



Article

A Machine Learning Strategy Based on Kittler's Taxonomy to Detect Anomalies and Recognize Contexts Applied to Monitor Water Bodies in Environments

Maurício Araújo Dias ^{1,*}, Giovanna Carreira Marinho ¹, Rogério Galante Negri ², Wallace Casaca ³,
Ignácio Bravo Muñoz ⁴ and Danilo Medeiros Eler ¹

¹ Department of Mathematics and Computer Science, Faculty of Sciences and Technology, Campus Presidente Prudente, São Paulo State University (UNESP), Sao Paulo 19060-900, Brazil; g.marinho@unesp.br (G.C.M.); danilo.eler@unesp.br (D.M.E.)

² Department of Environmental Engineering, Sciences and Technology Institute, Campus São José dos Campos, São Paulo State University (UNESP), Sao Paulo 12247-004, Brazil; rogerio.negri@unesp.br

³ Department of Energy Engineering, Campus Rosana, São Paulo State University (UNESP), Sao Paulo 19274-000, Brazil; wallace.casaca@unesp.br

⁴ Polytechnic School, University of Alcalá (UAH), 28805 Alcalá de Henares, Spain; ignacio.bravo@uah.es

* Correspondence: ma.dias@unesp.br



Citation: Dias, M.A.; Marinho, G.C.; Negri, R.G.; Casaca, W.; Muñoz, I.B.; Eler, D.M. A Machine Learning Strategy Based on Kittler's Taxonomy to Detect Anomalies and Recognize Contexts Applied to Monitor Water Bodies in Environments. *Remote Sens.* **2022**, *14*, 2222. <https://doi.org/10.3390/rs14092222>

Academic Editors: Juan Ignacio Arribas and Claudio Picciarelli

Received: 25 February 2022

Accepted: 3 May 2022

Published: 6 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Environmental monitoring, such as analyses of water bodies to detect anomalies, is recognized worldwide as a task necessary to reduce the impacts arising from pollution. However, the large number of data available to be analyzed in different contexts, such as in an image time series acquired by satellites, still pose challenges for the detection of anomalies, even when using computers. This study describes a machine learning strategy based on Kittler's taxonomy to detect anomalies related to water pollution in an image time series. We propose this strategy to monitor environments, detecting unexpected conditions that may occur (i.e., detecting outliers), and identifying those outliers in accordance with Kittler's taxonomy (i.e., detecting anomalies). According to our strategy, contextual and non-contextual image classifications were semi-automatically compared to find any divergence that indicates the presence of one type of anomaly defined by the taxonomy. In our strategy, models built to classify a single image were used to classify an image time series due to domain adaptation. The results 99.07%, 99.99%, 99.07%, and 99.53% were achieved by our strategy, respectively, for accuracy, precision, recall, and F-measure. These results suggest that our strategy allows computers to recognize contexts and enhances their capabilities to solve contextualized problems. Therefore, our strategy can be used to guide computational systems to make different decisions to solve a problem in response to each context. The proposed strategy is relevant for improving machine learning, as its use allows computers to have a more organized learning process. Our strategy is presented with respect to its applicability to help monitor environmental disasters. A minor limitation was found in the results caused by the use of domain adaptation. This type of limitation is fairly common when using domain adaptation, and therefore has no significance. Even so, future work should investigate other techniques for transfer learning.

Keywords: remote sensing; Kittler's taxonomy; anomaly detection; machine learning; time series; pattern recognition

1. Introduction

One of the greatest challenges to resolve problems related to environmental monitoring is the large quantity of heterogeneous data in images from remote sensing that need to be analyzed in different contexts by humans or machines. The analysis of river water quality conditions in images from remote sensing is an example of such challenges [1–6]. Regarding contexts, an example is the spatial context, in which the analysis is performed

by considering where abnormalities in the river water quality conditions are present in a single image. Another example is the temporal context, in which the analysis is performed by considering when abnormalities in river water quality conditions occurred by using a set of images. Both contexts are important for analyzing river water quality conditions. However, the conditions of river water over time can only be determined if this problem is analyzed in the context of a time series [7–9], i.e., a sequence of data (e.g., satellite images) acquired during a period. Therefore, a more accurately recognized context leads to an improved response to solve the problem. Moreover, context recognition is a relevant issue and applicable to a variety of areas, not only environmental monitoring.

Humans can accurately recognize the context in which a problem is inserted because they have developed skills to learn how to recognize different contexts [10]. Consequently, humans can offer the best solution to a problem found in images. A researcher with experience working with problems in different contexts can propose, for example, the use of spectral analysis [11] (spectral index [12] or slope ratio [13]) to detect river water pollution in a single image from remote sensing. Unfortunately, humans have difficulties analyzing large quantities of heterogeneous data in images, contrary to computers. For this reason, the focus of this research is on machines rather than humans. However, computers generally have difficulties recognizing different contexts because current machine learning approaches [11–26] are mainly based on trial and error [25] or probabilistic models. In other words, computers generally learn by making mistakes or testing each of all available probable solutions in order to recognize a context or choose an adequate solution for solving a problem.

Alternatively, machine learning can be improved when a learning approach is based on a taxonomy that provides support for context recognition, as it enhances the organization of learning processes. This improvement by context recognition is important because it can allow computers to make better choices when solving problems. A taxonomy is a framework that machines can use when they are making decisions. In such cases, machines act as a decision-making system (DMS) installed in a computer [27–30]. A DMS is software responsible for choosing a specific solution to solve each type of problem automatically or semi-automatically. The expectation is that a DMS can recognize the context in which each problem is inserted. Therefore, a framework, such as Kittler’s taxonomy of anomalies [27], is important because it can be used as a “scaffold” [10] by the DMS. In other words, the taxonomy can provide a supporting structure for DMSs to enhance machine learning and bring more “intelligence” to computers.

For example, it is now possible for DMSs to base their learning on Kittler’s taxonomy [27] by using a new anomaly detection strategy [31]. Such a strategy would allow the DMS to find appropriate solutions for analyzing river pollution in images from remote sensing when a problem is inserted in the spatial context [31]. However, this computational capability remains a challenge when a problem is inserted in the temporal context.

1.1. Kittler’s Taxonomy and Anomaly Detection

A taxonomy of anomalies was proposed by Kittler et al. in [27]. Kittler’s taxonomy is well-established among researchers from many different research areas, and its application on synthetic [28–30] and non-synthetic [31] data has been studied to find solutions for problems. Kittler et al. established a set of principles and rules to identify outliers, while considering that outliers can be distinguished according to categories of anomalies. Outliers [32–37] are observations which appear isolated from other observations because they are inconsistent with the rest of the data series, according to statistical analysis [32]. Once an outlier is categorized in accordance with a taxonomy, it is considered an anomaly of a specific type. The concept of an anomaly was expanded by Kittler et al. in [27] to a meaning more complex than the concept of outlier. For example, Kittler’s taxonomy defines some types of anomalies such as *unexpected structure and structural components*, *unknown structure*, *unknown object*, *unexpected structural component*, *measurement model drift*, and *component model*

drift. The greater the number of different types of anomalies that the DMS can deal with, the more qualified the DMS is to perform the detection of anomalies for solving problems.

The detection of anomalies [27,38] and incongruences [27–30,39] are activities applied to identify and categorize patterns that present some unexpected behavior, such as the outliers [27,38]. These two activities use powerful tools from pattern recognition (PR) [25, 40] and computer vision (CV) [40,41], such as classifiers [28–30,39,42]. PR and CV are important fields of computer science and heavily involve investigating regularities and patterns in data [25,27–30,38–40,43–45] and obtaining information from images [40,41], respectively. A classifier is software that uses statistical models to categorize things into one of a set of classes. Kittler et al. in [27] proposed the use of contextual [46] and non-contextual [26,47–49] classifiers [28–30,39,42], an incongruence indicator [28–30], and sensory data quality assessment [50] to detect and categorize each type of anomaly.

A relevant example of anomaly detection based on Kittler’s taxonomy that was used for recognizing real-world problems in the spatial context was presented by Dias et al. in [31]. In that study, they proposed an incongruence-based anomaly detection strategy for analyzing water pollution in images from remote sensing. The strategy can provide opportunities for DMSs to perform semi-automatic detection of anomalies such as *unexpected structure and structural components*. Incongruence (disagreement) between contextual and non-contextual classifiers is used as part of the strategy. The incongruence was used to recognize the presence of brown mud in river water. The strategy was designed to analyze river water quality conditions using a set of models to classify a single satellite image rather than a time series. Therefore, a DMS can use anomaly detection based on Kittler’s taxonomy for learning river pollution recognition in the spatial context.

River water quality conditions are typically analyzed using the same set of models to classify each image of a satellite image time series. In such cases, if classifiers agree while classifying one image but disagree on a later image, a DMS could use anomaly detection based on Kittler’s taxonomy to recognize river pollution in the temporal context. However, the design of an anomaly detection strategy based on Kittler’s taxonomy for recognizing real-world problems in a satellite image time series provides a challenge.

1.2. Contextualizing the Problem

On one hand, from the viewpoint on remote sensing of many researchers, machine learning is generally seen as a tool or set of tools that can be applied to achieve some objective in their scientific area; for example, for detecting river water pollution. Therefore, we suppose that researchers would expect to find in this paper only the description of a novel tool or its innovative use in remote sensing. With this in consideration, the same researchers would also likely expect to find comparisons between classifiers, e.g., those which are based on CNN, as well as the use of new datasets or different domain adaptation strategies. Although this paper deals with the detection of river water pollution, the main objective of this work is not to merely describe a tool or its innovative use.

On the other hand, from the viewpoint on computer science of researchers, machine learning is a scientific area of study that deals with the challenge of making machines learn things. In such context, it is possible to compare some computer science researchers and computers to teachers and their students in a classroom. Teachers use educational taxonomies (such as Bloom’s taxonomy, Ausubel’s taxonomy, etc. [10]) in their classes to improve students’ learning processes; similarly, computers learn more when learning is based on a taxonomy such as Kittler’s. Programming computers to apply learning based on Kittler’s taxonomy on problems related to Earth observation has been shown to be an efficient new strategy, as described by the study in [31]. To extend that same strategy for use with time series is challenging and innovative for both researchers and computers. Different from humans, computers have not naturally evolved for thinking. Therefore, one of the greatest challenges in machine learning is making machines learn how to distinguish contexts.

For machine learning in computer science, high quantities of similarities shared by different contexts in two studies result in increased challenges for computers to learn how to distinguish those contexts, as computers can interpret that there is a single context involved in both studies. Consequently, this leads to more difficulties for researchers to achieve success. Therefore, the more similarities between two studies in which a computer has learned to recognize two different contexts, the more evident is the success achieved by the researchers involved in those studies. Otherwise, doubts can arise regarding whether the success of the studies were achieved either because their strategies are efficient or because the differences between them are so evident that they make the task of distinguishing contexts easy for computers; this is the reason why it is important to maintain this study and the one described in [31] as similar as possible.

However, many differences exist between this study and the one described in [31]. These differences are inevitable when working with the anomaly such as *component model drift* [27], as it is only possible to detect this type of anomaly by analyzing time series. Table 1 shows examples of some important differences between our study and the one described in [31], taking into consideration only the area commonly studied by both studies. Based on Table 1, it is possible to infer that some features, such as those related to land cover, atmospheric conditions, shadows, etc., vary from one image to others because the scene in the images are influenced by the change in seasons, illumination from sunlight and changes caused by human activities over time, etc. Therefore, this study makes a major contribution to research on Earth observation.

As shown in Table 1, although the images in this study refer to the same geographical coordinates, the differences they register over the time series are subtle to the human eye, but strong enough for a single classifier to require a different model per image. In machine learning, different models that are useful for a single classifier represent different domains. Since the study on *component model drift* requires the use of only one main model for all images in the time series, domain adaptation is needed to provide transfer learning.

In this subsection, we have explained that a study describing a machine learning strategy based on Kittler's taxonomy to detect anomalies for analyzing an image time series in comparison to the study described in [31] is important to show that computers can learn how to distinguish two different contexts related to Earth observation. Moreover, a study describing how computers are able to distinguish between two different types of anomalies, such as *unexpected structure and structural components* and *component model drift*, without erroneously interchanging them, even when the same problem (i.e., river water pollution caused by an environmental disaster) is analyzed under very similar conditions, is equally important.

1.3. The Proposal

This paper describes the design and implementation of a machine learning strategy based on Kittler's taxonomy to detect anomalies for analyzing water pollution in an image time series from remote sensing. Data for this research were acquired by the Landsat 8 satellite and obtained from the United States Geological Survey (USGS) repository [50]. The methodological approach taken in this research is a mixed methodology based on PR, CV, and Kittler's taxonomy. Our strategy uses incongruent results from contextual and non-contextual classification to analyze high-quality images. We propose this strategy for monitoring the environment, detecting unexpected conditions that may occur (i.e., detecting outliers), and identifying those outliers in accordance with Kittler's taxonomy (i.e., detecting anomalies). This research is unique because the proposed strategy allows machines to base their learning process on Kittler's taxonomy to detect anomalies in the temporal context. Our study focuses only on detecting anomalies such as *component model drift* [27] when this anomaly is related to river water pollution. However, the proposed strategy is not specific to detecting anomalies in the water quality of rivers. Our strategy can also be applied to other water bodies if necessary.

Table 1. Examples of some important differences between our study and the one presented by Dias et al. In [31], taking into consideration only the area commonly studied by both studies.

Issue	Study	
	[31]	Ours
Type of anomaly investigated	Unexpected structure and structural components	Component model drift
Main context analyzed	Spatial	Temporal
Applicable to time series	No	Yes
Number of Landsat 8 images studied with resolution of $15,705 \times 15,440$ (height and width in pixels)	One	Eight
Analyze drift	No	Yes
Number of images analyzed with resolution of 151×193 (height and width in pixels)	8400	67,200
Seasons of the year studied (regarding the Southern Hemisphere)	Spring	Spring, summer, fall, and winter
Presence in the scene of effects caused by an environmental disaster registered in the image used to create the main models for classifications	Yes	No
Years studied	2015	2013, 2014, 2015, 2016, 2017, 2018, 2019, and 2020
Atmospheric conditions	Stable	Variable
Year of the reference image	2015	2013
Perform domain adaptation	No	Yes
Condition of the land cover analyzed	Stable	Variable
Number of samplings related to the domains studied	One per domain	One for all domains
Year of the image in which the sampling was performed	2015	2013
Perform transfer learning	No	Yes
Presence of changes in the scene caused by human activities	Unnoticeable	Multiple
Modeling	Two particular models per domain	Two main models for all domains (they are adapted to each domain)
Angle of incidence of the sunlight	Single	Multiple
Number of years needed to finish the study with success	Two	Three

In other words, the main objective of this study is to describe the design and implementation of a strategy that helps computers learn to distinguish two different contexts related to Earth observation by recognizing them. This learning is important in order to support computers when they have a large number of data to be analyzed in different contexts. In this study, the aforementioned learning objective is achieved by computers when they are able to deal either with the anomaly *unexpected structure and structural components* or the anomaly *component model drift*, without confusing them, even when the analyzes occur under very similar conditions. This is the first study to compare the experiences of applying Kittler’s taxonomy to the spatial and temporal contexts in remote sensing, to the knowledge of the authors.

Another objective has been to improve the monitoring of environmental disasters. Consequently, this research provides an important description of how the proposed strategy can contribute to the monitoring of river water quality conditions over time. The innovation of this research, to the knowledge of the authors, is the introduction of the first use of a strategy based on Kittler’s taxonomy for detecting anomalies to analyze river water pollution by a satellite image time series from remote sensing. Remote sensing was chosen for this research because of its capabilities for real application of Kittler’s taxonomy to help solve real-world problems [31].

This paper has been divided into six sections. The second section presents a background with two main subsections: theoretical foundation and related work. The third section describes the materials and methods used for performing the experiments. The fourth section reviews the achieved results. The fifth section discusses the results. The sixth section presents the conclusions.

2. Background

2.1. Theoretical Foundation

2.1.1. Domain and Image Time Series

The term domain will be used in this research to refer to each image in an image time series. In image analysis, a domain is determined by the set of characteristics associated with a certain image, such as viewing angle, overlay contamination by cloud cover, illumination conditions, and the conditions of the environment. Consequently, multiple images are multiple different domains, no matter whether the images were acquired from the same environment or not. Therefore, in this research, an image time series is a set composed of multiple domains acquired from the same environment at different dates.

2.1.2. Classifier

The term “classifier” is used in this paper to refer to a computational tool, such as an expert or a learner, that is responsible for performing classification. In other words, a classifier categorizes things into one class among predefined others according to their similarities [51]. In this sense, a classifier can be used, for example, to create models that represent environments [28]. Such models are generally created based on samples and prior knowledge about the problem. Samples are aggregations of pixels selected from the application domain [28,39] to represent features of interest, such as water, forest, plantation, etc. If samples for classifiers are selected by users, then the classifications are referred to as “supervised” or “semi-automatic.” In such cases, the user selects samples from an image and classifies the image based on the use of a classifier trained with these samples [52].

