Machine Learning Methods for Forest Image Analysis and Classification: A Survey of the State of the Art

CLOPAS KWENDA[ID]**, MANDLENKOSI GWETU, AND JEAN VINCENT FONOU DOMBEU**[ID]
School of Computer Science, Statistics and Mathematics, University of KwaZulu-Natal, Pietermaritzburg 3209, South Africa

Corresponding author: Clopas Kwenda (221072651@stu.ukzn.ac.za)

ABSTRACT The advent of modern remote sensors alongside the development of advanced parallel computing has significantly transformed both the theoretical and real implementation aspects of remote sensing. Several algorithms for detecting objects of interest in remote sensing images and subsequent classification have been devised, and these include template matching based methods, machine learning and knowledge-based methods. Knowledge-driven approaches have received much attention from the remote sensing fraternity. They do, however, face challenges in terms of sensory gap, duality of expression, vagueness and ambiguity, geographic concepts expressed in multiple modes, and semantic gap. This paper aims to review and provide an up-to-date survey on machine learning and knowledge driven approaches towards remote sensing forest image analysis. It is envisaged that this work will assist researchers in coming up with efficient models that accurately detect and classify forest images. There is a mismatch between what domain experts expect from remote sensing data and what remote sensing science produces. Such a mismatch or disparity can be reduced or alleviated by adopting an ontology paradigm methodology. Ontologies should be used to support the future of remote sensing in forest object classification. The paper is presented in five parts: (1) a review of methods used for forest image detection and classification; (2) challenges faced by object detection methods; (3) analysis of segmentation techniques employed; (4) feature extraction and classification; and (5) performance of the state-of-the-art methods employed in forest image detection and classification.

INDEX TERMS Feature extraction, ontology, segmentation, remote sensing.

I. INTRODUCTION

Remote sensing science is rapidly growing. The evolution of high spatial resolution remote sensors in conjunction with advanced computing has significantly transformed the specification and practice of remote sensing [1]. Remote sensing images are characterized by high spatial resolution and provide more explicit information on the earth's surface as compared to middle and coarser resolution images [2]. Machine learning methods for analyzing and classifying forest images are continuously evolving to provide more advanced automatic land cover pattern recognition on aerial images. This paper surveyed existing methods for forest ecosystem image classification. In particular, machine learning classifiers and

deep learning techniques for forest image classification are reviewed.

There are several algorithms that are geared towards detecting objects of interest in remote sensing images, for the further regional analysis and classification. These algorithms are categorized into three groups, namely template matching based, knowledge based, and machine learning based methods [3]. The taxonomy of image classification methods is depicted in Figure 1.

(a) Template matching based detection methods

The template matching method determines whether a picture or an image contains a previously defined object or whether a predefined sub image (template) has an exact match in an image. Although this method provided one of the first approaches for object analysis [3], its dependence on handcrafted matching criteria limited its applicability to complex object recognition. Once a suitable template is deter-
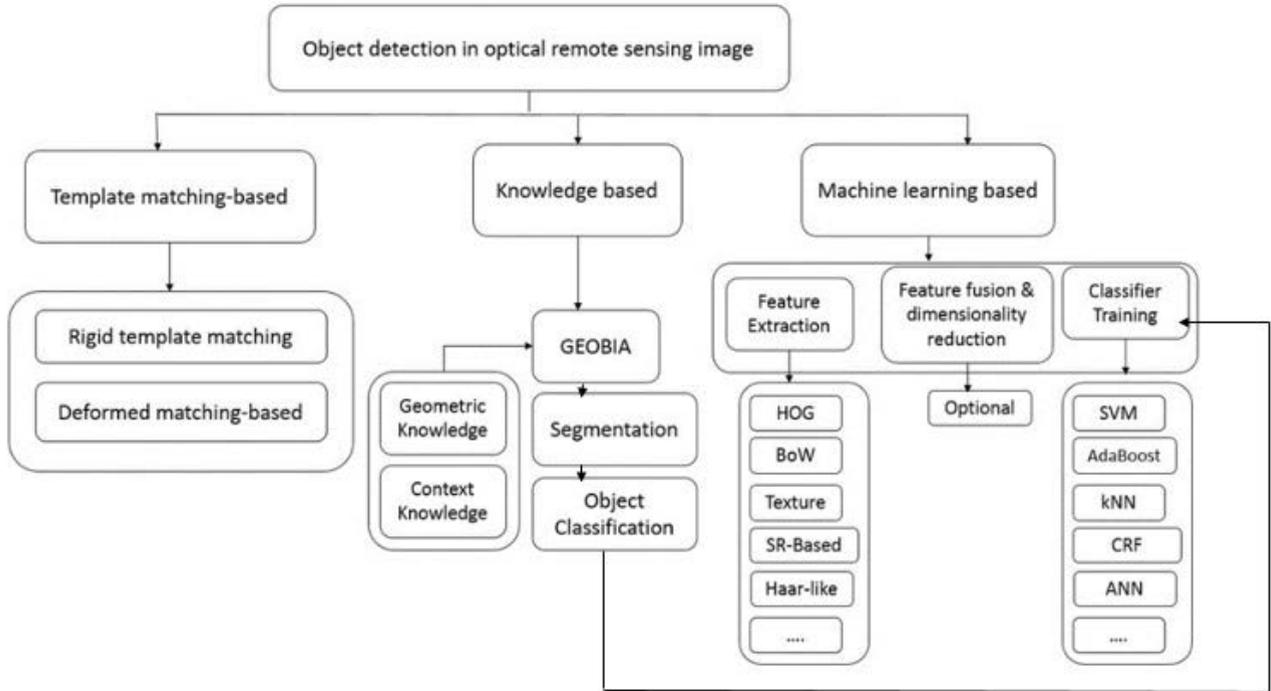
FIGURE 1. Methods for object detection in optical sensing images [3].

mined, a measure of matching between the template and every possible location in the image is calculated, and a classification decision is made based on the measure of certainty. The most popular metric based measures are the Euclidean distance, squared difference, and cross correlation, defined in Equations (1) to (3):

Euclidean distance

$$E(m, n) = \sqrt{\sum_i \sum_j [g(i,j) - t(i-m, j-n)]^2 24} \quad (1)$$

Squared Difference Measure

$$\begin{aligned} E^2(m, n) \\ = \sum_i \sum_j [g(i,j)^2 - 2g(i,j)t(i-m,j-n)) + t(i-m,j-n)^2] \end{aligned} \quad (2)$$

Cross Correlation Measure

$$R(m, n) = \sum_i \sum_j g(i,j) + t(i,j)\, t(i-m, j-n) \quad (3)$$

There are two types of templates, namely, global and local templates. When a template is used to reference the whole (global) object in an image, it is referred to as a global template. However, when object features (local features of an object) in an image are referenced with multiple or several templates, these templates are referred to as local templates [4]. Figure 2 shows stages to be followed to determine the best templates for object detection. The challenge
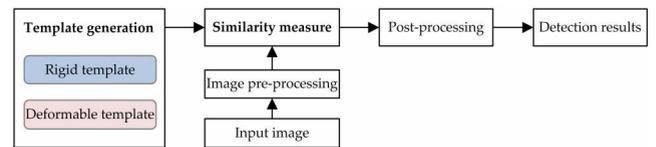

FIGURE 2. Template matching based criteria [3].

with this approach is that the method does not cater for the scale and orientation of the template [5]. It fails due to occlusions and distortions on the boundary [6]. The method is very sensitive to shape and viewpoint change. The solution suggested was to have a unique representation of a template orientation and scale that varies, but the solution becomes computationally expensive.

