False, \mathcal{T}_{corn} \rangle \\ answer_{rice} &= \langle False, \mathcal{T}_{rice} \rangle \\ answer_{cotton} &= \langle True, \mathcal{T}_{cotton} \rangle \\ answer_{cotton} &= \langle False, null \rangle \end{aligned}$$

The algorithm returns *cotton* because it corresponds to the exactly one *True* answer $answer_{cotton}$. Both $answer_{corn}$ and $answer_{rice}$ produce False answers because their complement is warranted, producing dialectical trees having undefeated nodes as roots. The $answer_{soybean}$ can not produce any dialectical tree having an undefeated root, thus producing a False answer too.

6.2. Dialectical analysis of a debatable instance

This section explains the dialectical analysis approach on one example of a debatable instance from the crop dataset.

Instance 32 of the crops dataset is classified as *cotton* by the decision tree classifier and *soybean* by the neural network and the support vector machine classifiers. By employing the voting resolution strategy, *soybean* would be declared the winning class with two votes against one. However, the actual class of the instance is *cotton*, correctly pointed by the described DeLP inference.

The feature values of the debatable instance are $g = 2.02, m = 1.85, ndvi = -1.16$. Expert knowledge is augmented with the phenological properties, *plant* on *May 10, harvest* on *Oct 15* with a true value for *harvest_color_change*. The rules extracted from the three statistical classifiers are identical with the ones in Examples 3, 4, and 5. The classifier rules are merged with the expert rules defined in Section 5.2 to form the DeLP program.

Finally, four queries, one for each crop type, are executed:

Corn query returns a False answer because the complement of *corn* is warranted. The argument structure $\langle \mathcal{A}_1, \sim corn \rangle$ is produced by the h_{dt} classifier, which believes that this instance should not be classified as corn:

$$\mathcal{A}_1 = \left\{ \begin{array}{l} \sim corn \prec decision\_tree(g, m, ndvi) \\ decision\_tree(g, m, ndvi) \prec m > -0.01, \\ \quad ndvi \in [-1.31, 0.07] \end{array} \right\}$$

$\langle \mathcal{A}_1, \sim corn \rangle$ is defeated by $\langle \mathcal{A}_2, corn \rangle$ and $\langle \mathcal{A}_3, corn \rangle$, argument structures produced by the expert rules confirming that plant date, harvest date and harvest state are specific for corn:

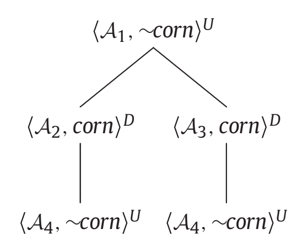
$$\mathcal{A}_2 = \left\{ \begin{array}{l} corn \prec expert\_corn(plant), expert\_corn(harvest) \\ expert\_corn(plant) \\ expert\_corn(harvest) \end{array} \right\},$$

$$\mathcal{A}_3 = \left\{ \begin{array}{l} corn \prec harvest\_color\_change \\ harvest\_color\_change \end{array} \right\}$$

$\langle \mathcal{A}_2, corn \rangle$ and $\langle \mathcal{A}_3, corn \rangle$ are in turn defeated by $\langle \mathcal{A}_4, \sim corn \rangle$, by the fact that green level does not fall in the expert defined range for corn:

$$\mathcal{A}_4 = \left\{ \begin{array}{l} \sim corn \prec \sim expert\_corn(g) \\ \sim expert\_corn(g) \end{array} \right\}$$

There is no other argument that can be constructed to defeat $\langle \mathcal{A}_4, \sim corn \rangle$, thus $\langle \mathcal{A}_1, \sim corn \rangle$ is reinstated. The dialectical tree that warranted $\sim corn$ is:



Rice query returns a False answer. The complement of *rice* is warranted by a sole argument structure $\langle \mathcal{A}_5, \sim rice \rangle$, produced by the strict fact that rice is not changing significantly the color at harvest.

$$\mathcal{A}_5 = \left\{ \begin{array}{l} \sim rice \leftarrow harvest\_color\_change \\ harvest\_color\_change \end{array} \right\},$$

The corresponding dialectical tree is formed by a single node:

$$\langle \mathcal{A}_5, \sim rice \rangle^U$$

Cotton query produces a True answer because *cotton* is warranted. The argument structure $\langle \mathcal{A}_6, cotton \rangle$ is produced by the h_{dt} which believes that this instance should classified as cotton:

$$\mathcal{A}_6 = \left\{ \begin{array}{l} cotton \prec decision\_tree(g, m, ndvi) \\ decision\_tree(g, m, ndvi) \prec m > -0.01, \\ \quad ndvi \in] -1.31, 0.07] \end{array} \right\}$$

The h_{nn} and h_{svm} classifiers argue that the instance should be not classified as cotton, producing argument structures $\langle \mathcal{A}_7, \sim cotton \rangle$ and $\langle \mathcal{A}_8, \sim cotton \rangle$ that defeat the initial argument structure $\langle \mathcal{A}_6, cotton \rangle$:

$$\mathcal{A}_7 = \left\{ \begin{array}{l} \sim cotton \prec neural\_net(g, m, ndvi) \\ neural\_net(g, m, ndvi) \prec 0.21 * m - 1.44 * ndvi > 1.83 \end{array} \right\},$$

$$\mathcal{A}_8 = \begin{Bmatrix} \sim cotton \prec svm(g, m, ndvi) \\ svm(g, m, ndvi) \prec g > 0.001, ndvi \leq -1.1 \end{Bmatrix}$$

$\langle \mathcal{A}_9, cotton\rangle$ argument structure is produced by the expert rules and it defeats the h_{nn} and h_{svm} arguments $\langle \mathcal{A}_7, \sim cotton\rangle$ and $\langle \mathcal{A}_8, \sim cotton\rangle$. $\langle \mathcal{A}_9, cotton\rangle$ is derived from the fact that plant and harvest dates fit in the dates defined by the expert for planting and harvesting cotton:

$$\mathcal{A}_9 = \begin{Bmatrix} cotton \prec expert\_cotton(plant), expert\_cotton(harvest) \\ expert\_cotton(plant) \\ expert\_cotton(harvest) \end{Bmatrix},$$

Since there are no defeaters for $\langle \mathcal{A}_9, cotton\rangle$ the dialectical inference stops. The corresponding dialectical tree is:

$$\begin{array}{ccc} & \langle \mathcal{A}_6, cotton\rangle^U & \\ \langle \mathcal{A}_7, \sim cotton\rangle^D & & \langle \mathcal{A}_8, \sim cotton\rangle^D \\ | & & | \\ \langle \mathcal{A}_9, cotton\rangle^U & & \langle \mathcal{A}_9, cotton\rangle^U \end{array}$$

Soybean query produces a False answer because neither *soybean* nor $\sim$ *soybean* are warranted. DeLP inference engine produces five dialectical trees all having the root labeled as *Defeated*. Since dialectical analysis can not prove *soybean* as the class of the instance the query returns *False*.