Two types of classifiers are well-established: non-contextual and contextual [27]. A non-contextual classifier is a weak learner that performs general-level tasks. A contextual classifier is a strong learner that performs specific-level tasks. Non-contextual classifiers are stronger than random classifiers, but they are not as constrained and precise as the contextual ones [51]. Contextual classifiers are more dependent on training data or specific prior knowledge [46]. Moreover, the contextual classifier uses multiple weaker classifiers working collaboratively to perform more successful classifications. For both types of classifiers, the input data induces the production of class-posterior probabilities as results [39]. Additionally, it is expected that both types of classifiers exhibit similar probability estimates when classifying a common input, such as a scene (image), despite the environment [28]. Otherwise, incongruence occurs.

2.1.3. Incongruence

Throughout this paper, the term “incongruence” refers to the disagreement between contextual and non-contextual classifiers caused by contradictory evidence that results from the time series analysis. In other words, incongruence is a conflicting classification caused by a significant divergence between posterior probabilities [39]. It is important to emphasize that the disagreements between either two non-contextual [26,47–49] or two contextual [46] classifiers have no relation to incongruence. In such cases, the disagreement reveals an incorrect classification performed by one of the classifiers [27–30,39].

Incongruence can be understood, for example, in a context in which river water quality conditions are analyzed by comparing images of a river acquired in different periods. Incongruence occurs when the classifiers agree about the presence of water when classifying one image, but one of the classifiers does not recognize the presence of water when classifying a later image.

2.1.4. Outlier and Anomaly

Historically, the terms “outlier” and “anomaly” have been used interchangeably to mean non-conforming patterns [32–37]. However, according to Kittler et al. in [27], non-conforming patterns can be organized into categories of anomalies. In this sense, each type of anomaly is determined by a combination of conditions, such as the quality of the sensory data, the use of a time series, the contextual [46] and non-contextual [26,47–49] classifications [28–30,39,42], and the occurrence of incongruence [28–30]. Therefore, according to Dias et al. in [31], an anomaly can be defined as follows: a non-conforming pattern categorized in accordance with a taxonomy.

An anomaly can be, for example, a component model drift. This type of anomaly can be detected when an analysis focuses on features that allow classifiers to identify components of samples in a time series. The detection of this type of anomaly also depends on the combination of the occurrence of incongruence between the results achieved by contextual and non-contextual classifications of high-quality data. Other combinations of these conditions are needed to detect other types of anomalies, such as the one described by Dias et al. in [31]. An example of the component model drift anomaly found in a satellite image time series is related to monitoring high levels of turbidity in river water over time, caused by the presence of brown mud.

2.1.5. DMS

In this article, the abbreviation DMS is used to refer to a decision making system. A DMS is a computational tool that identifies problems in order to automatically choose the most appropriate solutions for resolution [27–30]. Generally, a DMS uses multiple classifiers to support its decisions. When a computational tool makes decisions based on the analyses of classifiers, it is expected that the classifiers will agree regarding hypotheses [29]. However, a DMS also must recognize conflicting classifications when it uses non-contextual and contextual classifiers in parallel [39].

In other words, a DMS must be able to detect incongruences [28–30,39]. This is important, as the way a DMS will act can be conditioned by the monitoring and detection of incongruence [28–30,39]. Some decisions depend on the types of anomalies the DMS can identify. For example, when a DMS detects an incongruence while classifying a high-quality Landsat 8 scene time series, the set of conditions allows the DMS to identify an anomaly such as *component model drift*. The identification of this type of anomaly can allow the DMS to learn how to recognize, for example, river water pollution caused by brown mud in the temporal context. In this case, the DMS could guide a computer to choose, for example, spectral analysis [11] to perform a complementary analysis of the water pollution.

2.1.6. Transfer Learning and Domain Adaptation

It is necessary here to clarify the definition of transfer learning [53]. Selecting samples (i.e., sampling) from a domain for training classifiers is a difficult and time-consuming task. This fact makes sampling each domain unfeasible. This drawback can be overcome by transfer learning, i.e., models built from training classifiers for a domain and used by classifiers to analyze other similar domains. However, transfer learning usually fails when it is performed without any adaptation. Therefore, some domain adaptation is required to make transfer learning feasible and useful.

According to Tuia et al. in [53], Domain Adaptation (DA) is defined as an approach for performing transfer learning. In this research, we used a semi-supervised DA approach, i.e., the approach considers that the sampling is performed only for a single domain in the time series (i.e., a reference image). This approach is based on statistics and machine learning.

Regarding statistics, our approach was inspired by the Maximum Mean Discrepancy (MMD) method presented by Gretton et al. in [54]. Our DA approach is a set of calculations based on mean values obtained between a reference image and each time series image. More detailed descriptions of our DA approach can be found in the Materials and Method section.

Regarding machine learning, classifiers can achieve the highest levels of accuracy only if they analyze the same domain for which they were trained, and their models were built. If transfer learning is required, DA approaches take into consideration that although domains are different, they need to be similar enough to allow the models built for one domain to be useful for solving the learning problem of another domain. However, DA can produce poor results, even for similar domains, because the spectra found in domains can differ from one another. As Tuia et al. in [53] points out: “DA aims to adapt models trained to solve a specific task to a new, yet related, task, for which the knowledge of the initial model is sufficient, although not perfect”.

2.2. Related Work

2.2.1. Research on Outlier Detection

A considerable amount of literature has been published on detections in time series (e.g., papers such as [7–9,16,55–60]). These studies are mostly related to outlier detection. A recent study published by Che et al. in [58] described the temporal extension of a machine-learning algorithm. The algorithm was applied on a Landsat image time series to detect patterns of surface water cover and change. In their study, an accurate long-term water-body dataset was produced to study seasonal changes. The study improves the understanding of the surface water dynamics at a regional scale. Their method, based on a decision tree model built using the C5 algorithm, achieved a very high level of accuracy.

In an investigation into outlier detection in a time series, Bormann et al. in [57] developed a new daily snow cover dataset that provides a satellite-based observational record. These snow cover observations allow researchers to detect outliers in the temporal context, e.g., short season duration, declines in snow cover extent, and anticipated melting of snow. They used a custom Melt Area Detection Index (MADI) algorithm adapted for the snow conditions of Australia. Bormann et al. provided a study focused on a Landsat data time series to help assess snow monitoring, which contributes to snow hydrology and water resource management. In that study, the method based on detection index achieved very high levels of accuracy and precision, but a low level of recall.

In a study conducted by Shoujing et al. in [7], a robust change detection method to detect outliers in time-series images from remote sensing was proposed. Their method based on Median Absolute Deviation (MAD) allowed them to achieve a very high level of accuracy and high levels of precision and recall. The method can detect temporal and spatial changes. Shoujing et al. concluded that changes caused by human activities, weather condition variation, vegetation phenology, sensor aging, emergencies (e.g., fire, insect, drought, pest, etc.), and climatic changes found in the images were outliers.

In an analysis of seasonal autocorrelation published in [8], Zhou et al. identified spatial-temporal dynamic processes of unexpected changes as outliers. Their method analyzes each image of a satellite time-series for identifying changes related to flooding. In that study, a method based on a seasonal autoregressive integrated moving average (SARIMA) model for autocorrelation analysis allowed them to achieve high levels of accuracy, precision, and recall.

One study by Chandola and Vatsavai proposed a change-detection algorithm for monitoring a periodic time series in [9]. Their algorithm finds differences between predictions and previous observations within a statistical control framework to identify changes. In their study, Chandola and Vatsavai used a Bayesian nonparametric predictive model based on a Gaussian Process for time series in an online mode. This method achieved satisfactory levels of accuracy, precision, and recall.

In a study investigating distributed outlier detection presented in [56], Bhaduri et al. examined whether it is possible to identify outliers in distributed earth science databases without moving all the data to a single location. The problem is that some outliers are missed when data are available only in a single database. In their study, Bhaduri et al. developed an algorithm to detect such outliers. The algorithm was applied on a time series

composed of satellite images acquired over a period of eight years. In [56], a method based on Support Vector Machine (SVM) provided a high level of accuracy.

The study of automatic post-disaster damage mapping for change detection was carried out by Sublime and Kalinicheva in [16]. In their case study of the Tohoku Tsunami in [16], Sublime and Kalinicheva developed an unsupervised deep-learning method. Their method detects the difference between trivial (e.g., changes in luminosity, vegetation or crops due to seasonal patterns) and non-trivial changes (e.g., damaged roads or buildings, and flooded areas). Sublime and Kalinicheva applied the method on a satellite image time series to detect the outliers. In that study, a method based on a joint fully convolutional auto-encoders (FC AE) model allowed them to achieve high levels of accuracy and recall, but a low level of precision.

In a study that set out to define four different categories of outliers or anomalous events, Qi Liu et al. presented an unsupervised outlier detection framework in [55]. The framework detects outliers and anomalous events even without prior knowledge. In their study, Qi Liu et al. wrote: “outliers are objects which have either low spatial or temporal coherence with their neighbors”. They also wrote: “an anomalous event is a group of outliers that share similar spatial and temporal anomalous behaviors”. In [55], a method based on an extended Expectation-Maximization (EM) algorithm provided high levels of accuracy, precision, and recall.

Although [59,60] achieved high levels of accuracy applying their methods, they did not report results in regard to recall and precision, as well as [58]. In these studies, the levels of accuracy were achieved after applying a method based on: (1) maximum likelihood classification in [60]. (2) a deep-learning method based on convolutional neural network (CNN) in [59].

These studies support the notion that scientific literature has been interested in discussing research on non-conforming patterns. However, such studies remain narrow in focus when dealing with outliers, as outlier detection does not allow systems to categorize different types of problems. This drawback can be overcome by a taxonomy of anomalies. Of the studies reviewed here, Qi Liu et al. [55] has the most focus on categories of outliers or anomalous events. Qi Liu et al. highlighted the need for categorizing detected anomalies into taxonomies.

Regarding advantages and disadvantages, all the aforementioned methods have benefits and drawbacks that can be compared against our strategy. The main advantage is that these methods can be applied to a time series of images, such as those of our strategy. The main disadvantage is that these methods detect only outliers rather than anomalies; our strategy detects both. The only exception is the method proposed by Qi Liu et al., because they define four different categories of outliers or anomalous events. Nevertheless, this number of categories is much lower than the ten categories from Kittler’s taxonomy, which limits the application of the Qi Liu et al. method to machine learning.

2.2.2. Research on Anomaly Detection

Analysis of anomaly detection in machine perception was first carried out by Kittler et al. in [27]. In their study, Kittler et al. presented a taxonomy and a framework. The taxonomy establishes different types of anomalies, such as *unexpected structure and structural components*, *unexpected structural component*, *unknown object*, *unknown structure*, *measurement model drift*, and *component model drift*. Kittler et al. argued that novelty detection of an object or an object primitive, noise, rare events, unexpected events, and distribution drift are contributing factors to the occurrence of different types of anomalies. The framework provides mechanisms to detect and categorize anomalies, such as non-contextual [26,47–49] and contextual [46] classifications [28–30,39,42], sensory data quality assessment [50], and incongruence indication [28–30].

In a recent study [31], Dias et al. reported that it is possible to use an anomaly detection strategy for analyzing water pollution in images from remote sensing. Their strategy semi-automatically recognizes anomalies such as *unexpected structure and structural components* in

single images. Dias et al. introduced the first solution for real-world problems [61–64] that was based on the practical application of Kittler’s taxonomy [27]. Their research provides an opportunity for DMSs to use Kittler’s taxonomy [27] as a reference to categorize real-world problems. Their study can also provide an opportunity for DMSs to recognize different contexts in which real-world problems are inserted, as each different context can be represented by a specific type of anomaly. In this sense, Dias et al. suggested that Kittler’s taxonomy [27] is a “scaffold” [10] that can provide an opportunity for DMSs to improve their learning abilities [25]. Their research also contributes to post-disaster damage mapping.

In both studies reviewed here, outliers are recognized as anomalies because both studies associate problems with at least one among ten different categories of anomalies of Kittler’s taxonomy defined in [27]. Additionally, the same two studies support the hypothesis that a DMS could learn how to recognize anomalies (e.g., those which can be associated with river water pollution) in different contexts by taking into account Kittler’s taxonomy. In their study [31], Dias et al. took the spatial context into consideration to analyze the presence of brown mud in images of rivers in order to monitor environmental problems. Their work was the first of a series of studies focused on understanding how anomaly detection based on Kittler’s taxonomy can be applied in practice to solve real-world problems. Dias et al. argue that the practical application of each type of anomaly has been studied individually, because each type is comprised of different facets that influence the detection and identification processes, e.g., the context in which a real-world problem is inserted. In any case, the potential application of anomaly detection to solve problems is increased by using incongruence.

2.2.3. Research on Incongruence

In 2012, a meaningful analysis and discussion on incongruence was presented by Weinshall et al. in [39]. In their research, Weinshall et al. presented a framework that represents incongruent events. Weinshall et al. claimed that an event is incongruent if divergence occurs between the probabilities associated with different classifications. Therefore, the framework takes conflicting predictions between strong and weak classifications into consideration to identify each type of event. The research published by Weinshall et al. in [39] provides supporting evidence for Kittler’s taxonomy [27].

In 2015, the study by Kittler and Zor published in [30] offered a new way of measuring incongruence based on delta surprise measure. Kittler and Zor argued that different classifiers supporting distinct hypotheses exhibit higher delta surprise values than when supporting the same hypothesis. Classifier confidence has its independence and symmetry satisfied by the measure.

In 2017, in a study of decision cognizant measure, Ponti and colleagues in [28] reported that the contribution of the minority classes to make the true degree of classifier incongruence obscure can be reduced by a novel divergence measure. That measure is a variant of the Kullback–Leibler (KL) divergence named Decision Cognizant Kullback–Leibler divergence (DC-KL). The advantage of DC-KL is its lower sensitivity to noise when compared to the classical KL divergence. Consequently, DC-KL enables pattern recognition systems to be more discriminate between classifier incongruence and congruence.

In 2019, in another study of decision cognizant measure, Kittler and Zor presented a novel delta divergence measure of classifier incongruence in [29]. The measure is based on total variation distance. Kittler and Zor argued that two classes predicted by two classifiers and the possibility that the true class is neither of the two classes are propositions which can be taken into consideration to assess classifier congruence. The measure is decision cognizant, symmetric, bounded, and decision confidence independent.

With regard to advantages and disadvantages of the methods mentioned in this subsection, they all have the advantage of supporting the idea of using incongruence to help detect anomalies, such as with our strategy. The disadvantage is that most of the

methods, in contrast to our strategy, are dedicated to measure divergences rather than offer any practical solution applicable to real-world problems.

Four important facts related to incongruence emerge from the studies mentioned in this subsection: (1) it has been studied for a long time; (2) it is a well-established technique in PR; (3) it is essential for detecting anomalies; (4) it attracts much attention from the scientific research community.

3. Materials and Methods

3.1. Introducing the Machine Learning Strategy based on Kittler's Taxonomy

A number of techniques, such as those mentioned in Section 2.1, have been developed to detect outliers. Despite being successful techniques, they do not provide support for a DMS to improve its learning abilities to recognize different contexts. This problem can be overcome by using a strategy to detect anomalies based on the conceptual framework and taxonomy proposed by Kittler et al. in [27]. The strategy proposed in this article extends the strategy presented by Dias et al. in [31], which is based on Kittler's taxonomy, to the temporal context due to the use of a time series.

Figure 1 presents a flowchart which summarizes the methodology of our strategy. All steps of our strategy were systematically applied on each time series image except for the last step of the data preprocessing and the first step of the learning and classification approaches. In the case of these two steps, sampling followed by training, validation, and test of classifiers were applied on a time series reference image (i.e., the oldest image of the time series). Otherwise, domain adaptation followed by importing models related to the reference image were applied on the other time series images. More detailed descriptions of our strategy can be found in Section 3.4 and Appendix A.

The use of our strategy offers some advantages. It helps to provide an appropriate response to deal with anomalies such as *component model drift*. The strategy can also allow a DMS to identify where and when an anomaly occurred. An example of practical application of our strategy is the identification of where and when high levels of turbidity occur in river water caused by the presence of brown mud as a contaminator. The present study has established a relationship between the presence of brown mud in river water and the occurrence of anomalies such as *component model drift*. The importance of these findings is that our strategy can greatly increase the ability of DMSs to offer solutions to problems related to environmental monitoring.

3.2. Materials

Each task of our strategy presented in Figure 1 was executed methodically using the Qgis 2.18.19 Las Palmas software with the Orfeo 6.4.0 toolbox for this research. Both are freely accessible in [65] for non-commercial purposes. The software and their respective versions were chosen because they are the same used in [31]. In addition, we chose Qgis because it provides powerful toolboxes to analyze images from remote sensing, which has given the software worldwide popularity in this field.

The strategy in Figure 1 was systematically evaluated using data sets obtained from the Earth Explorer Platform, which is provided by the USGS [66]. We chose to work with the USGS' material to perform the experiments in this research because of the following reasons: (1) this material is freely accessible for non-commercial purposes (see [66]); (2) the data are widely used by the scientific research community; (3) the USGS provides high-quality images (scenes) of different environments which show many geographical features in detail.