(b)   Machine Learning based approaches

An input image is subjected to the initial first phase where regions or objects are extracted. Then, for each object, features of interest are computed using Convolutional Neural Networks (CNN). Optimal features are obtained after a subsequent series of feature fusion and dimension reduction processes. Finally, classifiers such as Support Vector Machines (SVM), k-Nearest Neighbor (kNN), Sparse Representation based Classification (SRC), AdaBoost, Conditional Random Fields (CRF), and others are used to classify each region/object. Figure 3 shows the main important phases of machine learning object detection, i.e., feature extraction, feature fusion, and dimensionality reduction, as well as
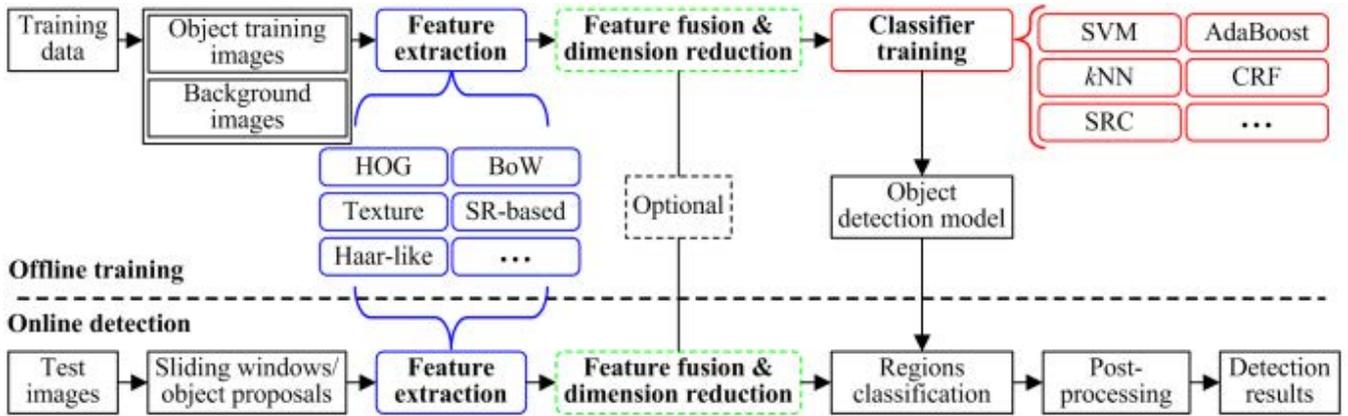
FIGURE 3. Machine learning methods [3].

the classification phase. Machine learning-based approaches coupled with innovative algorithms and higher performance computing seem to have gained popularity in remote sensing science because they produce better results considering the accuracy of the created maps [1]. As a result, they are used in big land cover applications that rely on pixel based statistical analysis of massive image data sets [7]–[9]. Pixel based approaches pose challenges in the analysis of high spatial resolution images [10] because they take into consideration the aspects of spectral information as a backdrop for analyzing and classifying high spatial resolution images, neglecting spatial and temporal information, which are of paramount importance. These methods are less efficient in dealing with symbolic knowledge, that is, when concepts are characterized by symbols, for instance, vegetation is made of grass [11]. They do not offer the function of creating a super class whilst classes have been defined. Suppose one has defined the following classes of interest; "trees", "grass", "road", and "building". It will then be impossible for the user to define a vegetation class unless it has been beforehand defined as a super class. The methods do not offer the facility to add spatial rules [12], for instance, "grass" cannot be found inside a building, but can be found in a field. Because of this reasoning limitation, data-driven approaches are unsuitable for use in applications areas such as ecology, that deal with earth observation.

(c) Knowledge based detection methods

These methods have been applied to land slide, crops, urban land change and forests [13]–[16]. Figure 4 shows the processes whereby an input image goes through a hypothesis generation phase, the hypothesis is validated and tested using the established knowledge and rules. Post processing of validation results will be subjected to machine learning for final object detection. Knowledge and rules from geometric information and context information will be used to test the validity of the hypothesis generated from an input image, if the hypothesis is valid it will be subjected to machine learning for object detection [17]. Generally, there are two
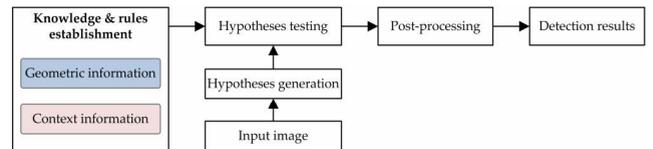

FIGURE 4. Knowledge based object detection systems [3].

types of knowledge that have been used on target objects, and these are geometric knowledge and context knowledge.

(a) Geometric Knowledge

This type of knowledge is the most important in a knowledge-based approach and is widely used for object detection. It encompasses generic shape models or parametric specifics. For instance, it is proposed in [18] that buildings are square or composed of rectangular segments and are utilized as conventional models of shapes to distinguish buildings.

(b) Context knowledge

Context knowledge is very important for key objects, and it is expressed by rules derived from relationships between objects of interest and their respective backgrounds [14], [19], [20]. For instance, shadow evidence has been used for building detection [21], the correlation between artificial structures such as buildings and their respective shadows has been used to project locations and shapes of buildings [22].

Recently, knowledge-driven approaches seem to be the direction taken by the remote sensing science community [3] since they incorporate domain expert knowledge. Geographic Object Based Image Analysis (GEOBIA), which classifies image objects based on apriori domain expert knowledge, is proving to be a key trend in remote sensing image analysis. GEOBIA is a classification technique that divides a remote sensing image into objects of interest and evaluates the objects based on their spectral, temporal and spatial characteristics. The generation of objects of interest is done using different segmentation approaches such as random walker, canny, histogram-based segmentation, etc. An algorithm is deemed effective in segmentation if and only if a segmented

image object completely matches the corresponding Actual Image Region (AIR) of a scene object. [2], proposed a blend of area coincidence methods and boundary coincidence methods for assessing segmentation quality. The area coincidence methods select an image that has the dominant or largest area of intersection with the AIR. The boundary coincidence methods calculate the distance between a point of interest in a segmented image and that of its corresponding point on the AIR. The segmentation quality is high when the measured distance is much closer to zero. Segmentation evaluation methods can either be Unsupervised or Supervised. Supervised techniques evaluate a segmented image based on a ground truth image also referred to as the reference image. The evaluation of unsupervised methods is solely dependant on the segmented image as it has to assess the extent to which the image matches the desirable features of a good segmented image. [23] proposed four metrics (Equations 4-7) for assessing supervised segmentation quality namely F-measure, SUM which should be less than 2, ED that indicates distance to point (0,0) in the space and ED' that indicates distance to point (1,1) in the space.

$$f = \frac{1}{\alpha \frac{1}{precision} + 1 - \alpha \frac{1}{recall}} \quad (4)$$

$$sum = precision + recall \quad (5)$$

$$ed = \sqrt{precision^2 + recall^2} \quad (6)$$

$$ed' = \sqrt{(1 - precision)^2 (1 - recall)^2} \quad (7)$$

Two other metrics that take into account the over and under segmentation errors, GOSE and GUSE, respectively, are proposed [24]. Rand Error(RE) is another widely used metric for evaluating supervised approaches. RE is a measure of is defined in Equation 8 [25]. Let R_1 and R_2 be segmentation regions of image S with t pixels and the following holds:

- n correspond to the number of pixels in image S that appear in both R_1 and R_2
- m correspond to the number of pixels in image S that are neither in R_1 and R_2

$$RE = \frac{n + m}{\binom{n}{2}} \quad (8)$$

A criterion for unsupervised technique that balances homogeneity and inter-segment heterogeneity is proposed by Wang *et al.* [26] as in Equation 9.

$$Z = T + \lambda D \quad (9)$$

where, T and D represents intra-segment homogeneity and inter-segment heterogeneity, respectively. Another metric for unsupervised techniques proposed by Gao *et al.* [27] is the Global Score (GS). GS incorporates weighted variance (WV) and Moron's I and is defined in Equation 10.
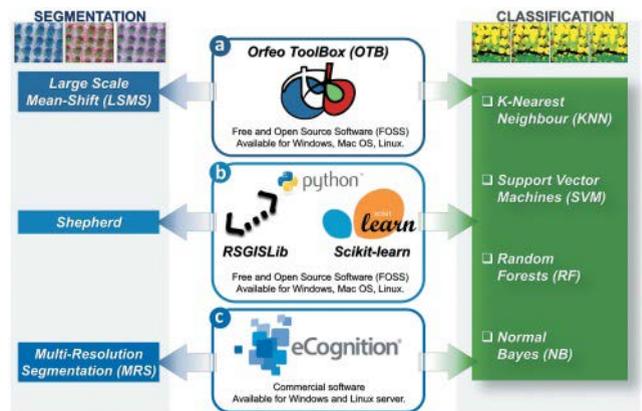
$$GS = V_{norm} + I_{norm} \quad (10)$$

FIGURE 5. GEOBIA WORKFLOW [28].

The final step of GEOBIA is image classification. The common image classifiers for GEOBIA are Random Forest (RF), Simple Vector Machines (SVM), k-Nearest Neighbor (kNN) and Naive Bayes (NB) [28].

Figure 5 shows GEOBIA workflow [28] that implemented three different algorithms, namely, Large Scale Mean Shift (LSMS) in OTB, the Shephered segmentation algorithm in RSGISLib and the Multi-resolution segmentation (MRS) algorithm in eCognition. However, GEOBIA solutions do not give answers to every segmentation problem. Even though GEOBIA is more efficient than pixel-based approaches, segmenting a multi-spectral image made up of thousands of mega pixels remains a challenging task [29]. Another drawback of GEOBIA is that it approximates, to some extent, computer-aided photo interpretation, which has been criticized as being highly subjective [5]. However, in the last decade, knowledge-driven techniques, like GEOBIA, have gained traction as a means of bridging the gap between implicit data representation and end-user needs. Knowledge-driven approaches consist of translating symbolic knowledge into a format understandable by humans into numerical knowledge.

Vegetation indices obtained from satellite images provide valuable information which is essential for the mapping of vegetation. The Normalised Difference Vegetation Index (NDVI) has proven to be a valuable tool, particularly in tropical dry forests, where it serves as a foundation for estimating overall green biomass, tree density, and species diversity [30]–[32]. NDVI is an indicator that determines the greenish component from the analyzed satellite images. NDVI provides a balance between the energy received and the energy emitted by objects on the earth's surface [32]. In the context of plant communities, it is an indicator that determines how greenish an area is, and that is influenced by the quantity of vegetation in that particular area and its state of health. The NDVI values range from -1 to $+1$. The values that are less than 0.1 correspond to water bodies and bare grounds, while higher values indicate the presence of agricultural activities, temperate forests, and rain
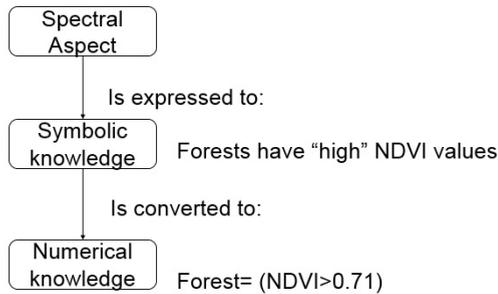
FIGURE 6. Symbolic to numerical knowledge conversion.

forests [32]. The NDVI values can be used to group the vegetation ecosystems into 4 major categories as follows [33]: forests made up of semi-deciduous and evergreen have NDVI $\geq$ 0.7, woodlands are defined by the range ($0.6 \leq$ NDVI < 0.71), a mixed class that is composed of a) shrub land, b) woodland/shrub land/exposed lands, and c) cactus forest have the range ($0.49 \leq$ NDVI < 0.61) and the dwarf woodland and shrub land assume the range of (NDVI < 0.49) [33]. For instance, a forest concept made up of semi-deciduous and evergreen is described by high NDVI values and when translated into numerical knowledge, it is implemented by the classification rule set: Forest = (NDVI $\geq$ 0.71). Figure 6 shows the symbolic to numerical knowledge conversion.

II. OVERVIEW OF THE CHALLENGES FACED WITH KNOWLEDGE BASED APPROACHES

A. DIFFERENT MODES OF DEFINING GEOGRAPHIC CONCEPTS

A geographical concept can be defined from different perspectives; the definition might be based on physical, historical, functional, or conventional mode [34]. Various methods of defining the same geographic concept bring about elective perspectives on the definition of the same concepts; for example, an idea can be characterized by elective definitions that are not basically the same, despite the fact that they are normally correlated [35]. From a functional viewpoint, the role of the forest primarily acts as a repository for storing carbon. This is correlated by the Net Primary Productivity (NPP) values. Forests can also be defined based on physical attributes such as vegetation cover, phenology, vegetation, age, etc. A tremendous effort is still in place to standardize land cover classes in land cover classification systems (LCCS) [36]. The term "forest" is defined differently by different organizations and countries; for instance, in Brazil an area that is regarded as a forest, has an area that exceeds one hectare, is characterised by a 30% canopy, and is composed of trees with a minimum height of 5m [37]. A forest in China is defined as an area larger than 0.67 hectares in size, with at least 20% crown cover and trees standing at least

2 meters tall. The Food and Agriculture Organization (FAO) standardised the definition of forest to refer to a land area spanning over 0.5 hectares enveloped by trees at 5m and above, with a canopy cover of 10%. This definition excludes land under agricultural or urban land use [38].

B. DUALITY OF GEOGRAPHIC CONCEPTS

Two important major terms arise from the concept of duality, that is, scene and image. A scene is real and exists on the ground, whereas an image is an assortment of spatially orchestrated estimations drawn from the scene [1]. Components obtained from images are regarded as abstractions of real objects in the ground scene [39]. Forest concepts can be viewed either from a real world perspective (a forest concept is characterized by high NPP values) or from image properties (a forest concept is defined by high NDVI values). In the case of forests, the assertion that NDVI is correlated to NPP is not always valid because NPP lacks information on attributes such as vegetation height, vegetation cover, and so forth. This anomaly is also referred to as the sensory gap. With this notion, sensory gaps cause improbability in describing geographic objects [40].

C. VAGUENESS AND AMBIGUITY OF GEOGRAPHIC CONCEPTS

The process of connecting attribute (e.g. NDVI) range of values (for instance, high) to geographic concepts (e.g. forest concept) is not easy [1]. The reason behind this is that the associated value "high" is qualitative in nature, so the obtained classification rule becomes vague and ambiguous. Some pixels (image objects) inside the image in Figure 7 are not classified as forest, though in nature they are constituted as forest areas. When qualitative terms like "high" are employed to identify objects with sharp, crisp boundaries, threshold ambiguity occurs. Qualitative description of geographic objects raises partiality issues. For instance, the symbolic classification rule "high NDVI" partially describes the forest because:

(a) It is very difficult to fish out only forests in areas that have other crops with the same NDVI values as in Figure 9.