Two of the *soybean* dialectical trees have a similar inference process, differing only by the argument structures corresponding to the root nodes:

$$\begin{array}{cccc} & \langle \mathcal{A}_{10}, soybean\rangle^D & & \\ & | & & \\ & \langle \mathcal{A}_{11}, \sim soybean\rangle^U & & \\ \langle \mathcal{A}_{12}, soybean\rangle^D & & \langle \mathcal{A}_{13}, soybean\rangle^D & \\ | & & | & \\ \langle \mathcal{A}_{14}, \sim soybean\rangle^U & & \langle \mathcal{A}_{14}, \sim soybean\rangle^U & \end{array} \qquad \begin{array}{cccc} & \langle \mathcal{A}_{15}, soybean\rangle^D & & \\ & | & & \\ & \langle \mathcal{A}_{11}, \sim soybean\rangle^U & & \\ \langle \mathcal{A}_{12}, soybean\rangle^D & & \langle \mathcal{A}_{13}, soybean\rangle^D & \\ | & & | & \\ \langle \mathcal{A}_{14}, \sim soybean\rangle^U & & \langle \mathcal{A}_{14}, \sim soybean\rangle^U & \end{array}$$

The root node arguments $\langle \mathcal{A}_{10}, soybean\rangle$ and $\langle \mathcal{A}_{15}, soybean\rangle$ are an outcome of the h_{svm} and h_{nn} learners, which both believe the instance should be classified as soybean:

$$\mathcal{A}_{10} = \begin{Bmatrix} soybean \prec svm(g, m, ndvi) \\ svm(g, m, ndvi) \prec g > 0.001, ndvi \leq -1.1 \end{Bmatrix},$$

$$\mathcal{A}_{15} = \begin{Bmatrix} soybean \prec neural\_net(g, m, ndvi) \\ neural\_net(g, m, ndvi) \prec 0.21 * m - 1.44 * ndvi > 1.83 \end{Bmatrix}$$

$\langle \mathcal{A}_{10}, soybean\rangle$ and $\langle \mathcal{A}_{15}, soybean\rangle$ are disputed by the h_{dt} classifier, which believes that the instance should not be classified as soybean, based on the argument $\langle \mathcal{A}_{11}, \sim soybean\rangle$:

$$\mathcal{A}_{11} = \begin{Bmatrix} \sim soybean \prec decision\_tree(g, m, ndvi) \\ decision\_tree(g, m, ndvi) \prec m > -0.01, \\ \quad ndvi \in]-1.31, 0.07] \end{Bmatrix}$$

The decision tree argument structure $\langle \mathcal{A}_{11}, \sim soybean\rangle$ is defeated by the expert derived arguments $\langle \mathcal{A}_{12}, soybean\rangle$ and $\langle \mathcal{A}_{13}, soybean\rangle$, which state that the expert green, moisture, ndvi levels and harvesting color state indicate soybean:

$$\mathcal{A}_{12} = \begin{Bmatrix} soybean \prec expert\_soybean(g), expert\_soybean(m), \\ \quad expert\_soybean(ndvi) \\ expert\_soybean(g) \\ expert\_soybean(m) \\ expert\_soybean(ndvi) \end{Bmatrix},$$

$$\mathcal{A}_{13} = \begin{Bmatrix} soybean \prec harvest\_color\_change \\ harvest\_color\_change \end{Bmatrix}$$

$\langle \mathcal{A}_{12}, soybean\rangle$ and $\langle \mathcal{A}_{13}, soybean\rangle$ are in turn defeated by the expert argument structure $\langle \mathcal{A}_{14}, \sim soybean\rangle$, pointing that planting date is outside of *soybean* planting time-frame:

$$\mathcal{A}_{14} = \begin{Bmatrix} \sim soybean \prec \sim expert\_soybean(plant) \\ \sim expert\_soybean(plant) \end{Bmatrix}$$

There are no argument structures that can be posed to $\langle \mathcal{A}_{14}, \sim soybean\rangle$, thus the argument is undefeated. The dialectical analysis ends for the two trees which started with arguments from the h_{svm} and h_{nn} classifiers.

Two more soybean dialectical trees with *Defeated* root nodes are produced based on expert argument structures $\langle \mathcal{A}_{12}, soybean\rangle$ and $\langle \mathcal{A}_{14}, \sim soybean\rangle$. The first argument states that soybean is a possible match because green, moisture, and ndvi levels correspond to soybean. The second argument opposes to soybean because the planting date does not fall in the expert defined time-frame. The two dialectical trees are:

$$\begin{array}{cc} \langle \mathcal{A}_{12}, soybean\rangle^D & \langle \mathcal{A}_{14}, \sim soybean\rangle^D \\ | & | \\ \langle \mathcal{A}_{14}, \sim soybean\rangle^U & \langle \mathcal{A}_{12}, soybean\rangle^U \end{array}$$

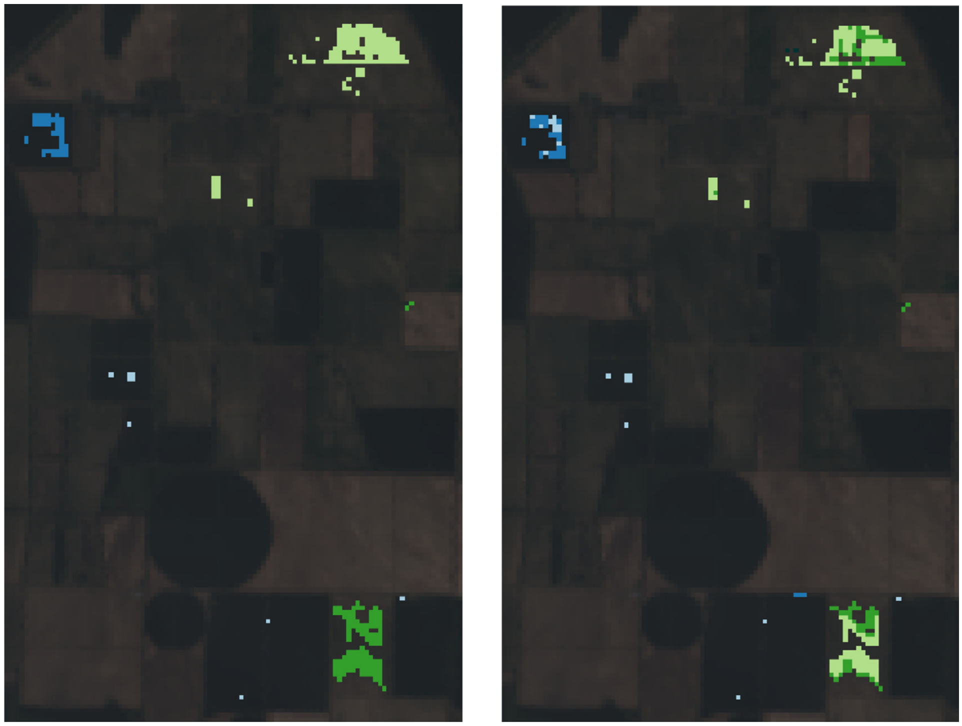
The last soybean dialectical tree fails to warrant $\sim$ *soybean* by starting with the argument structure induced by the decision tree learner $\langle \mathcal{A}_{11}, \sim soybean\rangle$. The h_{svm} classifier defeats $\langle \mathcal{A}_{11}, \sim soybean\rangle$ by using its argument $\langle \mathcal{A}_{10}, soybean\rangle$. The argument conveyed by h_{svm} is in turn defeated by the expert using $\langle \mathcal{A}_{14}, \sim soybean\rangle$ arguing that planting date is outside of soybean planting period. Finally $\langle \mathcal{A}_{14}, \sim soybean\rangle$ is defeated by $\langle \mathcal{A}_{12}, soybean\rangle$ expert argument which says that green, moisture and ndvi levels are specific for soybean. The dialectical tree is:

$$\begin{array}{c} \langle \mathcal{A}_{11}, \sim soybean\rangle^D \\ | \\ \langle \mathcal{A}_{10}, soybean\rangle^U \\ | \\ \langle \mathcal{A}_{14}, \sim soybean\rangle^D \\ | \\ \langle \mathcal{A}_{12}, soybean\rangle^U \end{array}$$

Out of the four queries: *corn, cotton, rice* and *soybean*, the only *True* answer is outputted by the *cotton* query, which corresponds to the actual class of the instance.

6.3. Experimental results

The results are presented first from the conflict resolution perspective, evaluating the resolution methods over the set of conflicting instances. Then, classification results are presented for the entire test crops dataset.



(a) Argumentation conflict resolution (99% accuracy). (b) Voting conflict resolution (58.1% accuracy)

Fig. 7. Classification results on conflict set Γ (306 instances). The color codes for *corn, rice, cotton* and *soybean* are light blue, blue, light green and green respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Conflict resolution accuracy, precision and recall on the conflict set Γ of 306 instances.

Method	Accuracy	Corn		Rice		Cotton		Soybean	
		P	R	P	R	P	R	P	R
Voting	58.1	45	75	90.6	72.5	57.6	70	50	36.8
DeLP	99	100	75	100	100	100	100	100	100

The set of conflicting instances is formed by 306 cases in which the statistical classifiers gave conflicting predictions. The conflict set accounts for 7% of the test dataset. Table 3 lists the results of conflict resolution methods employed on the conflict set. Resolving conflicts by voting gave an accuracy of 58.1%. On the other hand, resolving conflicts by DeLP argumentation, making use of expert and base classifiers knowledge, a much higher accuracy of 99% was obtained. Fig. 7 displays the classification result for each pixel of the conflict set.

As the voting system does not use human knowledge, while our ensemble-delp resolutor uses rules extracted from the agricultural domain, the improvement from 58.1% to 99% accuracy on the conflict set quantifies the impact of human knowledge in the classification. Hence, the difference of 99%-58.1% = 40.9% percents is due to the expert rules and the conflict resolution strategy of our argumentation method. This percent of 40.9% increasing represents the quantification of the relevance of the human knowledge in this domain.

DeLP conflict resolution produced a precision of 100% for each crop class, all predicted values being correctly classified. DeLP left three conflicting instances un-classified thus not achieving a perfect recall for *corn* (75%) class. Table 4 presents the confusion matrix for DeLP classification resolution on the conflict set. Three *corn* instances remained unclassified because DeLP inferred that none of the four classes is a match for these instances.

Table 4
Confusion matrix for DeLP resolution classifier on the conflict set Γ of 306 instances.

	Corn	Rice	Cotton	Soybean	Undecided	Recall
Corn	9	0	0	0	3	75
Rice	0	40	0	0	0	100
Cotton	0	0	140	0	0	100
Soybean	0	0	0	114	0	100
Precision	100	100	100	100		

Table 5
Confusion matrix for voting resolution classifier on the conflict set Γ of 306 instances.

	Corn	Rice	Cotton	Soybean	Recall
Corn	9	3	0	0	75
Rice	11	29	0	0	72.5
Cotton	0	0	98	42	70
Soybean	0	0	72	42	36.8
Precision	45	90.6	57.6	50	

Voting resolution method produced lower precision scores on the conflict set, especially for *corn* (45%), *soybean* (50%), and *cotton* (57.6%). Table 5 presents the confusion matrix for voting classification resolution on the conflict set. The low precision for *corn* is caused by incorrectly classifying more than half *corn* instances as *rice*. The low precision for *soybean* is caused by incorrectly classifying half of the *soybean* instances as *cotton*. DeLP resolution was able to settle all these confusions by making use of expert knowledge. For example *corn* is differentiated from *rice* by the significant change in color when harvested, while planting season for *soybean* can overpass with one month the *cotton* planting season.

Table 6 lists the evaluation of all classification methods on the test dataset (4,342 instances), best results are shown in bold face. Due to better conflict resolution, the Ensemble using DeLP produced a higher accuracy (98.4%) than the ensemble using voting (95.5%).

Table 6
Classification accuracy, precision and recall per each classification method on the test dataset (4,342 instances).

Classification Method	Acc.	Corn		Rice		Cotton		Soybean	
		P	R	P	R	P	R	P	R
Ensemble Voting	95.5	99.5	99.8	95.1	80.8	84.9	93.1	89.1	77.4
Ensemble DeLP	**98.4**	**99.8**	**99.9**	**100**	**95.8**	**93.3**	**98.6**	**98**	**90**

McNemar's test is employed for showing the statistical significance of the classification methods. (Foody, 2004) advocates for using the McNemar's test for remote sensing to compare classifiers built by using the same dataset. To compare the performance between two classification methods, a value z is computed according to the formula:

$$z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \tag{3}$$

where f_{12} represents the count of instances correctly classified by the first classifier and wrongly classified by the second, while f_{21} represents the count of instances correctly classified by the second classifier and wrongly classified by the first. According to (Foody, 2004), when $|z| > 1.96$ there is a difference in accuracy at a confidence level of 95%. For evaluating our classifiers, we computed the z score for the ensemble using argumentation and the ensemble using voting resolution. The z value was determined to be 11.1 (since $f_{12} = 125$ and $f_{21} = 0$), indicating a positive significance and thus a superior accuracy of the argumentation over the voting conflict resolution.

7. Discussion and research directions

7.1. Related approaches

We organise the discussion section based on two particularities of our method: i) integration of machine learning with argumentation and ii) rule extraction from base classifiers.