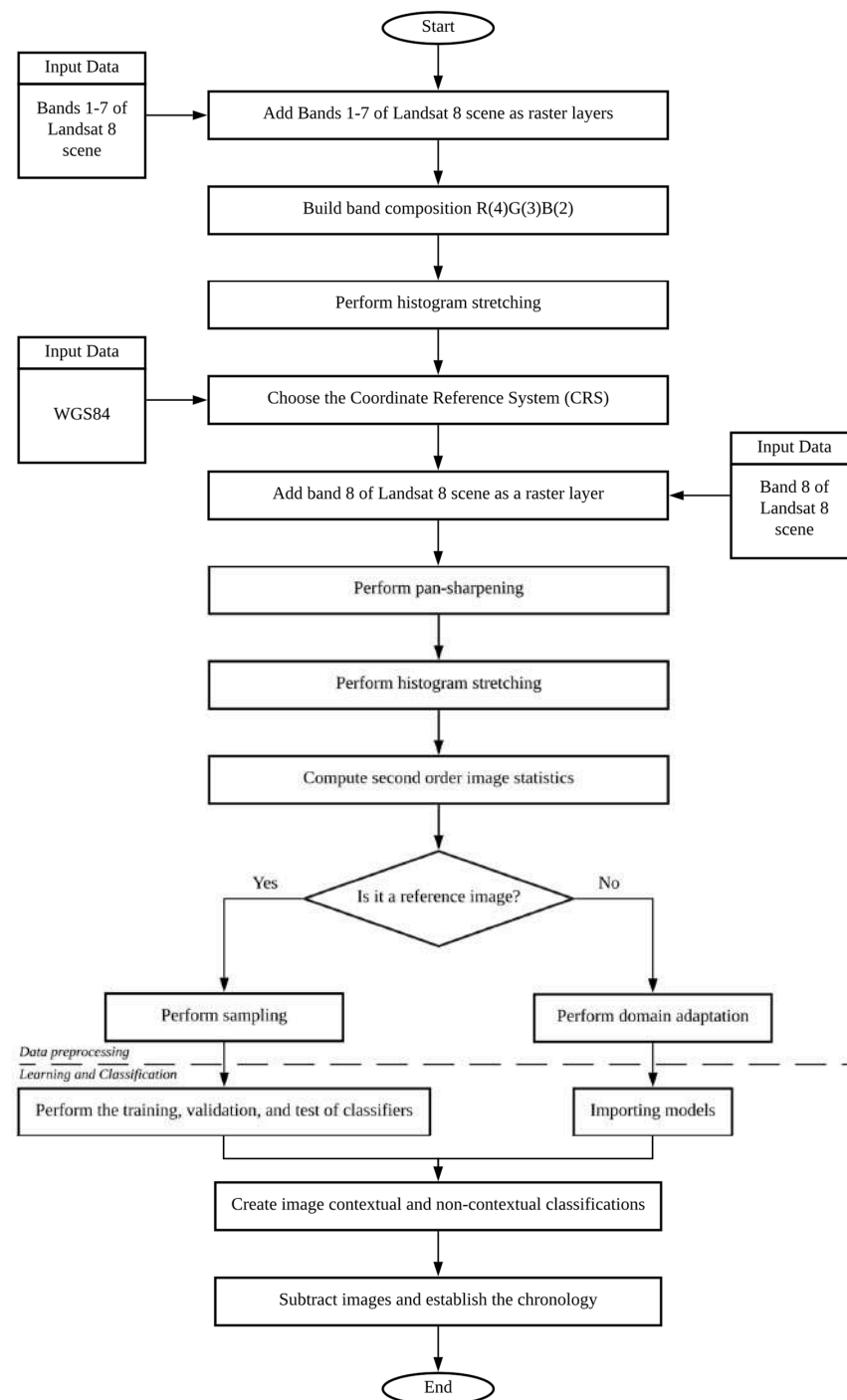


Figure 1. Flowchart of the proposed anomaly detection strategy, which is based on Kittler’s taxonomy. The border between the data preprocessing, learning, and classification approaches are set out in the figure. The adopted coordinate reference system (CRS) uses the WGS84 geodetic reference frame and the universal transverse mercator (UTM) map projection. Geographical features, such as rivers, lakes, waterways, artificial reservoirs, clouds, shadows of clouds, plantations, forests, fields, cities, and highways, are selected by the user (under input data) as samples. Landsat 8 band used, wavelength, and resolution: Band 1—coastal aerosol (0.43–0.45 μm) 30 m; Band 2—blue (0.45–0.51 μm) 30 m; Band 3—green (0.53–0.59 μm) 30 m; Band 4—red (0.64–0.67 μm) 30 m; Band 5—near infrared (NIR) (0.85–0.88 μm) 30 m; Band 6—SWIR 1 (1.57–1.65 μm) 30 m; Band 7—SWIR 2 (2.11–2.29 μm) 30 m, and Band 8—panchromatic (PAN) (0.50–0.68 μm) 15 m.

A Landsat 8 image time series was used to rigorously evaluate our strategy. The time series images were acquired by an instrument onboard the Landsat 8 satellite named Operational Land Imager (OLI). Each image size is almost 170 km north–south by 183 km east–west (106 mi by 114 mi). Landsat 8 images are ready for use in the tagged image file format (TIFF). TIFF images are generally gray scale, RGB, or indexed. The data characteristics of the images are: 15- to 30 m multispectral data, cubic convolution (CC) resampling, north up (MAP) orientation, UTM map projection (polar stereographic for Antarctica), WGS 84 datum, 12 m circular error, 90% confidence global accuracy for OLI, and 16-bit pixel values [50].

Eight Landsat 8 images were selected for this research. This is the number of Landsat 8 images studied with a resolution of $15,705 \times 15,440$ (height and width in pixels). The images were acquired over a period of eight years (2013 to 2020), i.e., since the Landsat 8 satellite started working. The selection criteria for images required that geographical features be as visible as possible. Therefore, all selected images had their conditions assessed by a visual check. Eligible images were identified by the presence of low overlay contamination by cloud cover. Only high-quality images were included in this research, i.e., all selected images are qualified with value “9,” which is the highest score possible in Landsat 8 image collections according to [50]. After being pre-processed in the first phase of our strategy, these images are used as test sets by the classifiers in the second phase of the strategy.

3.3. Study Area

Figure 2 presents a scene and the hydrographical map of part of the Doce River basin in Brazil.

The Doce River basin size is $83,400 \text{ km}^2$ ($32,201 \text{ sq mi}$), and the length of the Doce River is 853 km (530 mi). The predominant features in the environment are mountains and valleys. Mining is the main human activity in that region. Consequently, the potential threats to the environment are ore tailings reservoirs. This basin was chosen for this study because the Doce River is the location of Brazil’s worst environmental disaster, which occurred on 5 November 2015. Many research papers, such as [62], describe studies conducted in that area ($20^\circ 12,023.4'' \text{ S } 43^\circ 28,001.6'' \text{ W}$) to investigate the consequences of the disaster. According to previous studies, the disaster was caused by a dam breach. Studies have also reported that the Doce River received around 55–62 million m^3 of brown mud (iron ore tailings) as a result of the disaster.

The time series composed of eight Landsat 8 images acquired from the Doce River basin was used in the experiments. We evaluated each time series image separately. Table 2 lists information about each image in the time series. The image in the first row is a reference image. This image was chosen as a reference image because it is the first (oldest) image in the time series. The reference image is the one from which models are created to classify all-time series images. Models used to classify the images are created based on a set of samples, as samples are used as training data to model the environment.

Samples in our study are representative with respect to geographical features of the studied environment. The scene of the environment was challenging for classifiers to locate and analyze geographical features, as the time series images used in this present study had varied geographical features. There are parts of the time series images with numerous, small, hard to detect, and geographically spread features that differ from other parts of the time series images with large, easy to detect, and singular geographical features. The diversified geographical features present in the time series images help to assess the accuracy of our strategy while using classifiers to identify incongruences for detection of anomalies.

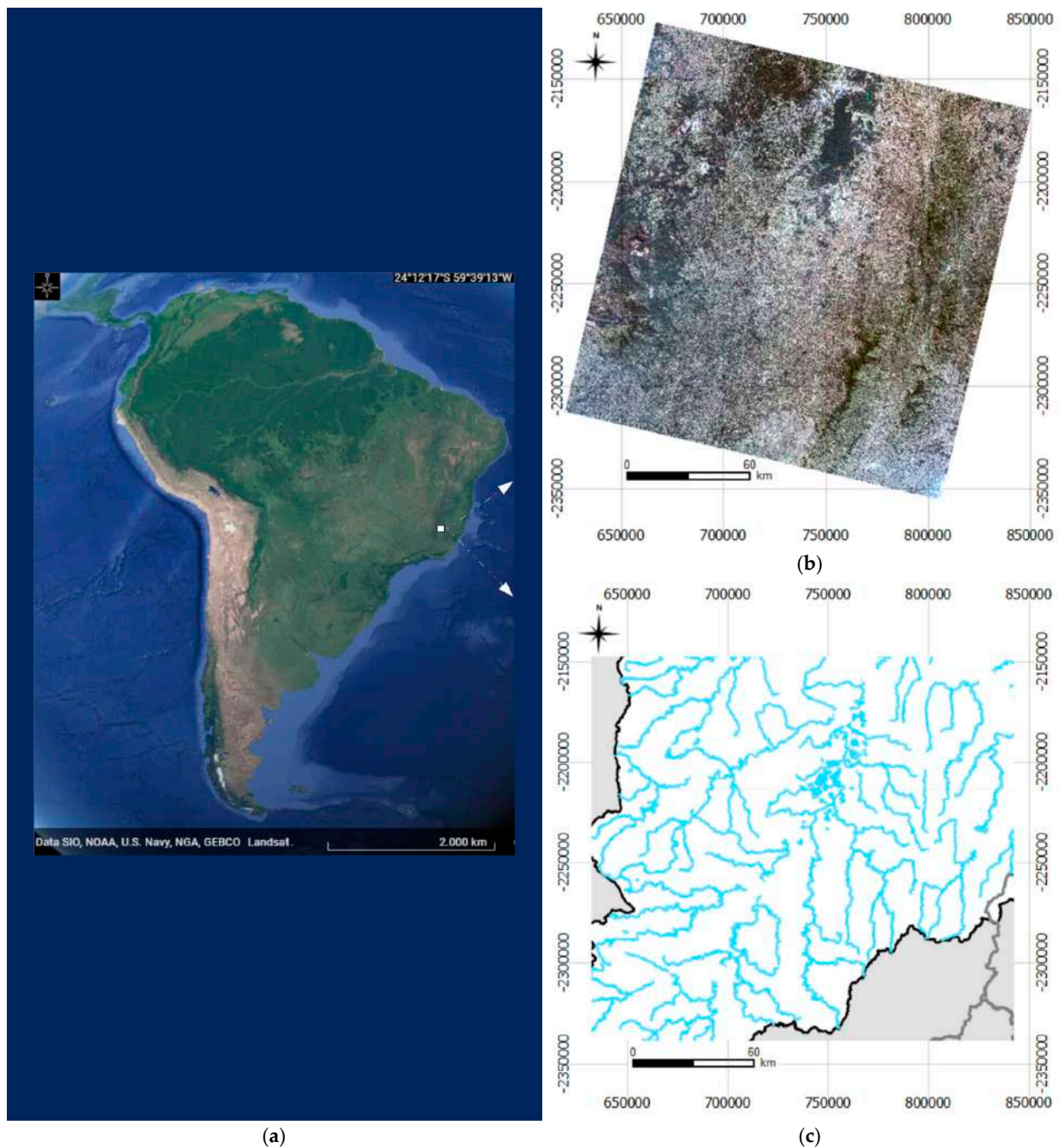


Figure 2. (a) South America, (b) the enhanced band composition (R(4)G(3)B(2)), and (c) the hydrographical map of part of the Doce River basin in Brazil. A band composition is an image composed of different bands of a satellite image.

The samples were selected only from the reference image in this research. A topographic map of the Doce River basin was adopted as a ground truth for the purpose of this research. The samples were carefully located and validated based on this map. Two groups of samples were selected, namely “water” and “no-water.” The first group was composed of open water bodies, such as rivers, lakes, artificial reservoirs, and waterways. The second group was composed of anything different from open water bodies, such as fields, trees, mountains, cities, highways, plantations, clouds, and shadows of clouds. Therefore, classifiers were expected to identify any open water body located in the time series images as “water” and anything else as “no-water.” In total, the samples consisted

of 250 geographical features, which were spread all over the reference image. Half of the samples were “water” and the other half were “no-water.” One part of these samples was used as a training set; their “water” and “no-water” labels were revealed to the classifier during training. Another part of these samples was used as a validation set, and their “water” and “no-water” labels were not revealed to the classifier in order to verify if the previous step (i.e., the training) was successful.

Table 2. Information about the time series images. Information in the first column refers to Landsat 8 OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor) Collection 1 Level-1 Data Products [50]. In the second column, the word ‘none’ refers to the absence of any environmental disaster at the respective date.

Landsat 8 Image (Scene)	Date of the: Image Acquisition/ Environmental Disaster
LC08_L1TP_217074_20130903_20170502_01_T1	3 September 2013 None
LC08_L1TP_217074_20140805_20170420_01_T1	5 August 2014 None
LC08_L1TP_217074_20151112_20170402_01_T1	12 November 2015 5 November 2015
LC08_L1TP_217074_20160810_20170322_01_T1	10 August 2016 None
LC08_L1TP_217074_20170829_20170914_01_T1	29 August 2017 None
LC08_L1TP_217074_20181222_20181227_01_T1	22 December 2018 None
LC08_L1TP_217074_20190904_20190917_01_T1	4 September 2019 None
LC08_L1TP_217074_20200501_20200509_01_T1	1 May 2020 None

3.4. Method

3.4.1. Data Preprocessing

Adding Bands 1–7 of Landsat 8 Scene as Raster Layers

Prior to analyzing the images using classifiers, each time series image was processed in accordance with the tasks outlined in the upper part of Figure 1. The first step in this process was to create a new project of Qgis and add seven bands from the time series image as raster layers to the project. This step is important because it organizes the bands, i.e., the input data, to be used in the next step. Table 3 shows the bands that were added to the new project and their specifications.

Table 3. Information about the bands that were added to the new project.

Landsat 8 Band Used.	Wavelength	Resolution
Band 1—Coastal Aerosol	0.43–0.45 μm	30 m
Band 2—Blue	0.45–0.51 μm	30 m
Band 3—Green	0.53–0.59 μm	30 m
Band 4—Red	0.64–0.67 μm	30 m
Band 5—Near Infrared (NIR)	0.85–0.88 μm	30 m
Band 6—SWIR 1	1.57–1.65 μm	30 m
Band 7—SWIR 2	2.11–2.29 μm	30 m

Building Band Composition R(4)G(3)B(2)

Next, to facilitate visual inspections, a band composition was created. A band composition is a single multiband raster built to represent the properties of objects in a scene. The band composition was created taking into account the attribution of different color values to the pixels of the raster. For example, we used the band composition R(4)G(3)B(2) in this work, which means that the raster presents natural colors. An image that resulted from this step is demonstrated in Figure 3.

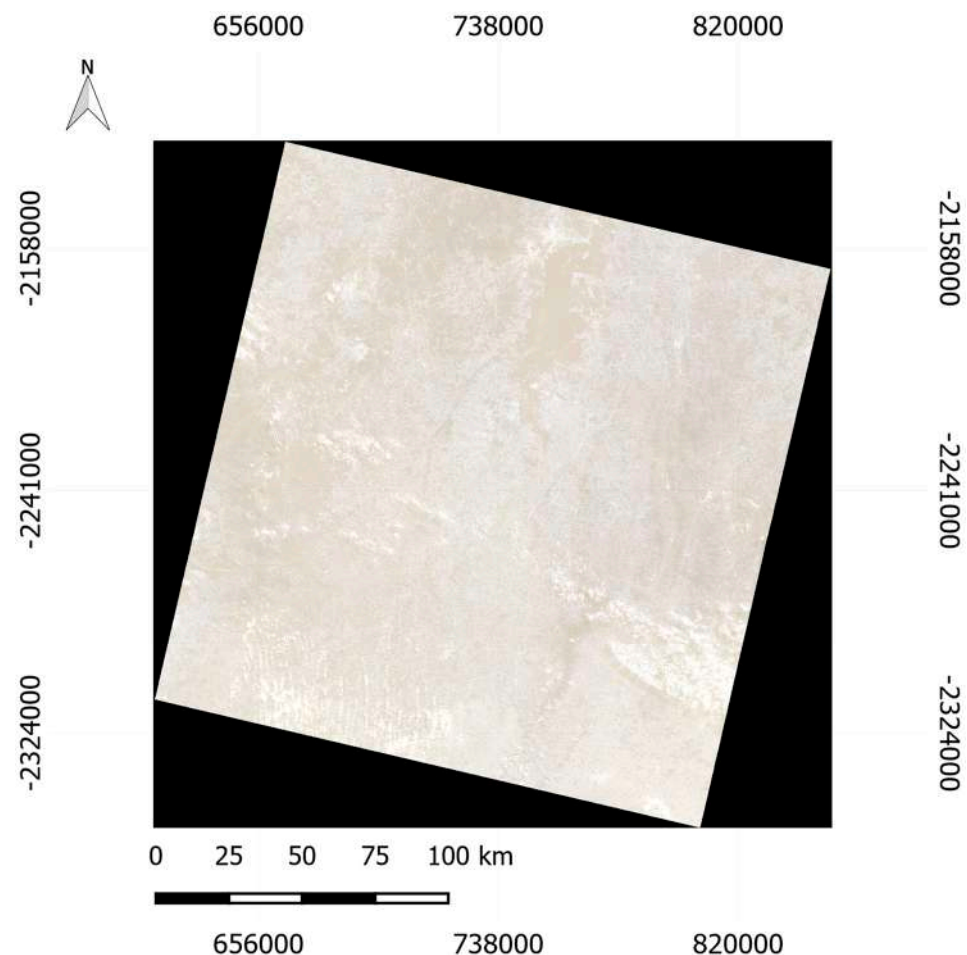


Figure 3. Example of a Landsat 8 scene related to the Doce River basin in Brazil presented as a band composition R(4)G(3)B(2). The Doce River and its tributaries are not visible because this image was not zoomed. In the coordinate grid, the UTM false northing (10,000 km) was not added to the N coordinate.