(b) It is also impossible to classify all the varying types of forests because, in some cases, there are some forests that have "low NDVI" values, such as the degraded forests in Figure 8.

Ambiguity arises in all cases where a natural language expression can have various meanings [41]. The usual one is lexical ambiguity, which emerges because of the homonym of regular language articulation, that is, an expression with more than one meaning, such that each meaning points only to one ontological concept unambiguously [41]. More than 800 different definitions of forest concepts are provided in [42]. Deep ambiguity, also referred to as open texture, exists where there is no clear boundary between concepts or terms or cases where the meaning of a concept changes over time, for
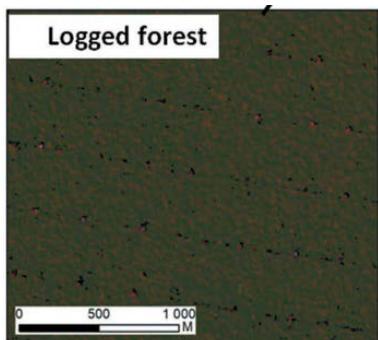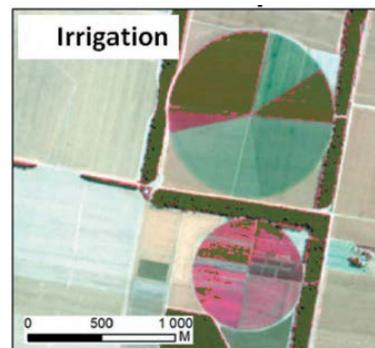
FIGURE 7. Concepts of vagueness.
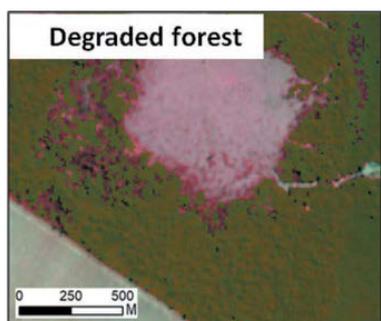

FIGURE 8. Forests having low NDVI values.

FIGURE 9. Other crops having high NDVI as forests.

FIGURE 10. Definition of knowledge [45].

instance, when a new technology appears or the physical or social context of the term evolves.

D. SEMANTIC GAP

It arises from the vagueness and ambiguity of geographic concepts. It is defined as a mismatch between data extracted on the basis of visual information and the interpretation drawn from the same data in a given situation [40]. This is so because converting visual data (from human perception) to computational representation is a very difficult task. The translation first requires expressing perception of visual data into a symbolic knowledge representation format (e.g. forests have high NDVI values). Such a conversion is a very difficult task since some concepts have vague meanings when expressed in natural language [43]. For instance, color may be considered a significantly important biophysical property [44], but its perception varies amongst humans and it is difficult to express.

III. INTRODUCTION ON ONTOLOGIES

Sharing knowledge among people is feasible only if people speak a common language [42]. The traditional definition of "knowledge is a subset of true beliefs" [45]. It is the intersection between truth and beliefs, as represented in Figure 10. Ontologies enable formal (machine-understandable) representation of knowledge. In computer science, ontology is defined as an explicit, formal specification of a shared conceptualization [46]. An ontology is a systematic description of existence, and this term is drawn from philosophy. What
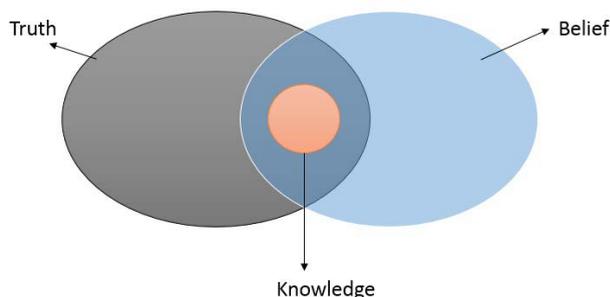
"exists" for Artificial Intelligence (AI) systems is that which can be represented. The following properties, with corresponding definitions, should be observed: (1) conceptualization, means that an ontology is an abstract model of a real world phenomenon; (2) explicit, implies that all ontology concepts must be clearly defined; (3) formal, implies that an ontology is machine understandable; and (4) shared, means that there should be consensus amongst a community of people about the knowledge represented by the ontology.

A. FORMAL ONTOLOGIES

Remote sensing science experts are conversant with working on numerical knowledge that has been derived from an image viewpoint [47]. Numerical knowledge representation by nature suffers from the problem of partiality and implicit knowledge representation, hence it becomes difficult to share the knowledge with other scientists, such as ecologist, agronomist, etc., who are used to working with symbolic knowledge in describing a geographic concept, for instance, a forest concept is defined by "High NDVI" values. Formal ontologies provide a road map that caters for the representation of both symbolic and numerical knowledge. Formal ontologies can be utilized to unequivocally portray a perception or observation from different perspectives, for instance, the extensible observation ontology (OBOE) is utilized to portray the semantics of scientific observations. An observation of an entity encompasses the characteristics (e.g. biomass) of the entity based on a measurement standard
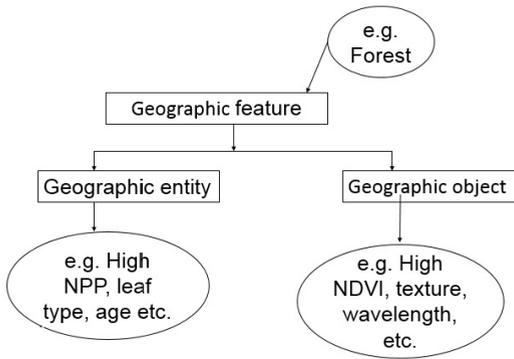
FIGURE 11. Dual representation of concepts.

FIGURE 12. Solving sensory gap challenge.

(grams). Ontologies for remote sensing science applications based on description logic offer the following advantages;

- Symbolic language - it binds/associates concepts with relevant sensing data and also promotes binding of related concepts.
- Knowledge sharing - it advocates for common conceptualization and adoption of standard ontology language such as web ontology language.
- Reasoning - description logic in ontology allows the inferring of new knowledge from explicit descriptions.

B. ONTOLOGY KNOWLEDGE BASED AS A SOLUTION

This section outlines how the adoption of ontologies in knowledge base approaches helps in alleviating the problems addressed in Section 2.0.

1) DUALITY OF GEOGRAPHIC CONCEPTS

Ontologies incorporate the concept of perspectivalism. That is, they allow the separate description of a field point of view of a forest concept. For instance, a forest concept can be specified in terms of attributes such as ''high'' NPP, leaf type, and so on. The other angle of description is from the point of view of an image of geographic objects. That is, a forest can be defined in terms of attributes such as ''high'' NDVI, texture, and wavelength. In general, it allows for the separate description of geographic entities and geographic objects alongside their characteristics. Figure 11 shows the dual representation of a geographic feature, that is, it can be described either from the perspective of a geographic entity or from the perspective of a geographic object.