Argumentation and machine learning. Comparing to voting-based ensemble learning, the human knowledge has lead to an increase of the classification accuracy from 58.1% to 99% on the conflict set. Moreover, the DELP argumentation machinery let us also include the arguments provided by the base learners in the reasoning process. We managed to ensemble the classification knowledge encapsulated by each learner intending to develop an argumentation framework for classification. From this perspective, our approach can be seen as a contribution on the research path opened by Amgoud and Serrurier (2008); Hao et al. (2015); Wardeh et al. (2012b), with two advancements: a realistic scenario and inclusion of rules mined from base learners in the dialectical reasoning.

In argumentation based multi-agent joint learning (Fomina, Morosin, & Vagin, 2014), argumentation is used to ensemble multiple classifiers (Xu et al., 2015). Differently from argumentation based joint learning (Xu et al., 2015), our scenario proposes the extraction of rules from a number of diverse statistical classifiers that learn to represent the data in different ways rather than using the same rule mining mechanism on different subsets of data. Moreover, we maintain the use of base classifiers in the decision making process contrary to building a single global knowledge base for classification that makes the base classifiers unnecessary. The scope of the argumentation framework is restricted to the situations where the base classifiers do not reach consensus.

Several collaborative knowledge models (Dalkey, Brown, & Cochran, 1969; Hao, Yao, Liu, & Wang, 2014; Wardeh, Coenen, & Bench-Capon, 2012a; Yao, Xu, Li, & Qi, 2012) integrate argumentation and machine learning. PISA model is used in Wardeh et al. (2012a) to solve the multi-classification problem using argumentation from experience. The resulting Prism method supports collaborative classification based on multiple classifiers and distributed data sets. The Arguing Prism extension (Hao et al., 2014) introduces multi-agent dialogue games to support classification in distributed environments. Arena dialectical analysis model (Yao et al., 2012) is used to allow agents to collaboratively classify individuals. Delphy collaborative method (Dalkey et al., 1969) assumes three learners. In the first step, knowledge is extracted from each learner. In the second step, knowledge of the other two learners is included in the third classifier. At the end, the learning algorithm is rerun, hoping that the new knowledge increases the accuracy of the classification. From a larger perspective, the above approaches are related to the knowledge spiral model. This collaborative model extracts organizational knowledge from the common knowledge of the members through communication. In line with Hao et al. (2014), we exploit the capabilities of argumentation technology for conflict resolution for different learners. Differently from Hao et al. (2014), we focus on knowledge extraction from classifiers, instead of dialogue games for exchanging arguments in Hao et al. (2014).

Integration of learning agents and argumentative agents is another promising line of research (Ontanón & Plaza, 2011; Xu et al., 2015) In Xu et al. (2015), agents perform data mining to construct classifiers independently at first. The assumption is that each agent has a different data set, hence a possible conflicting viewpoint with the other agents. In the second step, classifiers assess their individual knowledge by using argumentation. In the third step, global knowledge is extracted to generate an ensemble classifier. Integration of multi-agent inductive learning and argumentation is proposed in Ontanón and Plaza (2011). The source of inconsistency comes both from i) agents experience, represented by different training set and ii) different inductive method performed by each agent. It has been shown in Ontañón, Dellunde, Godo, and Plaza (2012) that inductive theories achieved by multi-agent induction plus argumentation are precisely the same as the inductive theories extracted by a single agent with all data.

Neural networks are augmented with defeasible logic programming in Gómez and Chesnevar (2004). The approach uses a Fuzzy ART neural network, a DeLP program, and the data for the instance to be classified. If the classification is not achieved by the Fuzzy ART neural network, the DeLP program performs a dialectical analysis to decide the class of the given instance. Given the instance and an assumption regarding the class, the system accepts the assumption (positive), rejects it negative (neg) or let it undecided (undec) if the encapsulated knowledge does not suffice for a warranted decision. Differently, in our case we apply the argumentation machinery on conflicts arising from an ensemble learner. Our strategy was driven by the goal to extract compact knowledge from the neural network in order to feed the argumentation system. The main conceptual difference is that we integrate expert knowledge with knowledge extracting from base learners in order to perform dialectical argumentation.

Similar to our approach, the system in Guerrero et al. (2013) exploits human knowledge to increase detection accuracy. The hybrid system in Guerrero et al. (2013) processes images taken from the ground, while we analyze remote images. Consequently, Guerrero et al. (2013) uses different agronomic vegetation indexes like excess green, color index of vegetation extraction or excess green minus excess red, while we use NDVI or short wave infrared. The constraint of Guerrero et al. (2013) to classify in real time is not applicable to our task.

The individuals that are likely to be misclassified are labeled as "warnings", which require a deeper analysis Luaces, Rodrigues, Meira, and Bahamonde (2011). In our case, this deeper analysis is performed within an argumentation framework, for the individuals labeled as "debatable". A side effect of both solutions is that when predicting a class for an individual which is not a "warning" in Luaces et al. (2011), respectively in the conflict set in our case, the confidence level of these classifications increases.

Rule extraction from machine learning. Our solution is based on defeasible rules automatically mined from three machine learning methods: neural networks, decision trees and support vector machines

Conceptual instrumentation for rule mining from neural networks includes neuro-symbolic integration (Hatzilygeroudis & Prentzas, 2015), formal concept analysis (Hasanah, Imai, & Nobuhara, 2010), decompositional approaches like TACO algorithm (Özbakir, Baykasoğlu, Kulluk, & Yapıcı, 2009), or formalism for extracting fuzzy rules (Kulluk, Özbakır, & Baykasoğlu, 2013). Knowledge in the form of heuristics is used to determine the optimum values of neural networks parameters (Kavzoglu & Mather, 2003). Guidelines to design artificial neural networks in remote sensing image classification have been also formalized (Kavzoglu & Mather, 2003). Our approach to tune the elements of the neural network is based on skeletal pruned neural networks (Setiono & Liu, 1997) aiming to derive simpler classification rules. Based on this approach, we select the configuration with a minimal number of edges from input to hidden neurons and a maximal accuracy.

Descriptive neural networks (DNNs) embed rules that have been discovered from previously trained networks (Guven, 2011). As DNN is a neural network with domain knowledge, the classifications are complemented with some form of explanation (Yao, 2005). Instead, after extracting rules from the neural network we rely on expert knowledge and argumentation to increase accuracy and on dialectical trees to explain the classification decision.

Decision trees classifiers produce classification rules by their very nature. Decision rules can be reduced to smaller sets by pruning subtrees from decision tree paths and replace them with leaves (Quinlan, 1987). Instead, we rely on using the entire set of rules derived from the decision tree as the tree contains only three decision nodes.