Considering that a Landsat 8 band is a two-dimensional array of pixels, if x and y are coordinates of the pixels, then a band n can be referred to as $bn(x, y)$. Therefore, the band composition R(4)G(3)B(2) joins Band 4 = $b4(x, y)$, Band 3 = $b3(x, y)$, and Band 2 = $b2(x, y)$. Since a band composition can be represented as a tridimensional array composed of three overlapped bands, Equation (1) symbolizes the composite R(4)G(3)B(2), which is referred to as bc in this work.

$$bc(x, y) = \begin{bmatrix} b4(x, y) \\ b3(x, y) \\ b2(x, y) \end{bmatrix} \quad (1)$$

Performing Histogram Stretching, Choosing the Coordinate Reference System, and Adding Band 8

The third step applies histogram stretching to bc in order to render the band composite. To perform the histogram stretching, Qgis builds a color table that is based on the mean and standard deviation calculated for the three bands of the composite R(4)G(3)B(2).

Equations (2) [67] and (3) [67] represent the mean and the standard deviation, which were calculated to build the color table. In both equations, the image formed averaging K images $g(x, y)$ that are referred to as $f(x, y)$. The function $g_i(x, y)$ represents an image, and the total number of images $g(x, y)$ is referred to as K . The identifier of each image $g(x, y)$ is represented by the index i . The coordinates of the pixels in the functions are represented by the pair x and y . The image formed calculating the standard deviation is referred to as $h(x, y)$ in Equation (3). An enhanced image [68] resultant from this step is exemplified in Figure 4.

$$f(x, y) = \frac{1}{K} \sum_{i=1}^K g_i(x, y) \quad (2)$$

$$h(x, y) = \sqrt{\frac{1}{K} \sum_{i=1}^K (g_i(x, y) - f(x, y))^2} \quad (3)$$

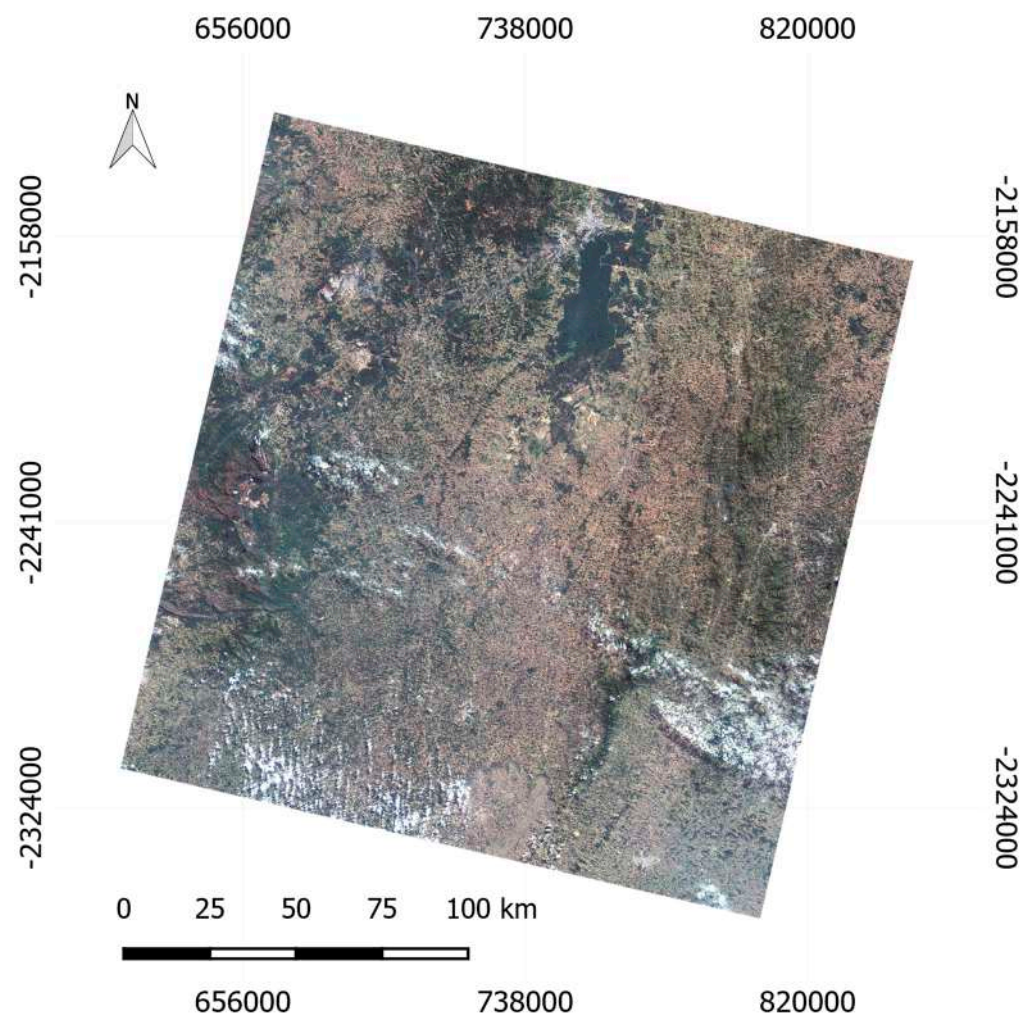


Figure 4. Example of an enhanced Landsat 8 scene of the Doce River basin in Brazil. The Doce River and its tributaries are not visible because this image was not zoomed.

In the following two steps, two images were added to a new project in Qgis: (1) the enhanced image; (2) the Band 8 panchromatic (PAN) (0.50–0.68 μm) 15 m from the same Landsat 8 scene. For this new project, we chose the coordinate reference system (CRS) WGS84 with the UTM map projection.

Performing Pan-Sharpener

The sixth step involved pan-sharpening. Performing pan-sharpening is important because its interpolation procedure makes the overlap of the panchromatic and the enhanced images at a finer scale. This finer-scale overlap was important because working with a 15 m image enabled the analysis of tributaries of the Doce River and other narrow rivers present in the scene, although the Doce River could be found in the image using the 30 m original spectral bands. We also used pan-sharpening because we repeated the experiments performed by Dias et al. in [31] in order to make a valid comparison.

Pan-sharpening is based on projection substitution [68] or component substitution (CS) [52]. The CS method is represented by Equations (4) [52] and (5) [69]. Table 4 shows the elements that compose Equations (4) and (5) and their respective meanings. The objective of the weight w in Equation (5) is to measure the degrees of spectral overlap among the multispectral and panchromatic channels.

Table 4. Information about the CS method expressed in Figure 5 and Equations (4) and (5).

Element of Equation	Meaning
MS	Multispectral image
$\hat{M}S$	Pan-sharpened image
$\widetilde{M}S$	Multispectral image interpolated at the scale of the panchromatic image
K	Subscript k indicates the k th spectral band
$g = [g_1, \dots, g_k, \dots, g_N]$	Vector of injection gains
P	Histogram-matched panchromatic image
$w = [w_1, \dots, w_i, \dots, w_N]$	Weight vector

The pan-sharpening process based on the CS method is exemplified in Figure 5. According to Figure 5, the pan-sharpening process requires four main procedures. In the first procedure, the multi-scale image is interpolated for matching the scale of the panchromatic image. The second procedure calculates the intensity component based on Equation (5). In the third procedure, the intensity component and the histograms of the panchromatic image are matched. The fourth procedure injects the extracted details based on Equation (4).

$$\hat{M}S_k = \widetilde{M}S_k + g_k (P - I_L), \quad k = 1, 2, \dots, N \quad (4)$$

$$I_L = \sum_{i=1}^N w_i \widetilde{M}S_i \quad (5)$$

Performing Histogram Stretching and Computing Second-Order Image Statistics

The seventh step involved the stretching of the histogram of the pan-sharpening resulting image. In this step, the histogram was stretched in the same way it was previously explained. Equations (2) [67] and (3) [67] were also calculated in the seventh step.

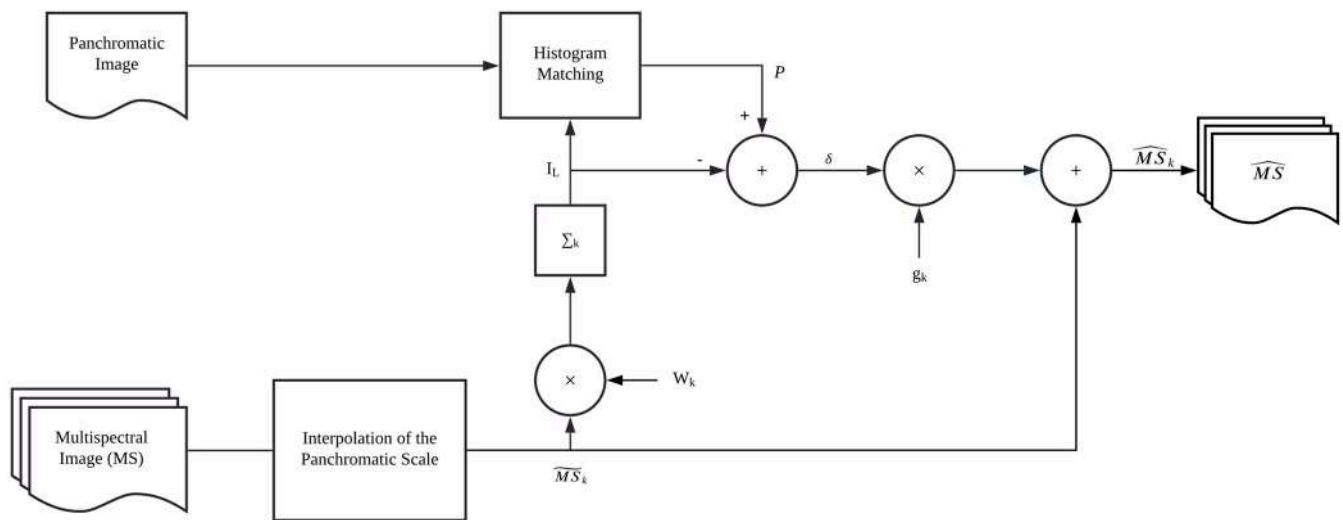


Figure 5. Flowchart of the component substitution method representing a generic pan-sharpening procedure. Adapted from [69]. In this sketch, $\delta = (P - I_L)$. This flowchart is a graphical representation of the procedures expressed in Table 4 and Equations (4) and (5).

In order to calculate the global mean and standard deviation for each band, the eighth step computed the second-order image statistics. These statistics take into consideration the spatial distribution of the pixels in an image to build a geometric model. Therefore, the model statistically represents the image. Enhancing parameters do not affect the statistics.

In second-order image statistics, the slope of the power spectrum tends to be close to negative two. Equation (6) represents the power spectrum of an M by M image [70]. In Equation (6), F is the Fourier transform of the image. Equations (7) and (8), respectively, represent the two-dimensional frequencies u and v in polar coordinates. In these equations, \varnothing are directions [70].

$$S(u, v) = \frac{|F(u, v)|^2}{M^2} \quad (6)$$

$$u = f \cos \varnothing \quad (7)$$

$$v = f \sin \varnothing \quad (8)$$

As Reinhard et al. wrote in [70]: ‘(...) averaging over all directions \varnothing and all images in the ensemble, it is found that on log-log scale power as function of frequency f lies approximately on a straight line (...) This means that spectral power as function of spatial frequency behaves according to a power law function. Moreover, fitting a line through the data point yields a slope α of approximately negative two for natural images:

$$S(f) \propto Af^2 = Af^{-2-\eta} \quad (9)$$

Here, $\alpha \approx -2$ is the spectral slope, η is its deviation from -2 , and constant A describes the overall image contrast.

Equation (10) represents how to obtain the contrast using image statistics, i.e., the standard deviation of all pixel intensities divided by the mean intensity (σ/μ).

$$\frac{\sigma^2}{\mu^2} = \sum_{(u, v)} S(u, v) \quad (10)$$

Performing Sampling or Domain Adaptation

The task that was performed in the last step of the data preprocessing depended on which image in the time series was processed. When the reference image was processed, sampling was performed as the last step. The details related to the selected samples have

already been presented in Section 3.3. Sampling is important because samples are used as training data to model the study area for contextual or non-contextual classifiers. When any other image in the time series was processed, DA was required.

DA is important for performing transfer learning, i.e., for making the model created for the reference image useful for application in any other image in the time series. The Maximum Mean Discrepancy method of DA [54] inspired us to take the global mean and standard deviation of each Landsat 8 band (see eighth step) into consideration to develop our DA approach. Equations (11)–(13) express our DA approach. To perform DA, we changed the model that statistically represents the reference image by recalculating the global mean and standard deviation of each Landsat 8 band (i.e., n). Equation (11) shows how to obtain the mean C_n between the results of the second order statistics of n in the reference image (i.e., A_n) and in the time series image currently being processed (i.e., B_n) for both the non-contextual and contextual classifiers.

$$C_n = \frac{1}{2}(A_n + B_n) \quad (11)$$

Equation (13) shows how to obtain a new global mean or standard deviation of band n for the contextual classifier, for which

$$D_n = A_n - C_n \quad (12)$$

$$E_n = \begin{cases} B_n + |D_n|, & \text{if } D_n < 0 \\ B_n - D_n, & \text{otherwise} \end{cases} \quad (13)$$

In other words, our DA approach calculated C_n and E_n , which are new values for the second order statistics of the band n in the reference image and are useful to perform, respectively, for non-contextual and contextual classifications of any time series image in this research.

End of Data Preprocessing

Preprocessing is important because it sets the data in an appropriate manner for being useful in the learning and classification phase. In order to avoid the samples that would bias the classifiers to classify the Doce River only as “water”, no areas located on the Doce River were selected as samples. Therefore, the care we took while selecting the samples is the main contribution of our preprocessing, which prevented the experiments to yield biased results.

3.4.2. Learning and Classification

Importing Models or Performing Training, Validation, and Test of Classifiers

The tasks that were performed in the first step of the learning and classification approach depended on which image in the time series was processed. When the reference image was processed, the training, validation, and test of classifiers were performed. In order to train the classifiers, the selected samples, the statistical model, and the enhanced 15 m image were used. The training is essential to provide the contextual or non-contextual classifiers with models. For classifiers, the models are used as references based on the selected samples. These models are used by the classifiers to categorize features of interest present in the 15 m time series images [71].

The models created for the reference image were made available to the other time series images. When any other image in the time series was processed, the training of classifiers was not performed because the models related to the reference image were imported for classification. The current paragraph described an example of the practical use of transfer learning.

Creating Image Contextual and Non-Contextual Classifications

Image contextual and non-contextual classifications were created. The image classifications are essential to allow a DMS to group features of interest into classes. Classes are defined in accordance with similarities of features of an interest's characteristics. As mentioned before, it is expected that our strategy can group features of interest into two different classes: "water" and "no-water". Each class is represented by a different color in the images that result from the classification processes, e.g., black or white. Therefore, the classifications generated only binary images [71] in this research, no matter if a classification was performed by a contextual or non-contextual classifier.

The contextual and non-contextual classifiers chosen for this research were Boost and Decision Tree (DT), respectively. Dias et al. in [31] concluded that the Boost and DT classifiers were the best choices to classify the image of the Doce River in order to satisfy the anomaly detection strategy. Therefore, this research applies the Boost and DT to classify the time series images in agreement with the study published by Dias et al. in [31]. In other words, Boost and DT allow our current strategy to achieve positive results and spend less time than other classifiers. The Boost and DT classifiers are described in detail as follows.

Boost classifier

The Boost classifier works by considering a weight distribution on the training samples and a set of iteratively weak classifiers. While working, the classifier attributes more weight to samples misclassified by the previous iterations [46]. The Boost combines the weak classifiers to be stronger. This combination followed by a threshold composes the final strong classifier. The following lines describe the procedures performed by the Boost algorithm.

- (1) Definition of training sets.

Here, $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ stands for a training set, for which $y_i \in \{-1, 1\}$. The t th distribution over training samples is represented by $D_t(x_i, y_i)$. The samples and their labels are, respectively, represented by x_i and y_i .