2) VAGUENESS AND AMBIGUITY OF GEOGRAPHIC CONCEPTS

Fuzzy logic is the most popular way of handling the vagueness of geographic concepts [48]. Processing of data is done using partial set membership rather than strict set membership. For example, a forest concept is not considered to be strictly ''green'', but rather is considered to belong partially to some degree to the set of things that are green. [49] defined
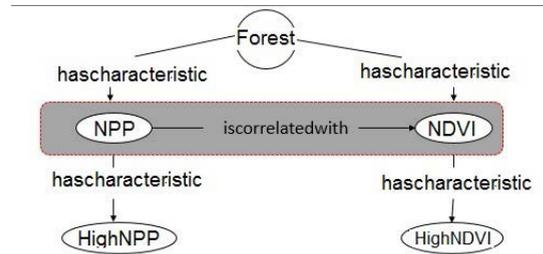
two thresholds i.e. ambiguity reject threshold and the distance reject threshold. Ambiguity reject threshold is defined by the rule $\alpha_{amp} \in | 0.5, 1 |$ and define the degree of confidence required to recognise an object. Distance reject threshold is defined by the rule $\alpha_{dist} \in | 0.1 |$, this means an object x is unlikely to belong to both classes C_k and $\neg$ C_k and might belong to a concept not yet learnt. Vagueness can also be addressed by adopting probability ontologies. They use probability sets to define concepts of interest. Attributes in the set properties have probabilities attached to them, and the statistical measure of the probability value of the geographic concept [50] is used to determine whether a geographic concept is a member of a class. Ambiguity in ontologies can be reduced by limiting the information that describes a concept [41].

3) SENSORY GAP

The discrepancy between real objects and their depiction in images is known as the sensory gap. As referenced by [40] sensory gap can be reduced by explicitly defining the domain and world knowledge in the system. Knowledge about physical laws, laws governing the behavior of objects, and how people perceive them will all be incorporated into the system in the hope of enhancing recognisers and thereby assisting machines in bridging the sensory gap [51]. However, in ontologies, real world description of forest entities is correlated with matching image point description of forest objects, i.e., NDVI is correlated with NPP [2]. Figure 12 shows how a real world description of a forest concept can be mapped to an object description in an image. An object description of an image is easily formalised on a computer.

4) SEMANTIC GAP

The semantic gap is the discrepancy between the high level descriptions of images by humans and the low-level detection used by machines to detect images [52]. On the other hand, adding captions and annotations to images solves the problem [50]. The method is time-consuming and costly because it requires a lot of effort, machine algorithm tweaking, and close attention to vocabulary and content to ensure that photos are appropriately labeled [50]. In ontologies, however, an image feature (e.g. NDVI) and its associated value (''HIGHNDVI'') are used to define a pixel (image object) of a forest concept. The ''HIGHNDVI'' concept is formalized as a result
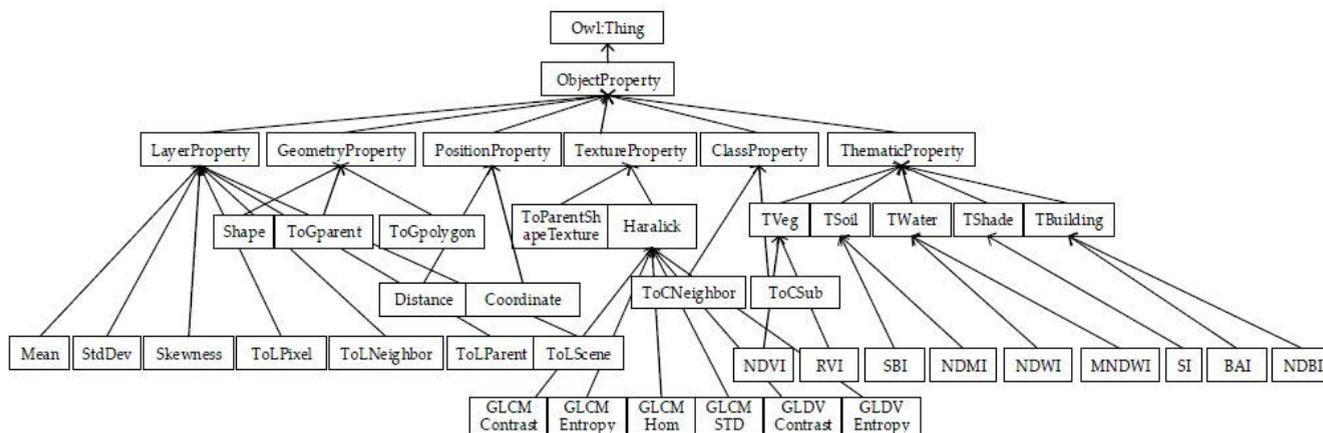
FIGURE 13. Object features hierarchy in ontology [55].

TABLE 1. Ontological framework for RSI concepts.

Item	Quantity	Image
Associated_to	Region	Image
from_band	Image	-
from_satellite	Sensor	Satellite
from_sensor	-	Sensor
has_spatial_resolution	-	spatial resolution
has_spectral_resolution	-	spectral resolution

of the established relationship between symbolic information (e.g. "HIGHNDVI") and numerical knowledge (e.g., NDVI>0.7), hence the semantic gap is reduced.

IV. ONTOLOGICAL FRAMEWORK FOR RSI (REMOTE SENSING IMAGE)

[53] proposed a novel framework for RSI. The framework is made up of important terms or concepts. These include satellite, sensor, image, spatial resolution, and spectral resolution. The elements are shown in the table 1. Slot is mainly concerned with the spatial and spectral resolutions, which relate to the scope, although there are no related elements in the range component. Spectral resolution is one of the most important concepts for the framework. It follows a top down approach method, where the concept is parceled into two sub-components, i.e. the visible part and the infrared part. The visible is made up of three color segments, i.e. the RGB (red, blue, and green). The infrared part is also made up of three segments, i.e. thermal infrared, near infrared, and far infrared. The parameters suited for the slot are explicitly defined and include has_spatial_resolution, has_spectral_resolution, etc. [47] developed a simple ontological approach for remote sensing image classification. The prototype was built upon the expert remote sensing knowledge expressed in [54].

A. ONTOLOGICAL FRAMEWORK FOR OBJECT FEATURE EXTRACTION

After an image goes through a segmentation process, each region is characterised by a set of features. The feature extraction process from eCognition software follows the general

upper ontology defined using the top down method [55]. The features are divided into six categories, namely LayerProperty, GeometryProperty, PositionProperty, TextureProperty, ClassProperty, and ThematicProperty. The selection of features of interest is performed by an expert to allow object detection. Figure 13 shows a hierarchical breakdown of object features from the six categories. GeometryProperty, TextureProperty, and ThematicProperty are important features in detecting forest objects [56].

B. ONTOLOGY MODEL OF THE LAND COVER CLASS HIERARCHY

The upper-level ontology is developed using concepts from land cover classification systems (LCCS) [36]. Figure 13 shows a hierarchically simplified way of representing classes of interest emanating from the main land cover class [53]. [55] designed an upper level ontology for the Chinese Geographic Condition Census Project [57]. Figure 14 depicts the design of an eight land cover ontology. The procedure was as follows:

1) The first step was to establish a set of important terms, in this case; Fields, Woodland, Grassland, Orchards, Bare land, Roads, Building and Water.
2) Classes and class hierarchies were then defined, A land cover class was defined through a top down approach.