Rule extraction from support vector machines is performed in Núñez, Angulo, and Català (2002) with an iterative process for identifying class centroids and define classification rules on top by using ellipsoids equations. Non-overlapping rules are extracted from linear SVM with an iterative process and a constrained optimization problem of few variables in Fung, Sandilya, and Rao (2005) . SVM rules can also be induced by using an additional white box learner. An eclectic rule extraction method has been presented in Barakat and Diederich (2005), where support vectors are classified by the C4.5 decision tree algorithm. In our approach we follow the eclectic method proposed by Barakat and Diederich (2005), employing the CART (Breiman et al., 1984) algorithm for the decision tree learner.

Inductive Logic Programming (Bayoudh, Roux, Richard, & Nock, 2015) has been used to extract 158 classification rules were extracted from 3 diachronic land cover maps. For 38 classes, a 10-fold cross-validation gave significant average values of 84.62%, 99.57% and 77.22% for classification, accuracy, specificity and sensitivity, respectively. These rules were formalized in first order logic, which makes them easily understandable by non-experts. The proposed instrumentation was used to monitor changes in land cover of the French Guiana coastline. Our 54 extracted rules from machine learning were augmented with 43 strict and defeasible expert rules. The rule set was formalized in defeasible logic in order to allow argumentative reasoning on them. The numerical results are not comparable since we focus on 4 classes instead of 38 in Bayoudh et al. (2015).

7.2. Research directions

The solution proposed in this paper for crop classification lays the groundwork for several extensions:

- Using a more expressive argumentation model, such as weighted argumentation systems (Dunne, Hunter, McBurney, Parsons, & Wooldridge, 2011) or probabilistic argumentation (Haenni, 2009).
- Exploiting the available formal knowledge in the crop domain, by importing various agricultural ontologies in the expert knowledge base (Beneventano, Bergamaschi, Sorrentino, Vincini, & Benedetti, 2015; Bonacin, Nabuco, & Junior, 2016; Caracciolo et al., 2013; Lauser et al., 2008; Li et al., 2013; Wang, Wang, Wang, Yuan, & Zhang, 2015).
- Investigating the behavior of the system in case of large numbers of base learners, towards large scale argumentation on crowds of learners (Lease, 2011).

More expressive argumentation models. A straightforward continuation of our work would be using weighted argumentation systems (Dunne et al., 2011) for conflict resolution. In this paper, defeasible argumentation was proposed as an improvement for voting in ensemble learning. In the same line, weighted argumentation systems might be an improvement of weighted voting method, in which based classifiers that are most sure vote with more conviction. That is, the surer the classifier, the stronger the argument.

In the same line, a possible improvement to our DeLP preference criteria is to match the order relation of defeasible rules generated by classifiers to their quality, expressed in term of classification accuracy. In this way, extracted rules gain more significance during the DeLP inference process.

A probabilistic argumentation system on top of an ensemble learner will benefit from the posterior probabilities of the classification decisions. Each base learner outputs a vector of continuous-valued measures representing estimates of class posterior probabilities that support for the possible classification hypotheses. The rationale behind computing the posterior probabilities of the decision is that more base learners agree with their classification, the more confidence in the ensemble's decision. As it has been shown that a properly trained ensemble decision is usually correct if its confidence is high (Muhlbaier, Topalis, & Polikar, 2005), a probabilistic argumentation system would bring benefits when the confidence is low. That is when the disagreement is high in the argumentation knowledge base. A relevant question would be how much the learners positions differ from each other in order to quantify disagreement in an argument-based system on top of an ensemble learner.

Complementarity in advantages and disadvantages of the combined methods are considered the basis for the success of hybrid systems (Hatzilygeroudis & Prentzas, 2015). In this line, more expressive argumentation models for ensemble learning can: i) modeling domain knowledge with the complete instrumentation provided by defeasible logic: defeasible rules, strict rules, defeaters, priorities; ii) evaluate the knowledge extracted from base classifiers to select only relevant global knowledge; iii) guide the process of feature selection in order to apply each base classifier only on relevant features; iv) assess the compatibility of a classifier with the input data by modeling machine learning heuristics as structured arguments An example of such meta-argument could be: "the k-nearest neighbors classifier does not work well on unbalanced data sets". This can be useful when defining priorities among rules extracted from different base learners.

Importing knowledge from agricultural repositories. Various online repositories contain structured knowledge on crop sciences. The rationale of supporting agricultural repositories is related to share research data or evaluation data sets. Efficient sharing and integration of agricultural knowledge is the main objective of agricultural semantic interoperability based on agricultural ontologies. As these ontologies can include local knowledge, applications can focus on different crops in different regions. The usage of these ontologies is twofold. One goal is to help farmers retrieve agricultural and practical technical information easily. A second goal is to facilitate local knowledge elicitation from farmers in different regions. These knowledge sources covers both general agricultural vocabularies (Agrovoc (Caracciolo et al., 2013), NAL (Lauser et al., 2008), OntoAgroHydro (Bonacin et al., 2016)) and specific crop ontologies like pepper (Li et al., 2013), citrus (Wang et al., 2015), or CEREALAB (Beneventano et al., 2015).

Comprehensive agricultural thesaurus such as Agrovoc (Caracciolo et al., 2013) or NAL (Lauser et al., 2008) are used by specialized knowledge-based applications in the agriculture or food domain. The Agrovoc (Agriculture with Vocabulary) is a controlled vocabulary developed by experts for the Food and Agriculture Organization of the United Nations. It contains 32,000 concepts available in 23 languages (Caracciolo et al., 2013), The Agrontology is an example of a specialized ontology that provides a set of domain properties for Agrovoc. The NAL Agricultural Thesaurus covers over 120,000 terms, and 57,000 cross references (Lauser et al., 2008). The main goal is to improve retrieval of agricultural information. The Rice Thesaurus is an example of an online tool fo rice-related terminology, that exploits the NAL vocabulary. Both Agrovoc and NAL are available as Linked Open Data to facilitate reused in various applications. Also a general ontology, OntoAgroHidro includes 8500 concepts and instances about impacts of agricultural activities and climatic changes on water resource (Bonacin et al., 2016).

Specific ontologies in the crop domain aim to facilitate community sharing crop-related information. The pepper ontology in Li et al. (2013) contains both domain knowledge and task knowledge related to pepper cultivation good practices. Domain knowledge models static information during the growing of pepper, such as soil, seed, and agricultural machines. Task-related knowledge is compliant with crop cultivation standards when formalizing plant processes like soil selection, seed selection, fertilization and irrigation. The citrus ontology in Wang et al. (2015) includes concepts related to effects of terrain on fertilization, and irrigation. Three hilly citrus decision services were developed based on this ontology. The nutrient imbalance service had a 98% accuracy, and the accuracy of our irrigation and drainage services reached 94%. The CEREALAB database stores both genotypic and phenotypic data specifically designed for plant breeding. The database has been semantically annotated (Beneventano et al., 2015) with concepts from Agrovoc, aiming to deploy data on Linked Open Data cloud infrastructure.