- (2) Initialization of weights.

Equation (14) expresses the initialization of weights [46].

$$D_0(x_i, y_i) = \frac{1}{N}, \quad i = 1, 2, \dots, N \quad (14)$$

- (3) Training loop. For $t = 1, 2, \dots, T$ (T is the maximum training number):

- (a) The training of a simple linear classifier h_j is performed for each feature j . A simple linear classifier is a classifier restricted to use a single feature. The simple linear classifier is represented by Equation (15) [46], for which the value of the j th feature of the sample x_i is denoted by $x_{i,j}$. Moreover, the threshold value of the j th feature of the sample x_i is expressed by $\theta_{i,j}$, and the direction of the inequality sign is decided by $p_j \in \{-1, 1\}$. The error ε_t is evaluated regarding $D_t(x_i, y_i)$, in accordance to Equation (16) [46].

$$h(x_i) = \begin{cases} 1, & \text{if } p_j x_{i,j} < p_j \theta_{i,j} \\ -1, & \text{otherwise} \end{cases} \quad (15)$$

$$\varepsilon_j = \sum_{i: y_i \neq h_j} D_t(x_i, y_i) \quad (16)$$

- (b) The weak classifier h_t with the lowest error ε_t is chosen.
- (c) The loop stops for $\varepsilon_t \geq 1/2$.

- (d) the weight α_t is calculated for $\epsilon_t < 1/2$. The α_t is the weight assigned to the classifier h_t . Equation (17) represents the weight α_t .

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right) \tag{17}$$

- (e) The way the weights are updated is expressed by Equation (18) [46]. In Equation (18), Z_t is a normalized constant. This constant is computed to insure that $D_t(x_i, y_i)$ represents a true distribution, in accordance with Equation (19) [46].

$$D_{t+1}(x_i, y_i) = \frac{D_t(x_i, y_i) e^{-\alpha_t y_i h_t(x_i)}}{Z_t} \tag{18}$$

$$\sum_{i=1}^N D_t(x_i, y) = 1 \tag{19}$$

- (4) Output of the final classifier.

Equation (20) represents this final procedure of the Boost [46].

$$H(x) = \text{sign} \left(\sum_{t=1}^{Tk} \alpha_t h_t(x) \right) \quad \text{sign}(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ 1, & x > 0 \end{cases} \tag{20}$$

DT classifier

The DT classifier finds the most probable decision for achieving an objective [48]. In other words, DT is important to perform decision analysis. Although decision trees can be built by many different algorithms, the use of information entropy and information gain to build decision trees is well-established in the scientific community. Information entropy and information gain are bases for developing the Iterative Dichotomiser 3 (ID3). The ID3 uses data sets to generate the decision trees. Equations (21) and (22) [49] represent, respectively, the information entropy and information gain.

$$\text{Ent}(D) = - \sum_{d=1}^k p_k \log_2 p_k \tag{21}$$

$$\text{Gain}(D, a) = \text{Ent}(D) - \sum_{v=1}^V \frac{|D_v|}{|D|} \text{Ent}(D_v) \tag{22}$$

In Equations (21) and (22), $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ represents the training sample, and the training sample number is represented by $|D|$. The entropy of D , i.e., $\text{Ent}(D)$, measures the amount of uncertainty in the data set D . In other words, the impurity of the collection of samples in D is characterized by $\text{Ent}(D)$. The ratio of every type for every current sample set is represented by p_k ($k = 1, 2, \dots, |D|$). The attribute set of D , e.g., texture, shape, color, etc., is represented by $A = \{a_1, a_2 \dots a_d\}$, for which $d = \{1, 2, \dots, k\}$. The expected reduction in the entropy of D regarding the choice of the attribute a is referred to as gain of D , i.e., $\text{Gain}(D, a)$. For each attribute a_i , V is a set of features. V is composed of different v values. Therefore, $V = \{a_i^1, a_i^2, \dots, a_i^v\}$ stands for a set of features, such as $a_1^1 = \text{red}$, $a_1^2 = \text{yellow}$, etc. The sample subset regarding the value a_i^v from a_i in D is referred to as D_v . The number of the current samples of the subset is referred to as $|D_v|$.

Subtracting Images and Establishing Chronology

Finally, to subtract the binary images generated by the contextual and non-contextual classifiers, the difference in all pairs of corresponding pixels from both images was computed [67]. The importance of the subtraction is related to the automatic identification of incongruence or congruence. There is incongruence wherever the subtraction results in a

pixel equal to one. Otherwise, there is congruence. The presence of an incongruence can expose the existence of an anomaly, according to [28]. Subtraction is the final step to detect an anomaly in the spatial context.

Equation (23) expresses the subtraction [67]. In Equation (23), the images resultant from contextual and non-contextual classifications are represented, respectively, by $h(x, y)$ and $f(x, y)$. The image resulting from the subtraction is represented by $g(x, y)$. The variables x and y are coordinates of the pixels.

$$g(x, y) = f(x, y) - h(x, y) \quad (23)$$

Additionally, by establishing the chronologies regarding the time series and the occurrence of the incongruence, it is possible to detect an anomaly in the temporal context, i.e., to determine the period during which the anomaly was present in a time series. The anomaly detected, analyzed, and categorized in this research was of the component model drift type, in accordance with [27].

End of Learning and Classification

Learning and classification are important, for example, to analyze image time series from remote sensing in different contexts, such as the spatial and temporal contexts. Our main contribution regarding learning and classification is completing research that can help machines detect and categorize occurrences of anomalies in image time series in accordance with Kittler's taxonomy [27].

Regarding methodological limitations, challenges posed by the time series can be resolved by selecting the correct domain adaptation method. Choosing the domain adaptation method is a hard task, because each domain requires a different adaptation in order to allow our strategy to detect anomalies. Moreover, in this research, the adaptation demanded empirical experiments that were carried out for each domain to find an effective solution. This demand suggests that the domain adaptation still depends on human decisions and, therefore, can limit the autonomy of a DMS to analyze the image time series in future studies. However, various methods of domain adaptation have been published by the scientific community that can help reduce the effort of researchers to find domain adaptation solutions for DMS autonomy problems.

4. Results

In order to qualify the anomaly in accordance with Kittler's taxonomy [27], three evaluation measure mechanisms were used. The three mechanisms are sensory data quality assessments, contextual and non-contextual classifiers, and incongruence indicators. Jointly, these three mechanisms are essential to successfully qualify anomalies.

The occurrence of the anomaly *component model drift* is associated with some specific conditions that need to be fulfilled, such as the following: (1) presence of an image time series; (2) high sensory data quality; (3) sample selection based on components, e.g., water and no-water; (4) the reuse of the model that was built based on sample selections of components in the reference image; (5) contextual and non-contextual classification; (6) incongruence. The absence of any of the six conditions makes the identification of this type of anomaly unfeasible. Condition (1) was confirmed because this research was based on the use of an image time series. Condition (2) was confirmed because the time series images are qualified with value "9" in the USGS. This value indicates the high quality of Landsat 8 images. Conditions (3) and (4) were confirmed performing the ninth and tenth steps of our strategy, respectively. Condition (5) was confirmed by the results achieved applying the boost and DT classifiers, respectively. Condition (6) was confirmed by performing the last step of our strategy.

Assessments were thoroughly and meticulously carried out taking into account the use of our strategy on a Landsat 8 image time series. In order to meet assessment requirements, each image in the time series was accurately cropped in order to generate a set of smaller image tiles. The total number of tiles in this research is 8400. Cropping is important because

assessment approaches described in the scientific literature are normally applied to small images. The image tile resolution in this research is 151×193 (height and width in pixels). In contrast, Landsat 8 images are large, with resolutions of $15,705 \times 15,440$ (height and width in pixels). Therefore, the assessment of a Landsat 8 image that has not been cropped is inadequate.

Our strategy was quantitatively validated by the use of the metrics accuracy, precision, recall [39], and F-measure [72]. Accuracy or Overall Accuracy (OA) measures the efficiency of results. Precision measures the relevancy of results. Recall measures the quantity of truly relevant results. F-measure measures the balance between precision and recall. Equations (24)–(27) express the accuracy, precision, recall, and F-measure, respectively. In these equations, the number of images is expressed by M ; images in which congruences are truly detected are represented by TP, i.e., true positives; images in which incongruences are truly detected are represented by TN, i.e., true negatives; images in which incongruences are falsely detected are represented by FN, i.e., false negatives; images in which congruences are falsely detected are represented by FP, i.e., false positives.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{M} \quad (24)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (25)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (26)$$

$$\text{F - Measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (27)$$

4.1. Experimental Results

Quantitative evaluations of congruence detections on time series images of the Doce River basin are presented as contingency tables [73–75] in Tables 5–12.

Table 5. Contingency table that assists with the detection of congruences in the Landsat 8 scene of the Doce River basin acquired in 2013 and consisting of 8400 tiles (TP—true positive, FP—false positive, FN—false negative, TN—true negative).

	Congruent Event	Incongruent Event
Detection congruent	TP = 8367	FP = 0
Detection incongruent	FN = 33	TN = 0

Table 6. Contingency table that assists with the detection of congruences in the Landsat 8 scene of the Doce River basin acquired in 2014 and consisting of 8400 tiles (TP—true positive, FP—false positive, FN—false negative, TN—true negative).

	Congruent Event	Incongruent Event
Detection congruent	TP = 8349	FP = 0
Detection incongruent	FN = 51	TN = 0

Table 7. Contingency table that assists with the detection of congruences in the Landsat 8 scene of the Doce River basin acquired in 2015 and consisting of 8400 tiles (TP—true positive, FP—false positive, FN—false negative, TN—true negative).

	Congruent Event	Incongruent Event
Detection congruent	TP = 8228	FP = 2
Detection incongruent	FN = 104	TN = 66

Table 8. Contingency table that assists with the detection of congruences in the Landsat 8 scene of the Doce River basin acquired in 2016 and consisting of 8400 tiles (TP—true positive, FP—false positive, FN—false negative, TN—true negative).

	Congruent Event	Incongruent Event
Detection congruent	TP = 8361	FP = 0
Detection incongruent	FN = 39	TN = 0

Table 9. Contingency table that assists with the detection of congruences in the Landsat 8 scene of the Doce River basin acquired in 2017 and consisting of 8400 tiles (TP—true positive, FP—false positive, FN—false negative, TN—true negative).

	Congruent Event	Incongruent Event
Detection congruent	TP = 8328	FP = 0
Detection incongruent	FN = 72	TN = 0

Table 10. Contingency table that assists with the detection of congruences in the Landsat 8 scene of the Doce River basin acquired in 2018 and consisting of 8400 tiles (TP—true positive, FP—false positive, FN—false negative, TN—true negative).

	Congruent Event	Incongruent Event
Detection congruent	TP = 8318	FP = 0
Detection incongruent	FN = 82	TN = 0

Table 11. Contingency table that assists with the detection of congruences in the Landsat 8 scene of the Doce River basin acquired in 2019 and consisting of 8400 tiles (TP—true positive, FP—false positive, FN—false negative, TN—true negative).

	Congruent Event	Incongruent Event
Detection congruent	TP = 8301	FP = 0
Detection incongruent	FN = 99	TN = 0

Table 12. Contingency table that assists with the detection of congruences in the Landsat 8 scene of the Doce River basin acquired in 2020 and consisting of 8400 tiles (TP—true positive, FP—false positive, FN—false negative, TN—true negative).

	Congruent Event	Incongruent Event
Detection congruent	TP = 8302	FP = 0
Detection incongruent	FN = 98	TN = 0

4.2. Interpretation of the Results

It is apparent from Tables 5–12 that the time series images are associated with a high quantity of true positives. Meanwhile, only a single image (Table 7) presents false positives, and their numbers are of little significance for this research. Moreover, no significant false negatives were found in the time series images. These are all encouraging results. These encouraging results were also favored by the expressive amount of: (1) Samples that were used as training and validation sets, i.e., 250 samples, 125 of “water” and 125 of “no-water”. (2) Pre-processed images that were used as test sets, i.e., eight entire Landsat 8 images. Those images have large amounts of heterogeneous data from remote sensing.

4.2.1. Accuracy, Recall, Precision, and F-Measure

Table 13 presents the results related to the accuracy, recall, precision, and F-measure of our strategy. Table 14 compares the accuracy, precision, recall, and F-measure obtained from other studies against the results achieved by this study, which is shown in the fifth

row. No significant differences were found between results achieved by our research when compared to the study that presented the best values (first row). The results, as shown in Table 14, indicate that there is a quantitative consistence between values found in this study and those found in the scientific literature.

Table 13. Accuracy, precision, recall, and F-measure of the proposed strategy for different Landsat 8 images. These metrics were calculated based on Equations (24)–(27) and the information presented in Tables 5–12, respectively.

Landsat 8 Scene	Accuracy	Precision	Recall	F-Measure
LC08_L1TP_217074_20130903_20170502_01_T1	99.61%	100%	99.61%	99.81%
LC08_L1TP_217074_20140805_20170420_01_T1	99.39%	100%	99.39%	99.69%
LC08_L1TP_217074_20151112_20170402_01_T1	98.74%	99.98%	98.75%	99.36%
LC08_L1TP_217074_20160810_20170322_01_T1	99.54%	100%	99.54%	99.77%
LC08_L1TP_217074_20170829_20170914_01_T1	99.14%	100%	99.14%	99.57%
LC08_L1TP_217074_20181222_20181227_01_T1	99.02%	100%	99.02%	99.51%
LC08_L1TP_217074_20190904_20190917_01_T1	98.82%	100%	98.82%	99.41%
LC08_L1TP_217074_20200501_20200509_01_T1	98.83%	100%	98.83%	99.41%
Results obtained averaging the values of this table (2014–2020)	99.07%	99.99%	99.07%	99.53%

Table 14. Comparison of the accuracy, precision, recall, and F-measure of this study with others. The information is ordered by level of accuracy. Our study is in the fifth row but presents the second highest levels of precision, recall, and F-measure in comparison with other studies. These results mean that our study is placed among the best when compared against the state-of-the-art.

Study	Accuracy	Precision	Recall	F-Measure
[31]	99.76%	100.00%	99.76%	99.88%
[58]	99.59%	————	————	————
[57]	99.20%	91.85%	53.55%	67.66%
[59]	99.14%	————	————	————
Ours	99.07%	99.99%	99.07%	99.53%
[7]	98.49%	83.84%	83.66%	83.76%
[56]	98.00%	————	————	————
[55]	91.20%	98.10%	95.7%	96.88%
[8]	88.68%	90.62%	79.62%	84.76%
[16]	84.00%	63.00%	81.00%	70.88%
[60]	81.18%	————	————	————
[9]	78.00%	82.00%	75.00%	78.34%

In this research, we reproduced experiments related to the Doce River performed by Dias et al. in [31], but by emphasizing the congruent results. The values presented in the first row of Table 14 are results related to congruences found while reproducing their experiments. These current values differ from the values originally found in [31] because Dias et al. emphasized the incongruent results [31] instead of the congruent ones. This change was needed because our research deals with congruences more often than with incongruences due to the presence of the same geographical settings in all images of the time series.

The highest values of accuracy, precision, recall, and F-measure among the presented studies were achieved by our research, except for [31]. However, the present research analyzed an image time series, as well as in [7–9,16,56–60,72], whereas Dias et al. investigated only a single image in [31]. In order to analyze the drift among the different images of the time series, we needed to use transfer learning in our strategy in contrast to [31]. Transfer learning can limit the results according to [53]. That is the reason why the results of our research have lower values in comparison with [31]. If we now turn to the results reported by [57–59], although the accuracy is higher than ours, it is not possible to claim that those

studies achieved better results than ours because the values of precision, recall, and F-measure were lower in [57] and not reported by [58,59]. Therefore, our research presented the highest values of precision, recall, and F-measure among the presented studies, for which the analyses were performed on an image time series.

4.2.2. Receiver Operating Characteristic and Precision-Recall Curves and Area under Curve Measurements

Figures 6 and 7 show the results obtained in another analysis. This analysis was carried out using the receiver operating characteristic (ROC) and Precision-Recall (Pr-Re) curves for classification results provided by the proposed methodology for each analyzed year. We can see again from the presented ROC profiles that the proposed method, independent of the input image, tends to deliver high True Positive Rates (TPR) and low False Negative Rates. Similarly, the Pr-Re curves usually show a good balance between the Precision and Recall measures. The Area Under Curve (AUC) measurements for both ROC and Pr-Re profiles are included in Figures 6 and 7 and quantify the high effectiveness of the introduced method.