1) ONTOLOGY MODEL OF THE DECISION TREE CLASSIFIER

Ontologies typically express two algorithms, namely decision trees and semantic rules [55]. [58], [59] used decision trees in the field of ontologies to cluster and classify image objects. Findings proved that decision trees enhance ontologies to granulate information, thereby increasing image classification accuracy. [59] uses decision trees to solve the problem of inconsistency between overlapping ontologies. [47] use decision trees for ontology matching; the matching process is purely based on derived decision tree rules for an ontology that are compared with rules for external ontologies. [55] designed an ontology model for decision tree classifier that
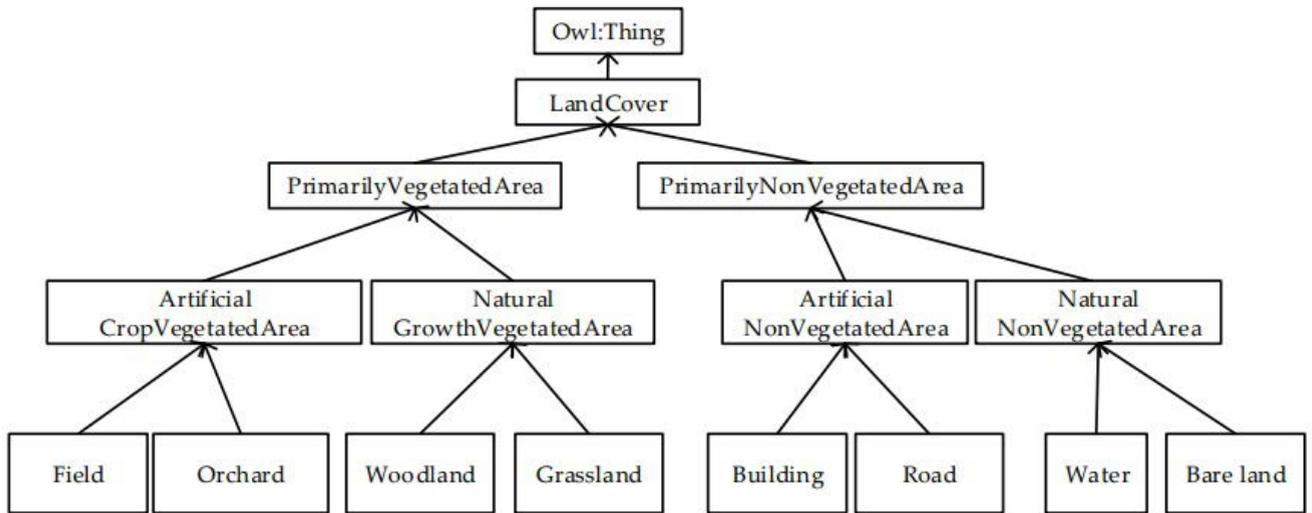
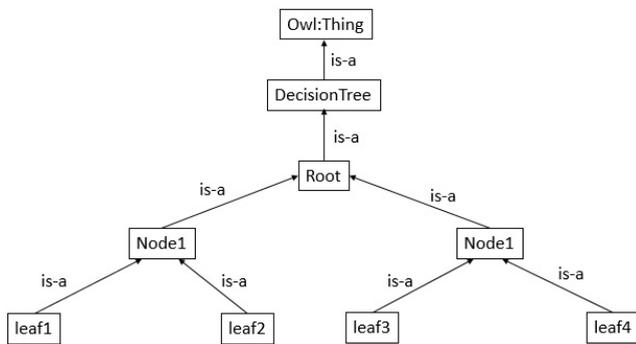**FIGURE 14. Land cover class hierarchy in ontology [55].**

FIGURE 15. Ontology model of the decision tree classifier [55].

consists of three parts; (1) a set of decision trees is composed of all essential terms and concepts, for instance, a node and a leaf; (2) a slot is defined by the following inequality symbols $> \geq, <, \leq$ (3). The final step is to create the nodes. Figure 15 shows the elements of the decision tree classifier.

2) ONTOLOGY MODEL OF THE SEMANTIC RULES

[55] followed a two phased approach to designing an ontology model for semantic rules; the first is the establishment of mark rules, followed by decision rules. Mark's rules convert low level features to semantic concepts. On the other hand, decision rules are inferred from mark rules and apriori knowledge.

- Ontology model for mark rules

The morphology of semantic notions is classified into strip and planar; the shape is regular and irregular; the texture is smooth and rough; the brightness is light and dark; the height is high, medium, and low; and the position relationship is adjacent, disjunct, and contained. The ontology model of the mark rules is shown in Figure 16

- Ontology model of the decision rules

Ontologies explicitly represent concepts in the same way humans describe concepts in their domain of interest. However, ontologies that are developed disregarding decision rules have proved to be computationally expensive [60]. This is due to their inability to capture the kinds of decision-making knowledge that arises in practice, such as those involving multiple ontologies. Decision rules on ontologies help in three ways, namely: [61], [62]; (a) they take into cognisance primitives from multiple ontologies as well as primitives that are not part of the rule framework; (b) they are time dependant (c) they incorporate default assumptions. Eight types of land cover obtained from the Chinese Geographic Census Project [57] were defined in terms of a rule as outlined in Figure 17.

C. SEMANTIC NETWORK MODEL

Semantic networks graphically represent knowledge in the form of nodes and links, whereby links provide hierarchical relationships between objects [63]. The semantic network model explicitly express knowledge through concepts and their corresponding semantic relations [55]. This is shown in Figure 18. The network bridges the gap between low-level characteristics and high-level semantics, reducing the semantic gap.

D. ONTOLOGIES FOR KNOWLEDGE MANAGEMENT

Framework ontologies and domain ontologies are the two most important types of ontologies. Frameworks, or foundation ontologies, consist of concepts explicitly expressed in high-level knowledge (for human understandability), and they are also not designed for a specific domain. A domain ontology has knowledge tailor-made for a specific domain, e.g., remote sensing. Domain ontology eave drops from framework ontology. Domain ontologies have a hierarchical structure of two levels; the first level is called the ABox,
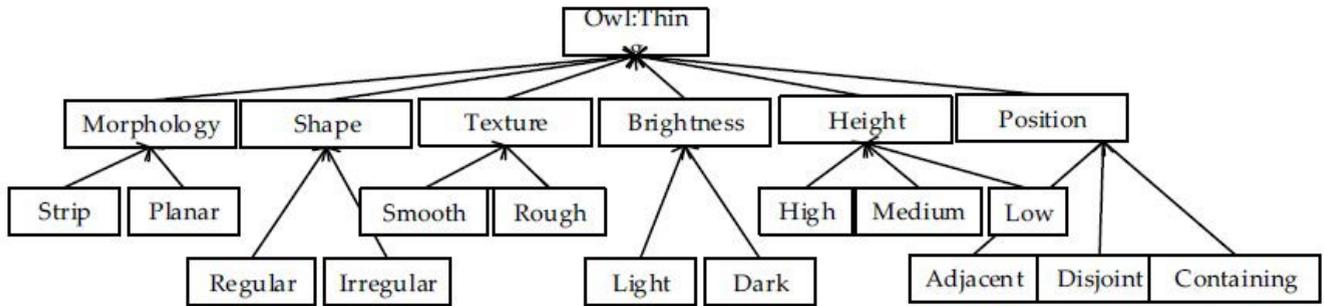
FIGURE 16. The mark rules in ontology [55].