Instead of exploiting these agricultural knowledge bases, we formalized expert knowledge from the USDA Agricultural Handbook (NASS, 2010). However, several technical instrumentations support the integration of ontological knowledge in our solution. First, we need to integrate defeasible rules on top of ontologies. Second, we need to include concepts related to satellite images in the existing crop ontologies. Relevant starting points for the first task are the current advancements in hybrid knowledge bases that compose ontologies and rules (Slota, Leite, & Swift, 2015) or the existing algorithms for inducing defeasible rules (Johnston & Governatori, 2003). Relevant to the second task is the support provided by the Linked Open Data technology or merging ontologies.

Large scale argumentation on crowd of learners.

A possible direction based on our approach is related to the current usage of the wisdom of crowds in machine learning (Lease, 2011). Relevant questions in the context of argumentation on top of a large ensemble learner would be: How does the argumentative base behave when rules are extracted from large numbers of learners? Do wisdom or higher accuracy emerge from a large number of small argumentative debates? How does the conflict set depends on the number of learners? There are also some assumptions for the wisdom of crowds, namely diversity, independence, and decentralization. The ensemble system should be analyzed from these three dimensions: diversity of the based classifiers, independence of the learners, and decentralization. Similarly, the corresponding knowledge base should quantify the independence and diversity of defeasible arguments. The efficiency of an ensemble learner also rests on the diversity of the base learners within the ensemble. Here diversity may be interpreted as making classifiers to manifest variety in order to avoid over-fitting. Brown (2010) distinguishes between explicit and implicit diversity. Implicit diversity occurs when each base classifier is feed with a different random subset of the training data (i.e., different instances in case of bagging, different features in case of random subspaces). Explicit diversity occurs when a metric is used to quantify that each learner is substantially different from each other (i.e, different weights for instances in case of the boosting method). For a comprehensive review on this topic, the reader is referred to Brown, Wyatt, Harris, and Yao (2005).

In this paper we did not consider more complicated practical dimensions which are very significant in crop classification, like multiple experts elicitation methods in agriculture (Léger & Naud, 2009), automatic agricultural knowledge generation from web resources (Wei, Wang, Hu, & Xue, 2012), unbalanced crop data sets (Spilke, Piepho, & Hu, 2005), crop forecasting (Johnson, Hsieh, Cannon, Davidson, & Bédard, 2016), or real time classification (Guerrero et al., 2013).

8. Conclusions

We developed a solution for conflict resolution is ensemble learning, and we successfully apply this solution for crop classification in the agriculture domain. Our hybrid system merges machine learning and symbolic argumentation with the scope of improving the classification of four crop classes in remote sensing: corn, soybean, rice and cotton. The machine learning pursuit is represented by an ensemble learner composed of three discriminative models: decision tree, neural network and support vector machine. Conflicting situations, characterized by instances for which base classifiers do not reach consensus, are resolved by using a symbolic argumentation process. Within the argumentation process, a dialectical analysis is performed on symbolic rules extracted from the base classifiers and knowledge defined by an expert. Expert knowledge guides the resolution process to reach definite decisions within a closed context defined from morphological and phenological profiles of the four crops. The proposed solution improved both the accuracy of resolution of conflicting instances and the accuracy of the ensemble learner as a whole. In conclusion, our argument-based conflict resolutor proved to be more effective than voting-based resolutor in ensemble learning. Moreover, the experiments clearly indicated the high impact of expert knowledge on resolving debatable classes in the agriculture domain.

The presented approach has several contributions in regards to the field of Expert and Intelligent Systems. To the best of our knowledge, this is the first approach that combines ensemble learning and argumentation in the agricultural domain. We developed a method for extracting defeasible rules from base learners to facilitate the integration of expert rules in the decision process. Moreover, the advantages the argumentation machinery brought on top of an ensemble classifier are: First, arguments helped us to introduce and use human knowledge during classification. Second, our experiments proved that argumentative reasoning represents a means to conflict resolution in ensemble learning, instead of voting-based methods. Third, by combining arguments with machine learning we managed to handle different types of information in a uniform way. Forth, argumentation increased transparency on our hybrid intelligent system. Hence, we consider that the conceptual instrumentation presented in this work can be used to take decisions in domains characterized by high data availability, robust expert knowledge, and a need for justifying the rationale behind decisions.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

We thank the reviewers for their valuable comments. There is no funding source for this research. The research has been conducted within the Intelligent Research Group at Technical University of Cluj-Napoca, Romania.

References

Aitkenhead, M., & Aalders, I. (2008). Classification of Landsat Thematic Mapper imagery for land cover using neural networks. *International Journal of Remote Sensing, 29*(7), 2075–2084.

Amgoud, L., & Serrurier, M. (2008). Agents that argue and explain classifications. *Autonomous Agents and Multi-Agent Systems, 16*(2), 187–209.

Bahrammirzaee, A. (2010). A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing and Applications, 19*(8), 1165–1195.

Barakat, N., & Diederich, J. (2005). Eclectic rule-extraction from support vector machines. *International Journal of Computational Intelligence, 2*(1), 59–62.

Bayoudh, M., Roux, E., Richard, G., & Nock, R. (2015). Structural knowledge learning from maps for supervised land cover/use classification: Application to the monitoring of land cover/use maps in french guiana. *Computers and Geosciences, 76*, 31–40. http://dx.doi.org/10.1016/j.cageo.2014.08.013.

Bedi, P., & Vashisth, P. (2014). Empowering recommender systems using trust and argumentation. In *Information sciences: 279* (pp. 569–586).

Bench-Capon, T. J. M., & Dunne, P. E. (2007). Argumentation in artificial intelligence. *Artificial Intelligence, 171*(10-15), 619–641. http://dx.doi.org/10.1016/j.artint.2007.05.001.

Beneventano, D., Bergamaschi, S., Sorrentino, S., Vincini, M., & Benedetti, F. (2015). Semantic annotation of the CEREALAB database by the AGROVOC linked dataset. *Ecological Informatics, 26*, 119–126.

Bonacin, R., Nabuco, O. F., & Junior, I. P. (2016). Ontology models of the impacts of agriculture and climate changes on water resources: Scenarios on interoperability and information recovery. *Future Generation Computer Systems, 54*, 423–434. http://dx.doi.org/10.1016/j.future.2015.04.010.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Briguez, C. E., Budán, M. C., Deagustini, C. A., Maguitman, A. G., Capobianco, M., & Simari, G. R. (2014). Argument-based mixed recommenders and their application to movie suggestion. *Expert Systems with Applications, 41*, 6467–6482.

Brown, G. (2010). Ensemble learning. In *Encyclopedia of machine learning* (pp. 312–320). Springer.