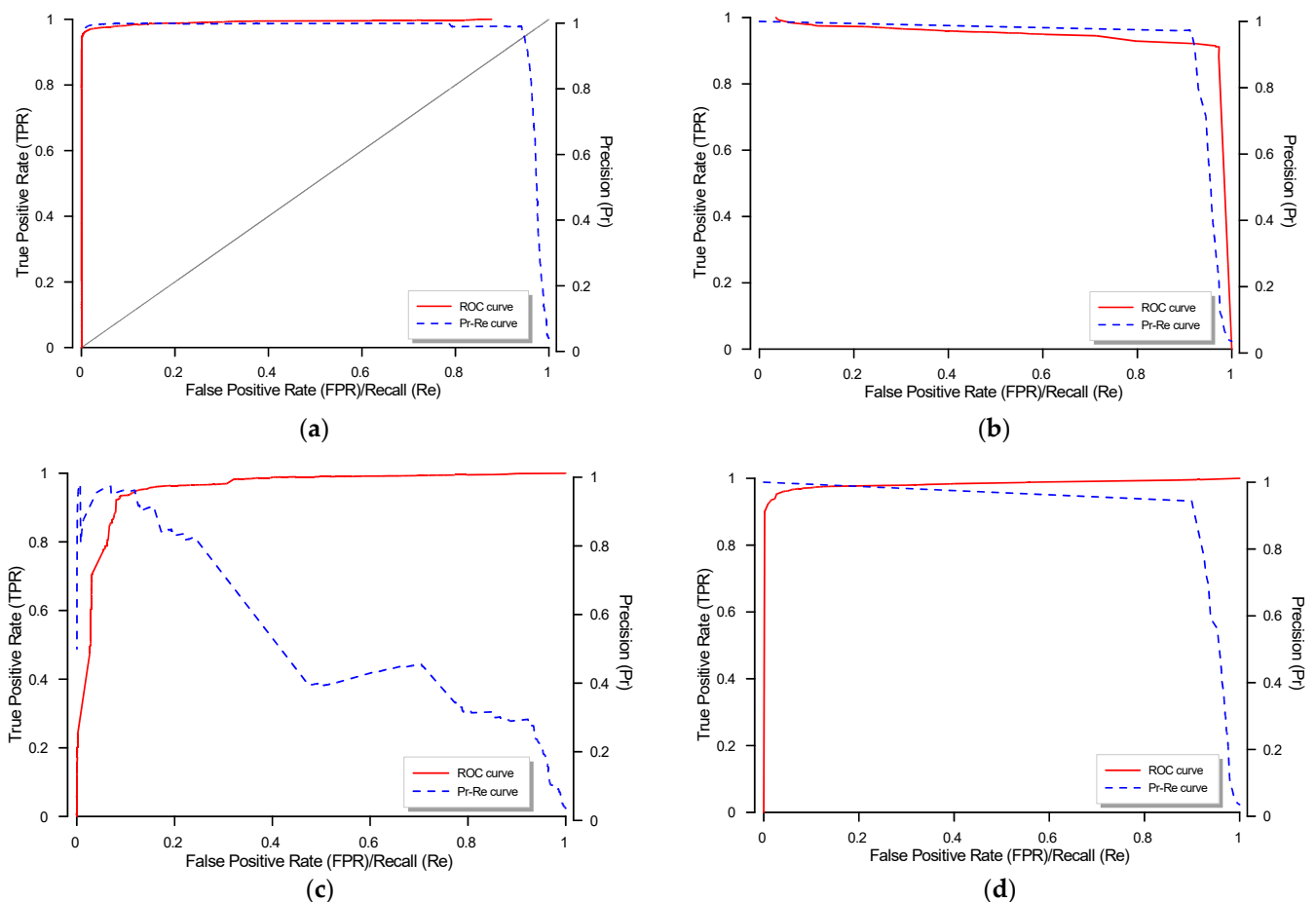


Figure 6. Receiver operating characteristic (ROC) and Precision-Recall (Pr-Re) curves for classification results provided by the proposed methodology for (a) 2013, (b) 2014, (c) 2015, and (d) 2016. According to the ROC profiles, the proposed method tends to deliver high True Positive Rates (TPR) and low False Negative Rates. According to the Pr-Re curves, there is a good balance between the Precision and Recall measures. The Area Under Curve (AUC) measurements for both ROC and Pr-Re profiles quantify the high effectiveness of the introduced method.

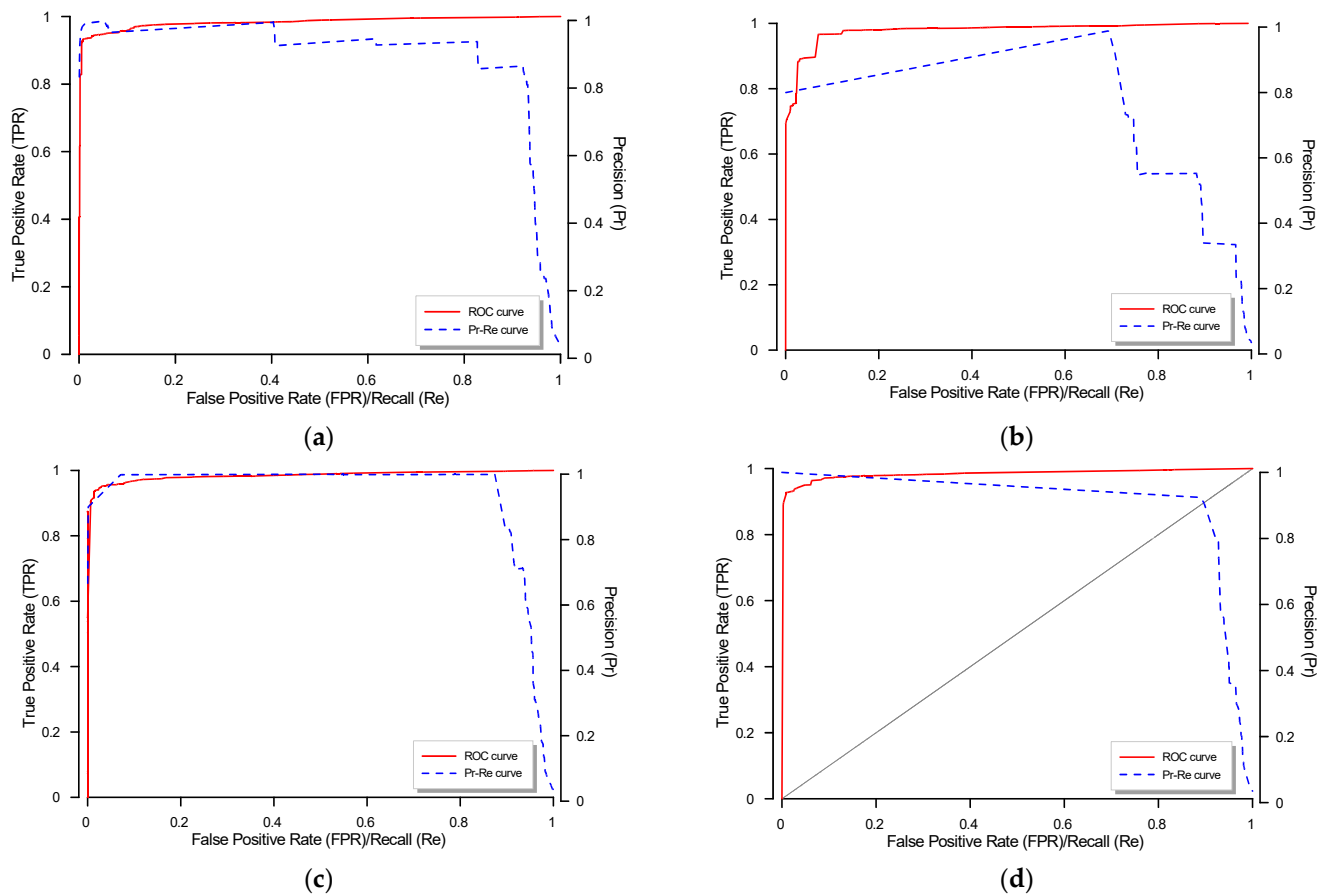


Figure 7. Receiver operating characteristic (ROC) and Precision-Recall (Pr-Re) curves for classification results provided by the proposed methodology for (a) 2017, (b) 2018, (c) 2019, and (d) 2020. According to the ROC profiles, the proposed method tends to deliver high True Positive Rates (TPR) and low False Negative Rates. According to the Pr-Re curves, there is a good balance between the Precision and Recall measures. The Area Under Curve (AUC) measurements for both ROC and Pr-Re profiles quantify the high effectiveness of the introduced method.

4.2.3. Overall Accuracy, Kappa Coefficient, and Its Variance

The results obtained from the application of the proposed strategy on time series images (tiles) taken from the Doce River basin are set out in Figures 8–10. These figures show the Doce River, respectively, before (in 2014), during (in 2015), and after the disaster (in 2016): (a) enhanced image; (b) real image; (c) contextual classifier output; (d) and (f) result overlay on the real image (if those results are present); (e) non-contextual classifier output. The outputs (c) and (e) are the images which are subtracted in the last step of our method in order to highlight incongruent results.

The three Figures 8–10 (a) exhibit part of the Doce River, which is the thick winding line starting at the bottom and ending at the top right corner, and part of a narrow and dark lake at the top, apart from other geographical features; (b) show the real images; (c) present that the classifier Boost is able to recognize both the presence and the absence of pollution in river water; (d) overlay results of the Boost on the real images; (e) present that the classifier decision tree is able to recognize only the absence of pollution in river water, e.g., Figures 8 and 10; (f) overlay results of the decision tree on the real images.

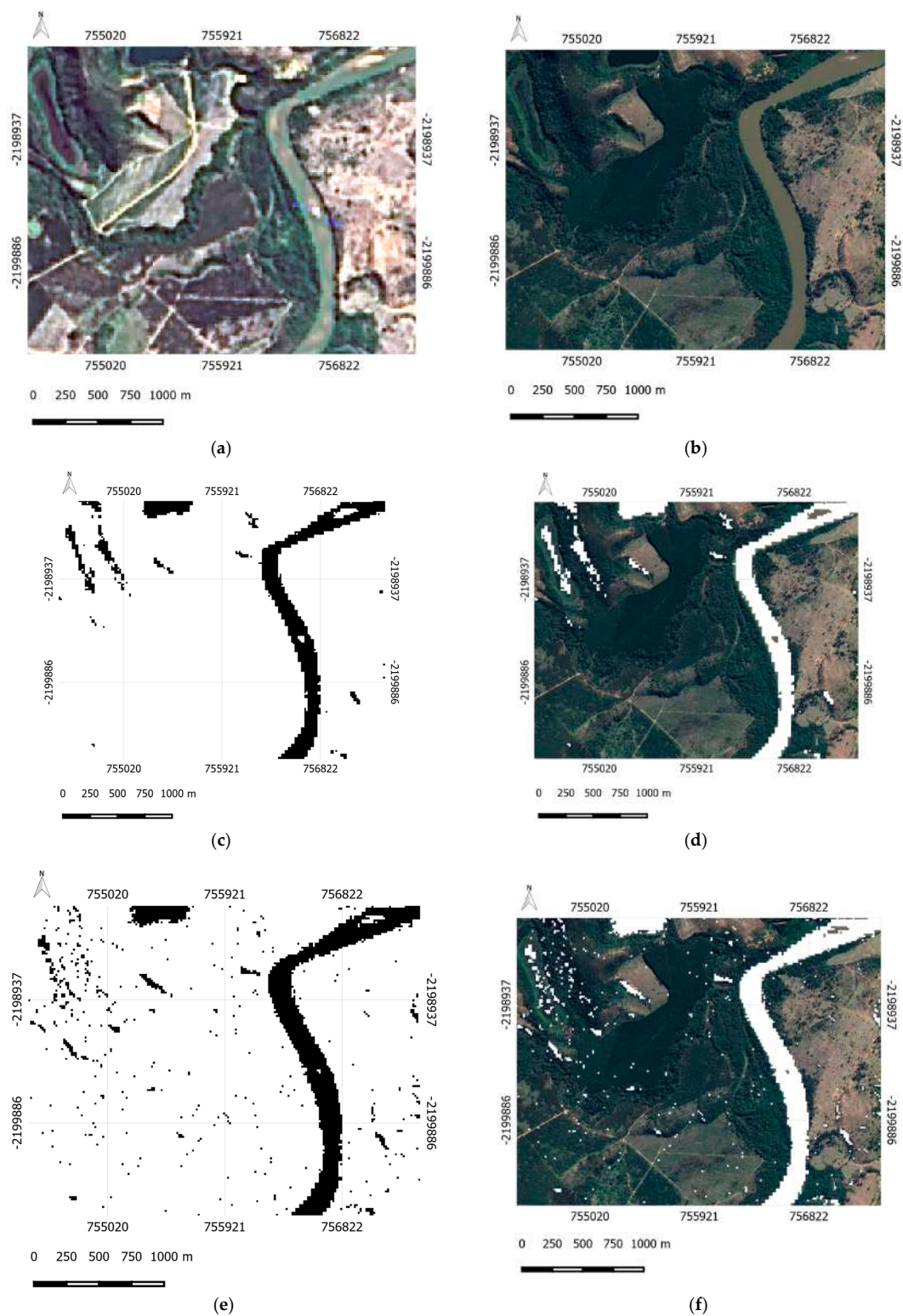


Figure 8. An example of the application of our strategy on an image tile taken from the image of the Doce River acquired by Landsat 8 in 2014: (a) enhanced image, in which there are two water bodies (the thick winding river and a narrow lake on the top); (b) the real image; (c) contextual classifier (boost) output; (d) result overlay of (c) (white) on the real image; (e) non-contextual classifier (decision tree) output; (f) result overlay of (e) (white) on the real image. In (c) and (e), black and white represent pixels classified as water and no-water, respectively.

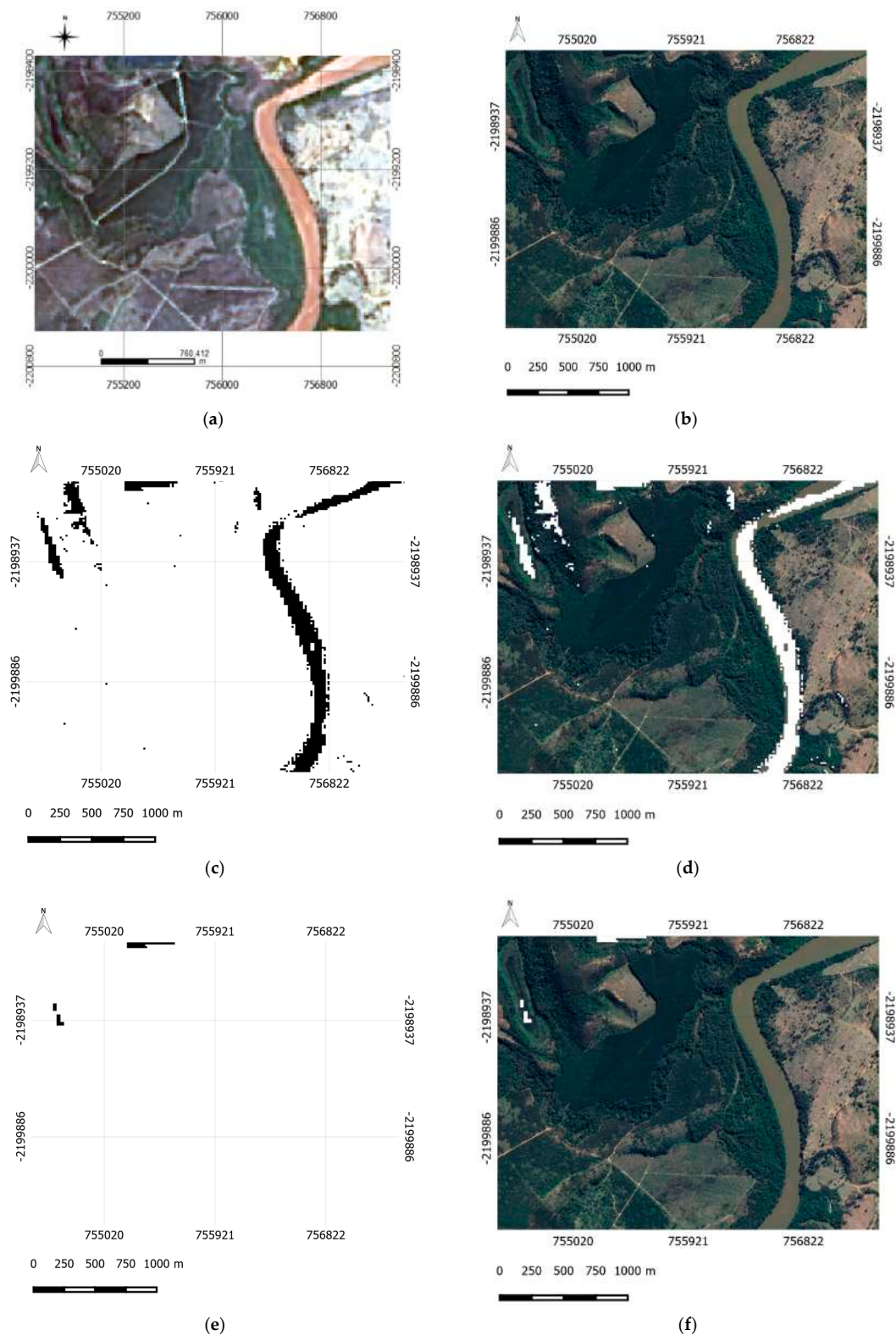


Figure 9. An example of the application of our strategy on an image tile taken from the image of the Doce River acquired by Landsat 8 in 2015: (a) enhanced image, in which there are two water bodies (the thick winding river and a narrow lake on the top); (b) the real image; (c) contextual classifier (boost) output; (d) result overlay of (c) (white) on the real image; (e) non-contextual classifier (decision tree) output; (f) result overlay of (e) (white) on the real image. In (c,e), black and white represent pixels classified as water and no-water, respectively.