1) Fieldland = Regular ∩ Planar ∩ Smooth ∩ Dark ∩ Low ∩ Adjacent to Road ;
2) Woodland = Irregular ∩ Planar ∩ High ∩ Rough ∩ Dark ∩ Adjacent to Fieldland;
3) Orchardland = Regular ∩ Smooth ∩ Planar ∩ Dark ∩ Adjacent to Fieldland;
4) Grassland= Irregular ∩ Planar ∩ Smooth ∩ Dark ∩ Low ∩ Adjacent to Building;
5) Building = Regular ∩ Planar ∩ Rough ∩ High ∩ Light∩ Adjacent to Road;
6) Road= Regular ∩ Strip ∩ Smooth ∩ Light ∩ Low;
7) Bareland = Irregular ∩ Planar ∩ Rough ∩ Lght ∩ Low;
8) Water =Irregular∩ Planar ∩ Smooth ∩ Dark ∩ Low ∩ Normal Differential Water Index(NDWI).

FIGURE 17. Decision rules based from ontology [55].

and the second level is called the TBox. ABox contains assertions (or rules) that comprise the theory that the ontology describes in its domain of application [64]. TBox is where experts conceptualise their knowledge in a specific scientific domain [47]. There are vast paradigms for modelling ontologies, but chief amongst them are Description Logics(DL) [65] and rule formalism. The DL formalism serves as a foundation for building ontologies using the web ontology language (OWL) [66]. Ontologies can be inferred from new knowledge using DL, which makes ontologies machine understandable.

E. MODULAR ONTOLOGICAL APPROACH

The modular approach is the best way of building complex ontologies from simpler (modular) ontologies in a constant and well-defined way [67]. Such an approach allows collaborative development by many different domain experts to build a single ontology through the integration of independently developed ontologies. The ontological approach is carried out in such a way that TBoxTs are not changed when elements of T' are reused in another TBoxT. Formalisation of such a property follows the conservative theorem [68].

Definition 1 (Conservative Extension): Let T and T' be TBoxes, Sig(α) be a signature of axiom α and Sig(T') be
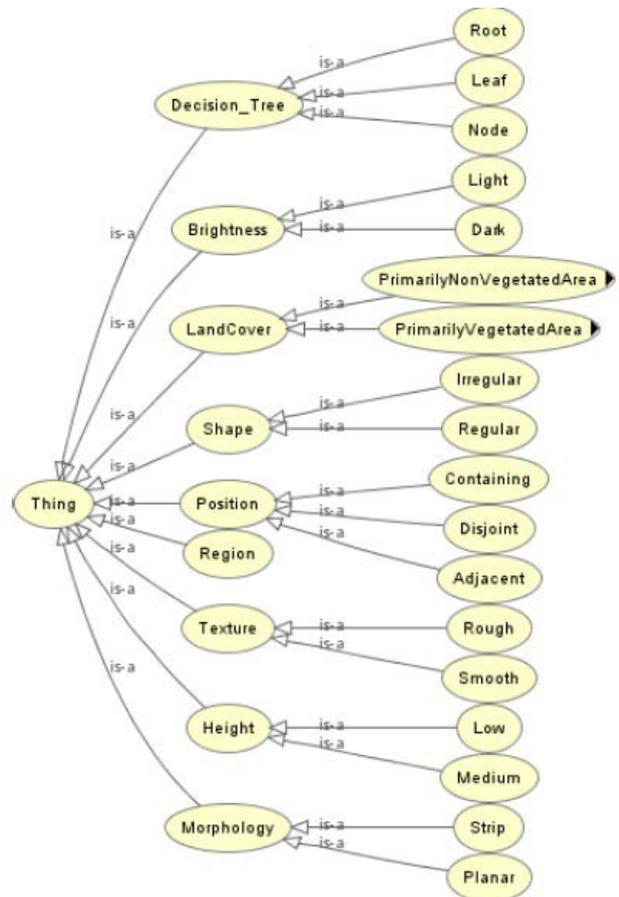
FIGURE 18. The semantic network model [55].

signature of TBoxes of T'. Then $T \cup T'$ is a conservative extension of T' if for every axiom α with Sig(α) ⊆ Sig(T') we have $T \cup T' \Rightarrow \alpha$ iff T' $\Rightarrow \alpha$ [67]. In addition, if two independent parts T_1 and T_2 of an ontology T, are constructed in a modular way, then T remains modular as well. These are formalised as follows [67]:

Definition 2 (Modularity): Let Loc(T) be a local signature T and Ext(T) be external signature. A set M of TBoxes T with Sig(T) = Loc(T) ⊎ Ext(T) is a modular class if the following condition holds:

FIGURE 19. Structure of knowledge base [47].

M1. if T ∈ M, then T ∪ T' is a conservative extension of every T' such that Sig(T') ∩ Loc(T) = θ

M2. If T_1, T_2 ∈ M, then T = T_1 ∪ T_2 ∈ M with Loc(T) = Loc(T_1) ∪ Loc(T_2)

Falomir at al [66] proposed three levels of knowledge that are imperative for designing a modular ontological approach: the reference conceptualisation (which p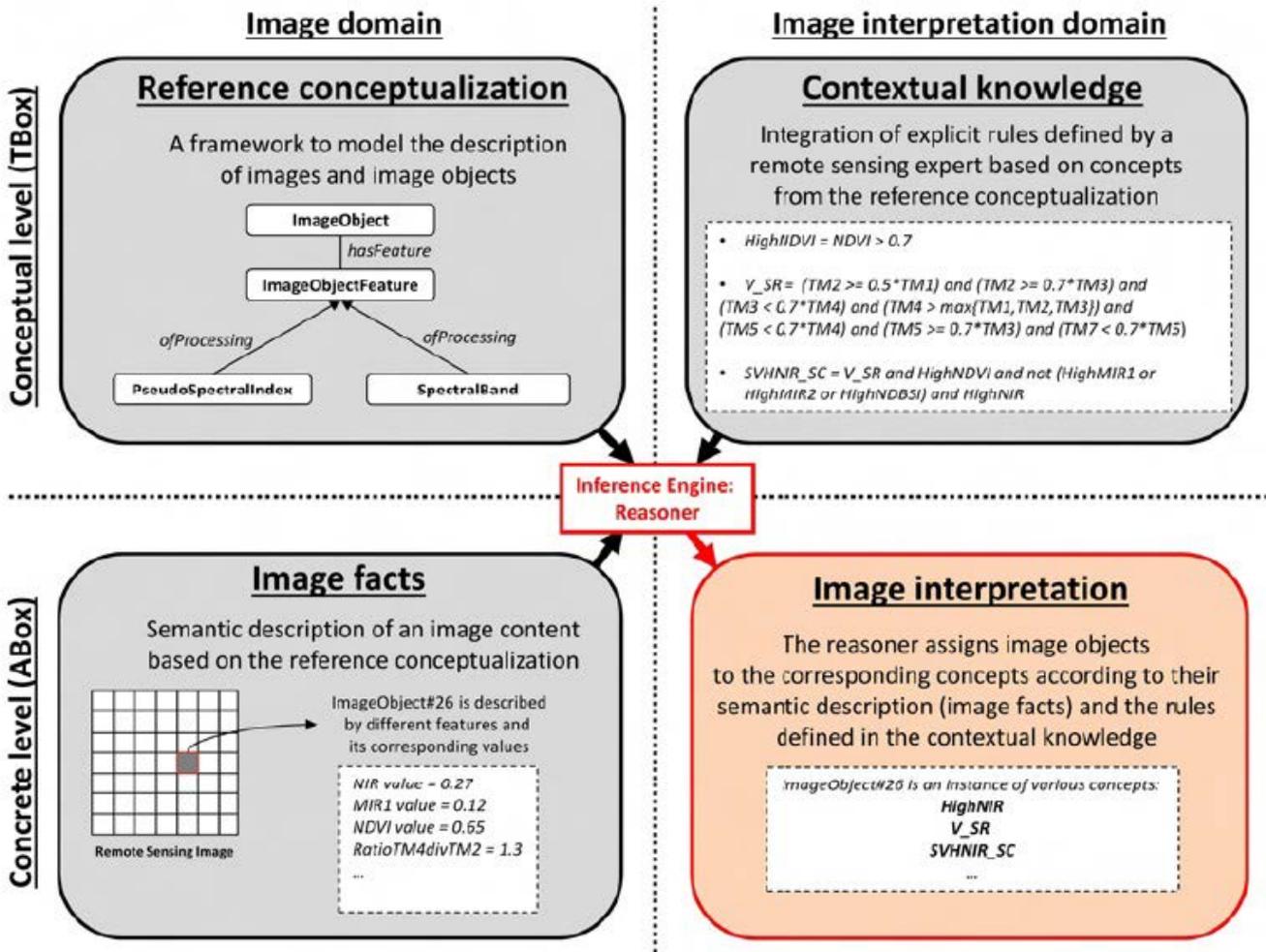rovides a description of images and image objects), the contextual knowledge (a set of rules defined by a domain expert) and the image facts (these are semantic descriptions of image content). Figure 19 illustrates how the reasoner assigns image objects to their corresponding concepts based on facts drawn from reference conceptualisation and contextual knowledge also drawn from reference conceptualisation.