Brown, G., Wyatt, J., Harris, R., & Yao, X. (2005). Diversity creation methods: a survey and categorisation. *Information Fusion, 6*(1), 5–20. http://dx.doi.org/10.1016/j.inffus.2004.04.004. Diversity in Multiple Classifier Systems

Caracciolo, C., Stellato, A., Morshed, A., Johannsen, G., Rajbhandari, S., Jaques, Y., et al. (2013). The AGROVOC linked dataset. *Semantic Web, 4*(3), 341–348.

Chow, H. K., Siu, W., Chan, C.-K., & Chan, H. C. (2013). An argumentation-oriented multi-agent system for automating the freight planning process. *Expert Systems with Applications, 40*, 3858–3871.

Cruz-Ramírez, M., Hervás-Martínez, C., Jurado-Expósito, M., & López-Granados, F. (2012). A multi-objective neural network based method for cover crop identification from remote sensed data. *Expert Systems with Applications, 39*(11), 10038–10048.

Dalkey, N. C., Brown, B. B., & Cochran, S. (1969). *The Delphi method: An experimental study of group opinion*: 3. Rand Corporation Santa Monica, CA.

Deagustini, C. A., Dalibon, S. E. F., Gottifredi, S., Falappa, M. A., Chenevar, C. I., & Simari, G. R. (2013). Relational databases as a massive information source for defeasible argumentation. *Knowledge-Based Systems, 51*, 93–109.

DeFries, R., & Townshend, J. (1994). NDVI-derived land cover classifications at a global scale. *International Journal of Remote Sensing, 15*(17), 3567–3586.

Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple classifier systems* (pp. 1–15). Springer.

Dunne, P. E., Hunter, A., McBurney, P., Parsons, S., & Wooldridge, M. (2011). Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artificial Intelligence, 175*(2), 457–486.

El Hajj, M., Bégué, A., Guillaume, S., & Martiné, J.-F. (2009). Integrating spot-5 time series, crop growth modeling and expert knowledge for monitoring agricultural practices - the case of sugarcane harvest on reunion island. *Remote Sensing of Environment, 113*(10), 2052–2061.

Fomina, M., Morosin, O., & Vagin, V. (2014). Argumentation approach and learning methods in intelligent decision support systems in the presence of inconsistent data. In *14th international conference on computational science (iccs 2014): 29* (pp. 1569–1579).

Foody, G. M. (2004). Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering & Remote Sensing, 70*(5), 627–633.

Friedl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment, 61*(3), 399–409.

Fung, G., Sandilya, S., & Rao, R. B. (2005). Rule extraction from linear support vector machines. In *Proceedings of the eleventh acm sigkdd international conference on knowledge discovery in data mining* (pp. 32–40). ACM.

García, A. J., & Simari, G. R. (2004). Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming, 4*(1+ 2), 95–138.

Gómez, S. A., & Chesnevar, C. I. (2004). A hybrid approach to pattern classification using neural networks and defeasible argumentation.. In *Flairs conference* (pp. 393–398).

Gómez, S. A., Chesñevar, C. I., & Simari, G. R. (2013). ONTOarg: A decision support framework for ontology integration based on argumentation. *Expert Systems with Applications, 40*, 1858–1870.

Guerrero, J., Guijarro, M., Montalvo, M., Romeo, J., Emmi, L., Ribeiro, A., et al. (2013). Automatic expert system based on images for accuracy crop row detection in maize fields. *Expert Systems with Applications, 40*(2), 656–664. http://dx.doi.org/10.1016/j.eswa.2012.07.073.

Guven, A. (2011). A multi-output descriptive neural network for estimation of scour geometry downstream from hydraulic structures. *Advances in Engineering Software, 42*(3), 85–93. http://dx.doi.org/10.1016/j.advengsoft.2010.12.005.

Gómez, S. A., Goron, A., Groza, A., & Letia, I. A. (2016). Assuring safety in air traffic control systems with argumentation and model checking. *Expert Systems with Applications, 44*, 367–385. http://dx.doi.org/10.1016/j.eswa.2015.09.027.

Haenni, R. (2009). Probabilistic argumentation. *Journal of Applied Logic, 7*(2), 155–176. http://dx.doi.org/10.1016/j.jal.2007.11.006. Special issue: Combining Probability and Logic

Hao, Z., Liu, B., Wu, J., & Yao, J. (2015). Exploiting ontological reasoning in argumentation based multi-agent collaborative classification. In *Intelligent information and database systems* (pp. 23–33). Springer.

Hao, Z., Yao, L., Liu, B., & Wang, Y. (2014). Arguing Prism: An argumentation based approach for collaborative classification in distributed environments. In *Database and expert systems applications* (pp. 34–41). Springer.

Hasanah, N., Imai, S., & Nobuhara, H. (2010). Application of formal concept analysis for rule mining in artificial neural networks. In *Scis & isis: 2010* (pp. 670–675). Japan Society for Fuzzy Theory and Intelligent Informatics.

Hatzilygeroudis, I., & Prentzas, J. (2015). Symbolic-neural rule based reasoning and explanation. *Expert Systems with Applications, 42*(9), 4595–4609. http://dx.doi.org/10.1016/j.eswa.2015.01.068.

Hsu, C.-W., & Lin, C.-J. (2002). A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE Transactions on, 13*(2), 415–425.

Huang, C., Davis, L., & Townshend, J. (2002). An assessment of support vector machines for land cover classification. *International Journal of remote sensing, 23*(4), 725–749.

Johnson, M. D., Hsieh, W. W., Cannon, A. J., Davidson, A., & Bédard, F. (2016). Crop yield forecasting on the canadian prairies by remotely sensed vegetation indices and machine learning methods. *Agricultural and Forest Meteorology, 218-219*, 74–84. http://dx.doi.org/10.1016/j.agrformet.2015.11.003.

Johnston, B., & Governatori, G. (2003). Induction of defeasible logic theories in the legal domain. In *Proceedings of the 9th international conference on artificial intelligence and law* (pp. 204–213). ACM.

Kavzoglu, T. (2009). Increasing the accuracy of neural network classification using refined training data. *Environmental Modelling & Software, 24*(7), 850–858.

Kavzoglu, T., & Mather, P. (2003). The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing, 24*(23), 4907–4938.

Kulluk, S., Özbakır, L., & Baykasoğlu, A. (2013). Fuzzy DIFACONN-miner: A novel approach for fuzzy rule extraction from neural networks. *Expert Systems with Applications, 40*(3), 938–946.

Kuncheva, L. I. (2004). *Combining pattern classifiers: Methods and algorithms*. John Wiley & Sons.

Lauser, B., Johannsen, G., Caracciolo, C., van Hage, W. R., Keizer, J., & Mayr, P. (2008). Comparing human and automatic thesaurus mapping approaches in the agricultural domain. *Universitätsverlag Göttingen*, 43.