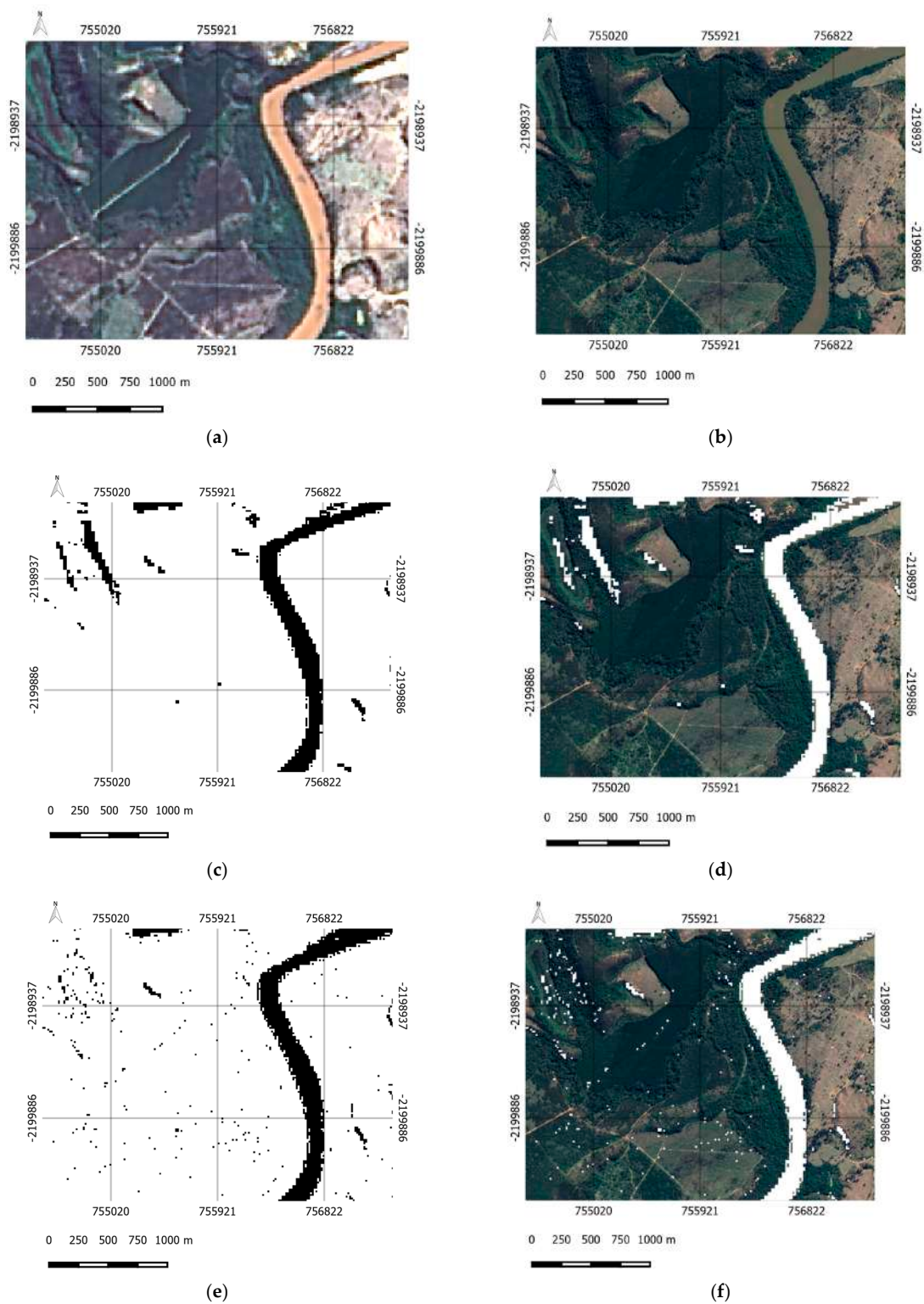


Figure 10. An example of the application of our strategy on an image tile taken from the image of the Doce River acquired by Landsat 8 in 2016: (a) enhanced image, in which there are two water bodies (the thick winding river and a narrow lake on the top); (b) the real image; (c) contextual classifier (boost) output; (d) result overlay of (c) (white) on the real image; (e) non-contextual classifier (decision tree) output; (f) result overlay of (e) (white) on the real image. In (c,e), black and white represent pixels classified as water and no-water, respectively.

As Figure 9 shows, there is a significant difference between the outputs (c) and (e). This difference reveals the occurrence of incongruence and indicates the presence of brown mud in the Doce River water (spatial context) in 2015 (temporal context). On the other hand, the outputs (c) and (e) are congruent in both other cases, i.e., Figures 8 and 10.

Artifacts, which are inherently generated in every stage of image processing, can be present in the results. Such artifacts can be caused, for example, by co-registration errors in the image time series. However, we observed during the experiments that this issue was attenuated due to the high precision of georeferencing of the time series images. If this issue occurs at high intensity, we suggest applying a smoothing filter to remove artifacts (e.g., salt and pepper noise) or smoothing the results. Morphological opening filters could have been applied to eliminate the small (either white or black) points spread on the figures. However, we decided not to apply any filter in order to preserve some narrow results related to incongruences that we detected in the images.

Table 15 presents the Overall Accuracy (OA), kappa coefficient (kappa) [76], and its variance (var.kappa) regarding the classification results depicted in Figures 8–10. Typically, the Boost method is assigned to higher OA and kappa values than the DT method (baseline). In order to verify the significance of the Boost in terms of the kappa coefficient, a statistical test with a 5% significance level was applied. These tests indicated that Boost was inferior to DT for the 2014 input image but superior regarding the other two images.

Table 15. Overall Accuracy (OA), kappa coefficient (kappa) and its variance (var.kappa) regarding the classification results depicted in Figures 8–10. Boost was inferior to DT for the 2014 input image but superior regarding the other two images.

Measure		2014	2015	2016
Boost	OA	0.95576344	0.98661989	0.94951084
	Kappa	0.58219884	0.8214931	0.5424097
	Var.Kapp	1.3828607×10^{-5}	1.0082267×10^{-5}	1.3873299×10^{-5}
DT	OA	0.96155206	0.84809506	0.93850362
	Kappa	0.61801729	0.26374463	0.49920605
	Var.Kapp	0.000013833488	0.0000062081885	0.000012732607
Significance test for kappa	<i>p</i> -value (two-sided)	0.00000000000244	0.00000000000000	0.00000000000000
	Significant?	Yes: Boost < DT	Yes: Boost > DT	Yes: Boost > DT

5. Discussion

Recall and precision are two evaluation mechanisms that measure the success of prediction for cases in which classes are very imbalanced. Both mechanisms present high values whenever the majority of all positive results and correct results are returned by classifiers. In this research, our anomaly detection strategy based on Kittler's taxonomy achieved high values of accuracy, precision, recall, and F-measure when applied to analyze river water pollution in an image time series from remote sensing. Therefore, our strategy efficiently detected anomalies such as *component model drift*, as can be seen by the quantity of truly relevant results. If the studied area was limited to the Doce River water instead of studying the whole Landsat 8 images, the recall would present higher values than the current ones. The reason for our results is that we found false negatives distant from the Doce River water.

This research has introduced a strategy to achieve results that differ from the strategies commonly presented in other research. However, the achieved results showed that our strategy is consistent with other research when it is applied to detect river water pollution in an image time series. The current research found that the strategy developed by Dias et al. in [31] can be expanded to detect anomalies in an image time series from remote sensing instead of only in a single image. The results corroborate the ideas of Dias et al. in [31], who suggested that Kittler's taxonomy can allow computers to detect anomalies by identifying the type of each anomaly based on the recognition of the context in which the anomaly is inserted. We applied our strategy to the same problem of the presence of brown

mud contaminating the Doce River, which was analyzed in [31]. It is important to analyze the same problem because this approach allowed us to evaluate whether our strategy is able to differentiate a single problem in accordance with the context which was recognized or not. Therefore, our research is in agreement with [31] because the results from both research suggest that Kittler's taxonomy can allow computers to recognize different contexts related to the same problem.

Outcomes resulting from this research are in parallel with the ones from [27,39], because, similar to Weinshall et al. in [39] and Kittler et al. in [27], we also studied the potential use of incongruence applied to detect anomalies. For example, a DMS can use incongruence and other conditions to help computers automatically identify if the anomaly detected is a *component model drift* anomaly and the moment when the anomaly occurred. The correct identification of the type of anomaly detected is important, for example, to help computers provide quick and appropriate solutions to analyze the extent of an environmental disaster that resulted in river water pollution. Moreover, many studies, such as [1–6,77], were broadened by our strategy, because all of them provide tools to detect and monitor water pollution while analyzing images from remote sensing.

One unanticipated finding was that the values of accuracy, precision, recall, and F-measure were higher than we expected. We did not expect meaningful results because, according to [53], poor results can be produced by DA even if the domains involved in the transfer learning are similar. However, the differences observed between the best results (see the first row of Table 14) and the ones achieved by this study were not significant.

To the knowledge of the authors, this is the first research which describes the use of Kittler's taxonomy [27] to analyze an image time series from remote sensing. This research supports evidence from previous observations (e.g., [27,31]) which deal with the multifaceted nature of anomalies. The findings of this research differ from the ones presented by other detection methods, such as [7–9,16,33–37,55–60,72,78–83], because those methods deal with anomalies in a time series as if the anomalies were all simple outliers. Moreover, the number of categories of anomalies related to this research is much higher compared to those of other studies, such as [55], due to the use of Kittler's taxonomy.

Regarding negative results, although our strategy used pan-sharpening to better analyze the tributaries of the Doce River, many false negative results were found in the tributaries in many time series images. Future studies on the current topic are therefore recommended. This limitation is probably caused mostly by the resolution of some tiles that is not enough to present narrow tributaries in a clear manner. However, this fact is of little significance for this study, since the occurrence of water pollution in the Doce River was accurately detected in the time series. Nevertheless, the overall results indicate that anomalies such as *component model drift* are detectable by our strategy and consonant with [27].

6. Conclusions

The main goal of the current research was to determine if it would be possible for computers to recognize occurrences of river water pollution in two different contexts by using our machine learning strategy for anomaly detection. This research has found that in order to detect anomalies in an image time series from remote sensing, the machine learning process can be based on Kittler's taxonomy by using our strategy.

While composing Kittler's taxonomy, the anomalies *component model drift* and *unexpected structure and structural components* had their practical applications in remote sensing investigated in this research and in [31], respectively. Each of these studies found a relation between the anomaly detected and one problem in common: the presence of brown mud contaminating the Doce River after an environmental disaster. These findings have significant implications for the understanding of how a machine, such as a DMS, can learn to detect different types of anomalies by differing contexts; for example, spatial and temporal contexts, in which they are inserted, even taking into account the same problem and the same environment. This fact explains why, instead of presenting a solution to a specific

problem in remote sensing, this research focused on developing a strategy for machines to learn how to deal with anomalies in a different context in comparison to the previous research published by Dias et al. in [31]. This is the reason why the strategy proposed in this research is needed instead of applying the strategy described in [31] to each time series image. Therefore, the findings of this investigation complement those of earlier studies, such as [27,31].

The main contribution of this research has been to confirm the potential use of Kittler's taxonomy by machines to detect anomalies in the temporal context. Moreover, this study is the first research that compares the experiences of applying Kittler's taxonomy to the spatial and temporal contexts in remote sensing, to the knowledge of the authors. The relevance of this research is the support this study provides to improve machine learning capabilities, as we proposed and investigated a strategy that uses Kittler's taxonomy in order to bring a more organized and structured learning process to machines. Therefore, this research helps computers overcome their difficulties to recognize different contexts. Additionally, the proposed strategy is relevant because it allows computers to analyze large quantities of heterogeneous data in images from remote sensing.

The most important limitation of this research lies in the fact that this research showed that domain adaptation is needed when anomaly detection is applied to an image time series. Additionally, this research suggests that the method which must be used for domain adaptation depends on the environment that is being investigated. Therefore, the use of our strategy in automatic systems may be somewhat limited by the need for choosing a particular domain adaptation method for each environment. However, this is of little significance, since semi-automatic systems are able to detect the presence of brown mud where there are high levels of turbidity in the water.

Since the choice of the domain adaptation method is a challenge, future work should examine how a computer can select the most appropriate domain adaptation method automatically for each domain. Moreover, further research should focus on determining how many different domain adaptation methods are needed to detect anomalies in a huge set composed of an image time series acquired from a variety of environments.

Future work will investigate other types of anomalies defined by Kittler's taxonomy, such as *measurement model drift*, which still need to be examined in remote sensing. Future work will also develop and describe a single system able to use different strategies to detect the types of anomalies based on Kittler's taxonomy.

The strategy proposed in this study can be applied to improve the learning capabilities of machines, in order to allow computers to learn how to recognize temporal context to detect anomalies such as *component model drift*. Our strategy can also be applied to monitor river water pollution caused by environmental disasters related to brown mud. It is expected that our study will encourage other researchers to apply our anomaly detection strategy to conserve natural resources, in addition to other scientific areas.

Author Contributions: Conceptualization, M.A.D., G.C.M., R.G.N., W.C., I.B.M. and D.M.E.; Funding acquisition, M.A.D. and D.M.E.; Investigation, M.A.D. and G.C.M.; Methodology, M.A.D.; Resources, R.G.N., W.C., I.B.M. and D.M.E.; Validation, M.A.D.; Writing—original draft, M.A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by FOUNDATION FOR RESEARCH SUPPORT OF THE STATE OF SÃO PAULO (Fundação de Amparo à Pesquisa do Estado de São Paulo—FAPESP), grant numbers: 2020/06477-7, 2016/24185-8, 2021/01305-6, and 2021/03328-3. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES). CNPq (National Council for Scientific and Technological Development): 316228/2021-4.

Data Availability Statement: The Landsat 8 data set was provided by the United States Geological Survey (USGS): <https://earthexplorer.usgs.gov/>. The QGIS software was provided by the QGIS Development Team: <https://www.qgis.org/>. The Orfeo ToolBox was provided by the Open-source Geospatial and the OTB Communities: <https://www.orfeo-toolbox.org/>. Part of Figure 2 was provided by Google Earth.

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

Appendix A

Parameters

Some default parameters were used for training all classifiers: *zero* for default elevation, *1000* for maximum training sample size per class, *1000* for maximum validation sample size per class, *one* for bound sample number by minimum, *0.5* for training and validation sample ratio, *class* for the name of the discrimination field, and *zero* for set user defined seed. Other default parameters, which were used for training a specific classifier, are discriminated as follows. For the DT classifier: *65535* for maximum depth of the tree, *10* for minimum number of samples in each node, *0.01* for termination criteria for regression tree, *10* for cluster possible values of a categorical variable into $K \leq \text{cat}$ clusters to find a suboptimal split, *10* for K-fold cross-validations, set use *1seRule* flag to *false*, set *TruncatePrunedTree* flag to *false*, *dt* for classifier to use for training, and set *off* edge pixel inclusion. For the Boost classifier: *one* for maximum depth of the tree, *100* for weak count, *0.95* for weight trim rate, *real* for boost type, *boost* for classifier to use for training, and set *off* edge pixel inclusion. More detailed description of the TrainImagesClassifiers tool, as well as of each parameter used here, can be easily found in [52].

References

1. Nazeer, M.; Nichol, J.E. Combining Landsat TM/ETM+ and HJ-1 A/B CCD sensors for monitoring coastal water quality in Hong Kong. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1898–1902. [\[CrossRef\]](#)
2. Ha, N.T.T.; Koike, K.; Nhuan, M.T.; Canh, B.D.; Thao, N.T.P.; Parsons, M. Landsat 8/OLI two bands ratio algorithm for Chlorophyll-A concentration mapping in hypertrophic waters: An application to West Lake in Hanoi (Vietnam). *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4919–4929. [\[CrossRef\]](#)
3. Chen, J.; Zhu, W.-N.; Tian, Y.Q.; Yu, Q. Estimation of colored dissolved organic matter from Landsat-8 imagery for complex inland water: Case study of Lake Huron. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2201–2212. [\[CrossRef\]](#)
4. Kotchi, S.O.; Brazeau, S.; Turgeon, P.; Pelcat, Y.; Légaré, J.; Lavigne, M.-P.; Essono, F.N.; Fournier, R.A.; Michel, P. Evaluation of Earth observation systems for estimating environmental determinants of microbial contamination in recreational waters. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3730–3741. [\[CrossRef\]](#)
5. Chang, N.-B.; Vannah, B.; Yang, Y.J. Comparative sensor fusion between hyperspectral and multispectral satellite sensors for monitoring microcystin distribution in Lake Erie. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2426–2442. [\[CrossRef\]](#)
6. Li, Z.; Liu, H.; Luo, C.; Li, P.; Li, H.; Xiong, Z. Industrial wastewater discharge retrieval based on stable nighttime light imagery in China from 1992 to 2010. *Remote Sens.* **2014**, *6*, 7566–7579. [\[CrossRef\]](#)
7. Shoujing, Y.; Qiao, W.; Chuanqing, W.; Xiaoling, C.; Wandong, M.; Huiqin, M. A robust anomaly based change detection method for time series remote sensing images. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing Ltd.: Bristol, UK, 2014; Volume 17, p. 012059.
8. Zhou, Z.-G.; Tang, P.; Zhou, M. Detecting Anomaly Regions in Satellite Image Time Series Based on Seasonal Autocorrelation Analysis. In Proceedings of the ISPRS Annals of the Photogrammetry, Remote Sensing Spatial Information Science (XXIII ISPRS Congress), Prague, Czech Republic, 12–19 July 2016; pp. 303–310.
9. Chandola, V.; Vatsavai, R.R. A Gaussian Process Based Online Change Detection Algorithm For Monitoring Periodic Time Series. In Proceedings of the 2011 SIAM International Conference on Data Mining, Society for Industrial and Applied Mathematics, Mesa, AZ, USA, 28–30 April 2011; pp. 95–106. [\[CrossRef\]](#)
10. Committee on Developments in the Science of Learning; Committee on Learning Research and Educational Practice & National Research Council. *How People Learn: Brain, Mind, Experience, and School*, Expanded ed.; National Academy Press: Washington, DC, USA, 2000.
11. Mustard, J.F.; Sunshine, J.M. Spectral analysis for earth science: Investigations using remote sensing data. *Remote Sens. Earth Sci. Man. Remote Sens.* **1999**, *3*, 251–307.
12. Verstraete, M.M.; Pinty, B. Designing optimal spectral indexes for remote sensing applications. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 1254–1265. [\[CrossRef\]](#)
13. Brezonik, P.L.; Olmanson, L.G.; Finlay, J.C.; Bauer, M.E. Factors affecting the measurement of CDOM by remote sensing of optically complex inland waters. *Remote Sens. Environ.* **2015**, *157*, 199–215. [\[CrossRef\]](#)
14. Xie, X.; Li, B. A Unified Framework of Multiple Kernels Learning for Hyperspectral Remote Sensing Big Data. *J. Inf. Hiding Multimed. Signal. Process.* **2016**, *7*, 296–303.
15. Li, Y.; Li, J.; Pan, J.-S. Hyperspectral Image Recognition Using SVM Combined Deep Learning. *J. Internet Technol.* **2019**, *20*, 851–859.

16. Sublime, J.; Kalinicheva, E. Automatic post-disaster damage mapping using deep-learning techniques for change detection: Case study of the Tohoku tsunami. *Remote Sens.* **2019**, *11*, 1123. [[CrossRef](#)]
17. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A Survey on Deep Transfer Learning. In Proceedings of the 27th International Conference on Artificial Neural Networks, 4–7 October 2018, Rhodes, Greece; Springer: Cham, Switzerland, 2018; pp. 270–279.
18. Lu, J.; Behbood, V.; Hao, P.; Zuo, H.; Xue, S.; Zhang, G. Transfer learning using computational intelligence: A survey. *Knowl. Based Syst.* **2015**, *80*, 14–23. [[CrossRef](#)]
19. Shao, L.; Zhu, F.; Li, X. Transfer learning for visual categorization: A survey. *IEEE Trans. Neural Netw. Learn. Syst.* **2014**, *26*, 1019–1034. [[CrossRef](#)] [[PubMed](#)]
20. Cook, D.; Feuz, K.D.; Krishnan, N.C. Transfer learning for activity recognition: A survey. *Knowl. Inf. Syst.* **2013**, *36*, 537–556. [[CrossRef](#)]
21. Sun, S. A survey of multi-view machine learning. *Neural Comput. Appl.* **2013**, *23*, 2031–2038. [[CrossRef](#)]
22. Xu, Q.; Yang, Q. A survey of transfer and multitask learning in bioinformatics. *J. Comput. Sci. Eng.* **2011**, *5*, 257–268. [[CrossRef](#)]
23. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
24. Taylor, M.E.; Stone, P. Transfer learning for reinforcement learning domains: A survey. *J. Mach. Learn. Res.* **2009**, *10*, 1633–1685.
25. Bishop, C.M. *Pattern Recognition and Machine Learning*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2006.
26. Ma, L.; Crawford, M.M.; Tian, J. Local manifold learning-based k-Nearest-Neighbor for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 4099–4109. [[CrossRef](#)]
27. Kittler, J.; Christmas, W.; de Campos, T.; Windridge, D.; Yan, F.; Illingworth, J.; Osman, M. Domain anomaly detection in machine perception: A system architecture and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 845–859. [[CrossRef](#)]
28. Ponti, M.; Kittler, J.; Riva, M.; de Campos, T.; Zor, C. A decision cognizant Kullback–Leibler divergence. *Pattern Recognit.* **2017**, *61*, 470–478. [[CrossRef](#)]
29. Kittler, J.; Zor, C. Delta divergence: A novel decision cognizant measure of classifier incongruence. *IEEE Trans. Cybern.* **2018**, *99*, 1–13. [[CrossRef](#)] [[PubMed](#)]
30. Kittler, J.; Zor, C. A Measure of Surprise for Incongruence Detection. In Proceedings of the 2nd International Conference on Intelligent Signal Processing (ISP), London, UK, 1–2 December 2015; IET: London, UK, 2015; pp. 1–6. [[CrossRef](#)]
31. Dias, M.A.; Silva, E.A.d.; Azevedo, S.C.d.; Casaca, W.; Statella, T.; Negri, R.G. An incongruence-based anomaly detection strategy for analyzing water pollution in images from remote sensing. *Remote Sens.* **2020**, *12*, 43. [[CrossRef](#)]
32. Chandola, V.; Banerjee, A.; Kumar, V. Outlier detection: A survey. *ACM Comput. Surv. (CSUR)* **2007**, *14*, 1–83.
33. Gupta, M.; Gao, J.; Aggarwal, C.C.; Han, J. Outlier detection for temporal data: A survey. *IEEE Trans. Knowl. Data Eng.* **2013**, *26*, 2250–2267. [[CrossRef](#)]
34. Zimek, A.; Schubert, E.; Kriegel, H.P. A survey on unsupervised outlier detection in high-dimensional numerical data. *Stat. Anal. Data Min. ASA Data Sci. J.* **2012**, *5*, 363–387. [[CrossRef](#)]
35. Gogoi, P.; Bhattacharyya, D.K.; Borah, B.; Kalita, J.K. A survey of outlier detection methods in network anomaly identification. *Comput. J.* **2011**, *54*, 570–588. [[CrossRef](#)]
36. Niu, Z.; Shi, S.; Sun, J.; He, X. A Survey of Outlier Detection Methodologies and Their Applications. In Proceedings of the International Conference on Artificial Intelligence and Computational Intelligence, Taiyuan, China, 24–25 September 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 380–387.
37. Hodge, V.; Austin, J. A survey of outlier detection methodologies. *Artif. Intell. Rev.* **2004**, *22*, 85–126. [[CrossRef](#)]
38. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A survey. *ACM Comput. Surv. (CSUR)* **2009**, *41*, 1–72. [[CrossRef](#)]
39. Weinshall, D.; Zweig, A.; Hermansky, H.; Kombrink, S.; Ohl, F.W.; Anemuller, J.; Bach, J.-H.; Gool, L.V.; Nater, F.; Pajdla, T.; et al. Beyond novelty detection: Incongruent events, when general and specific classifiers disagree. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1886–1901. [[CrossRef](#)] [[PubMed](#)]
40. Chen, C.H. *Handbook of Pattern Recognition and Computer Vision*, 5th ed.; World Scientific: Singapore, 2015.
41. Jahne, B. *Computer Vision and Applications: A Guide for Students and Practitioners*, 1st ed.; Elsevier: London, UK, 2000.
42. Li, M.; Zang, S.; Zhang, B.; Li, S.; Wu, C. A review of remote sensing image classification techniques: The role of spatio-contextual information. *Eur. J. Remote Sens.* **2014**, *47*, 389–411. [[CrossRef](#)]
43. Schowengerdt, R.A. *Remote Sensing: Models and Methods for Image Processing*, 3rd ed.; Elsevier: London, UK, 2006.
44. Richards, J.A.; Jia, X. *Remote Sensing Digital Image Analysis*, 1st ed.; Springer: Berlin, Germany, 1999.
45. Asht, S.; Dass, R. Pattern recognition techniques: A review. *Int. J. Comput. Sci. Telecommun.* **2012**, *3*, 25–29.
46. Shen, L.; Li, C. Water Body Extraction from Landsat ETM+ Imagery Using Adaboost Algorithm. In Proceedings of the IEEE 2010 18th International Conference on Geoinformatics, Beijing, China, 18–20 June 2010. [[CrossRef](#)]
47. Blanzieri, E.; Melgani, F. Nearest Neighbor Classification of remote sensing images with the maximal margin principle. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1804–1811. [[CrossRef](#)]
48. Swain, P.H.; Hauska, H. The decision tree classifier: Design and potential. *IEEE Trans. Geosci. Electron.* **1977**, *15*, 142–147. [[CrossRef](#)]
49. Yi-Bin, L.; Ying-Ying, W.; Xue-Wen, R. Improvement of ID3 Algorithm Based on Simplified Information Entropy and Coordination Degree. In Proceedings of the Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017; IEEE: Toulouse, France, 2017; pp. 1526–1530. [[CrossRef](#)]

50. Earth Resources Observation and Science (EROS) Center. Landsat Data Dictionary. Available online: https://www.usgs.gov/centers/eros/science/landsat-data-dictionary#image_quality_landsat_8 (accessed on 30 March 2022).
51. Tan, P.-N.; Steinbach, M.; Karpatne, A.; Kumar, V. *Introduction to Data Mining*, 2nd ed.; Pearson Education: Noida, India, 2018.
52. OTB Development Team. The Orfeo ToolBox Cookbook, a Guide for Non-Developers Updated for OTB-5.6.0. 2011. Available online: <https://www.orfeo-toolbox.org/packages/archives/Doc/CookBook-5.6.0.pdf> (accessed on 30 March 2022).
53. Tuia, D.; Persello, C.; Bruzzone, L. Domain adaptation for the classification of remote sensing data: An overview of recent advances. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 41–57. [[CrossRef](#)]
54. Gretton, A.; Borgwardt, K.M.; Rasch, M.J.; Schölkopf, B.; Smola, A. A kernel two-sample test. *J. Mach. Learn. Res.* **2012**, *13*, 723–773.
55. Liu, Q.; Klucik, R.; Chen, C.; Grant, G.; Gallaher, D.; Lv, Q.; Shang, L. Unsupervised detection of contextual anomaly in remotely sensed data. *Remote Sens. Environ.* **2017**, *202*, 75–87. [[CrossRef](#)]
56. Bhaduri, K.; Das, K.; Votava, P. Distributed Anomaly Detection Using Satellite Data from Multiple Modalities. In Proceedings of the 2010 Conference on Intelligent Data Understanding, CIDU 2010, Mountain View, CA, USA, 5–6 October 2010.
57. Bormann, K.J.; McCabe, M.F.; Evans, J.P. Satellite based observations for seasonal snow cover detection and characterization in Australia. *Remote Sens. Environ.* **2012**, *123*, 57–71. [[CrossRef](#)]
58. Che, X.; Feng, M.; Sexton, J.; Channan, S.; Sun, Q.; Ying, Q.; Liu, J.; Wang, Y. Landsat-based estimation of seasonal water cover and change in arid and semi-arid Central Asia (2000–2015). *Remote Sens.* **2019**, *11*, 1323. [[CrossRef](#)]
59. Chen, Y.; Fan, R.; Yang, X.; Wang, J.; Latif, A. Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning. *Water* **2018**, *10*, 585. [[CrossRef](#)]
60. Natesan, S.; Armenakis, C.; Benari, G.; Lee, R. Use of UAV-borne spectrometer for land cover classification. *Drones* **2018**, *2*, 16. [[CrossRef](#)]
61. Yang, K.; Li, M.; Liu, Y.; Cheng, L.; Huang, Q.; Chen, Y. River detection in remotely sensed imagery using Gabor Filtering and path opening. *Remote Sens.* **2015**, *7*, 8779–8802. [[CrossRef](#)]
62. Fernandes, G.W.; Goulart, F.F.; Ranieri, B.D.; Coelho, M.S.; Dales, K.; Boesche, N.; Bustamante, M.; Carvalho, F.A.; Carvalho, D.C.; Dirzo, R.; et al. Deep into the mud: Ecological and socio-economic impacts of the dam breach in Mariana, Brazil. *Braz. J. Nat. Conserv.* **2016**, *14*, 35–45. [[CrossRef](#)]
63. Mielke, C.; Boesche, N.K.; Rogass, C.; Kaufmann, H.; Gauert, C.; de Wit, M. Spaceborne mine waste mineralogy monitoring in South Africa, applications for modern push-broom missions: Hyperion/OLI and EnMAP/Sentinel-2. *Remote Sens.* **2014**, *6*, 6790–6816. [[CrossRef](#)]
64. Rosell-Melé, A.; Moraleda-Cibrián, N.; Cartró-Sabaté, M.; Colomer-Ventura, F.; Mayor, P.; Orta-Martínez, M. Oil pollution in soils and sediments from the Northern Peruvian Amazon. *Sci. Total Environ.* **2018**, *610*, 1010–1019. [[CrossRef](#)]
65. QGIS Development Team. Available online: <https://www.qgis.org/> (accessed on 25 April 2018).
66. USGS—The United States Geological Survey, “Earth Explorer”. Available online: <https://earthexplorer.usgs.gov/> (accessed on 25 April 2018).
67. Gonzales, R.C.; Woods, R.E. *Digital Image Processing*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2002.
68. QGIS Project. QGIS User Guide Release 2.18. 2019. Available online: <https://docs.qgis.org/2.18/pdf/en/QGIS-2.18-UserGuide-en.pdf> (accessed on 30 March 2022).
69. Vivone, G.; Alparone, L.; Chanussot, J.; Mura, M.D.; Garzelli, A.; Licciardi, G.A.; Restaino, R.; Wald, L. A critical comparison among pansharpening algorithms. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2565–2586. [[CrossRef](#)]
70. Reinhard, E.; Shirley, P.; Ashikhmin, M.; Troschianko, T. Second Order Image Statistics in Computer Graphics. In Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization (APGV'04), Los Angeles, CA, USA, 7–8 August 2004; pp. 99–106.
71. OTB Development Team. The On-Line Orfeo ToolBox Cookbook, a Guide for Non-Developers Updated for OTB-3.10. 2011. Available online: https://www.orfeo-toolbox.org/packages/doc/tests-rfc-52/cookbook-3b41671/Applications/app_TrainImagesClassifier.html (accessed on 25 April 2018).
72. Chen, C.; Yang, B.; Song, S.; Peng, X.; Huang, R. Automatic clearance anomaly detection for transmission line corridors utilizing UAV-Borne LIDAR Data. *Remote Sens.* **2018**, *10*, 613. [[CrossRef](#)]
73. Congalton, R.G.; Mead, R.A. A review of three discrete multivariate analysis techniques used in assessing the accuracy of remotely sensed data from error matrices. *IEEE Trans. Geosci. Remote Sens.* **1986**, *GE-24*, 169–174. [[CrossRef](#)]
74. Marzano, F.S.; Scaranari, D.; Montopoli, M.; Vulpiani, G. Supervised classification and estimation of hydrometeors from C-Band dual-polarized radars: A Bayesian approach. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 85–98. [[CrossRef](#)]
75. Indu, J.; Kumar, D.N. Evaluation of precipitation retrievals from orbital data products of TRMM over a subtropical basin in India. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6429–6442. [[CrossRef](#)]
76. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*; CRC Press: Boca Raton, FL, USA, 2019.
77. Bernardo, N.; do Carmo, A.; Park, E.; Alcântara, E. Retrieval of suspended particulate matter in inland waters with widely differing optical properties using a semi-analytical scheme. *Remote Sens.* **2019**, *11*, 2283. [[CrossRef](#)]
78. Li, W.; Du, Q. A survey on representation-based classification and detection in hyperspectral remote sensing imagery. *Pattern Recognit. Lett.* **2016**, *83*, 115–123. [[CrossRef](#)]

79. Zhang, J. Advancements of outlier detection: A survey. *ICST Trans. Scalable Inf. Syst.* **2013**, *13*, 1–26. [[CrossRef](#)]
80. Frontera-Pons, J.; Veganzones, M.A.; Pascal, F.; Ovarlez, J.P. Hyperspectral anomaly detectors using robust estimators. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *9*, 720–731. [[CrossRef](#)]
81. Matteoli, S.; Diani, M.; Theiler, J. An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2317–2336. [[CrossRef](#)]
82. Matteoli, S.; Veracini, T.; Diani, M.; Corsini, G. Models and methods for automated background density estimation in hyperspectral anomaly detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *51*, 2837–2852. [[CrossRef](#)]
83. Matteoli, S.; Diani, M.; Corsini, G. A tutorial overview of anomaly detection in hyperspectral images. *IEEE Aerosp. Electron. Syst. Mag.* **2010**, *25*, 5–28. [[CrossRef](#)]