Lease, M. (2011). On quality control and machine learning in crowdsourcing.. *Human Computation, 11*, 11.

LeCun, Y. A., Bottou, L., Orr, G. B., & Müller, K.-R. (2012). Efficient backprop. In *Neural networks: Tricks of the trade* (pp. 9–48). Springer.

Léger, B., & Naud, O. (2009). Experimenting statecharts for multiple experts knowledge elicitation in agriculture. *Expert Systems with Applications, 36*(8), 11296–11303.

Li, D., Kang, L., Cheng, X., Li, D., Ji, L., Wang, K., et al. (2013). An ontology-based knowledge representation and implement method for crop cultivation standard. *Mathematical and Computer Modelling, 58*(3-4), 466–473. http://dx.doi.org/10.1016/j.mcm.2011.11.004. Computer and Computing Technologies in Agriculture 2011 and Computer and Computing Technologies in Agriculture 2012

Li, M., & Chung, S.-O. (2015). Special issue on precision agriculture. *Computers and Electronics in Agriculture, 112*. 1– http://dx.doi.org/10.1016/j.compag.2015.03.014.

Liaghat, S., Balasundram, S. K., et al. (2010). A review: The role of remote sensing in precision agriculture. *American Journal of Agricultural and Biological Sciences, 5*(1), 50–55.

Luaces, O., Rodrigues, L. H. A., Meira, C. A. A., & Bahamonde, A. (2011). Using nondeterministic learners to alert on coffee rust disease. *Expert Systems with Applications, 38*(11), 14276–14283.

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*(3), 247–259.

Muhlbaier, M., Topalis, A., & Polikar, R. (2005). Ensemble confidence estimates posterior probability. In *Multiple classifier systems* (pp. 326–335). Springer.

NASS, U. (2010). Field crops: Usual planting and harvesting dates. *USDA National Agricultural Statistics Service, Agriculural Handbook*, (628).

Noh, H., Zhang, Q., Shin, B., Han, S., & Feng, L. (2006). A neural network model of maize crop nitrogen stress assessment for a multi-spectral imaging sensor. *Biosystems Engineering, 94*(4), 477–485.

Núñez, H., Angulo, C., & Català, A. (2002). Rule extraction from support vector machines.. In *Esann* (pp. 107–112).

Ontañón, S., Dellunde, P., Godo, L., & Plaza, E. (2012). A defeasible reasoning model of inductive concept learning from examples and communication. *Artificial intelligence, 193*, 129–148.

Ontanón, S., & Plaza, E. (2011). Empirical argumentation: Integrating induction and argumentation in mas. In *Argumentation in multi-agent systems* (pp. 49–67). Springer.

Özbakir, L., Baykasoğlu, A., Kulluk, S., & Yapıcı, H. (2009). Taco-miner: an ant colony based algorithm for rule extraction from trained neural networks. *Expert Systems with Applications, 36*(10), 12295–12305.

Pal, M., & Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment, 86*(4), 554–565.

Pérez-Ortiz, M., Peña, J. M., Gutiérrez, P. A., Torres-Sánchez, J., Hervás-Martínez, C., & López-Granados, F. (2016). Selecting patterns and features for between-and within-crop-row weed mapping using uav-imagery. *Expert Systems with Applications, 47*, 85–94.

Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine, 6*(3), 21–45. doi:10.1109/MCAS.2006.1688199.

Pollock, J. L. (1995). *Cognitive Carpentry: A blueprint for how to build a person*. Bradford/MIT Press.

Quinlan, J. R. (1987). Simplifying decision trees. *International Journal of Man-Machine Studies, 27*(3), 221–234.

Rahwan, I., & Simari, G. R. (2009). *Argumentation in artificial intelligence*. Springer.

Raudys, Š., & Roli, F. (2003). The behavior knowledge space fusion method: Analysis of generalization error and strategies for performance improvement. In *Multiple classifier systems* (pp. 55–64). Springer.

Setiono, R., & Liu, H. (1997). Neurolinear: From neural networks to oblique decision rules. *Neurocomputing, 17*(1), 1–24.

Shoham, Y. (2015). Why knowledge representation matters. *Communications of the ACM, 59*(1), 47–49.

Simari, G., & Loui, R. (1992). A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence, 53*, 125–157.

Slota, M., Leite, J., & Swift, T. (2015). On updates of hybrid knowledge bases composed of ontologies and rules. *Artificial Intelligence, 229*, 33–104. http://dx.doi.org/10.1016/j.artint.2015.07.008.

Spilke, J., Piepho, H., & Hu, X. (2005). Analysis of unbalanced data by mixed linear models using the mixed procedure of the sas system. *Journal of Agronomy and Crop Science, 191*(1), 47–54.

Thimm, M. (2014). Tweety: A comprehensive collection of java libraries for logical aspects of artificial intelligence and knowledge representation. *KR 14*, 528–537.

United States Department of Agriculture (2012). *Census of agriculture, New Madrid County, Missouri*.

Wang, Y., Wang, Y., Wang, J., Yuan, Y., & Zhang, Z. (2015). An ontology-based approach to integration of hilly citrus production knowledge. *Computers and Electronics in Agriculture, 113*, 24–43. http://dx.doi.org/10.1016/j.compag.2015.01.009.

Wardeh, M., Coenen, F., & Bench-Capon, T. (2012a). Multi-agent based classification using argumentation from experience. *Autonomous Agents and Multi-Agent Systems, 25*(3), 447–474.

Wardeh, M., Coenen, F., & Capon, T. B. (2012b). PISA: A framework for multiagent classification using argumentation. *Data & Knowledge Engineering, 75*, 34–57.

Wei, Y., Wang, R., Hu, Y., & Xue, W. (2012). From web resources to agricultural ontology: A method for semi-automatic construction. *Journal of Integrative Agriculture, 11*(5), 775–783.

Wozniak, M., Graña, M., & Corchado, E. (2014). A survey of multiple classifier systems as hybrid systems. *Information Fusion, 16*, 3–17. http://dx.doi.org/10.1016/j.inffus.2013.04.006. Special Issue on Information Fusion in Hybrid Intelligent Fusion Systems

Xu, J., Yao, L., & Li, L. (2015). Argumentation based joint learning: A novel ensemble learning approach. *PloS one, 10*(5), e0127281.

Yang, X., & Lo, C. (2002). Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area. *International Journal of Remote Sensing, 23*(9), 1775–1798.

Yao, J. T. (2005). Knowledge extracted from trained neural networks: What's next? In *Defense and security* (pp. 151–157). International Society for Optics and Photonics.

Yao, L., Xu, J., Li, J., & Qi, X. (2012). Evaluating the valuable rules from different experience using multiparty argument games. In *Proceedings of the the 2012 ieee/wic/acm international joint conferences on web intelligence and intelligent agent technology-volume 02* (pp. 258–265). IEEE Computer Society.