(a)   The reference conceptualisation

It is a general model for describing image objects in remote sensing. It consists of two packages, namely, (1) the image structure package and (2) the image processing package [47]. The image structure package is superimposed with the ImageObjects concepts, which describe objects according to their characteristics, and the ImageObjectFeature concept, which links related concepts with associations such as "hasfeature". The image processing package is composed of the PseudoSpectralIndex and SpectralBand concepts. The concepts help remote sensing experts describe contextual knowledge. Concepts such as spectral bands and texture are used by remote sensing experts to interpret remote sensing images.

(b)   The Contextual Knowledge

Contextual knowledge's purpose is to represent remote sensing expert knowledge using DL, hence the name "contextual knowledge." The basis of this knowledge comes from the Remote Sensing Science expert. As a result, it is a "subjective" description of image rules rather than an "objective" depiction of image structure. Figure 20 shows the concepts, relations, and instances in conceptual knowledge.

(c)   The Image Facts

These are facts extracted from image analysis, and they are stored in the ABox [47]. The TBox contains the reference conceptualisation and the contextual knowledge [47]. Facts in ABox provide semantic descriptions of image objects, and
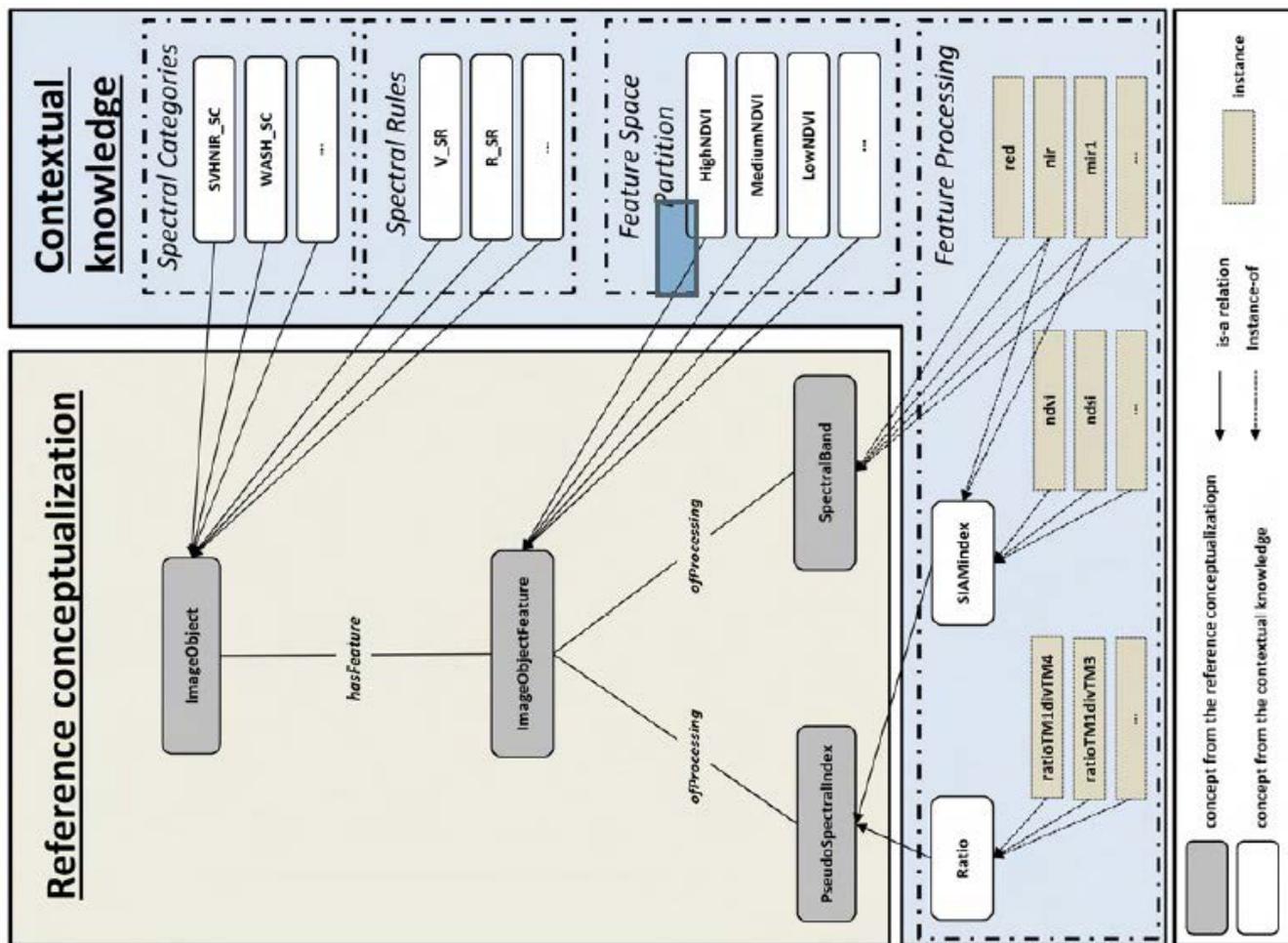
FIGURE 20. Conceptual knowledge showing concepts, relations and instances in an ontology [47].

the description is done with the help of reference conceptualisation and conceptual knowledge.

V. VEGETATION DETECTION
Unsupervised and supervised classification algorithms are very crucial in identifying vegetation areas.

A. UNSUPERVISED CLASSIFICATION INDICES
Spectral indices are used in these methods to detect vegetation areas. The Normalized Difference Vegetation Index (NDVI), which is calculated for each pixel in an image, is one of the indices utilized. The NDVI image is represented in a gray scale image. As shown in Figure 21: image (a) is a representation of an image using the RGB channel; image (b) is the representation of the same image in an NDVI format using the gray scale.

$$NDVI = \frac{\psi IR - \psi R}{\psi IR + \psi R} \quad (11)$$

Equation 4 illustrates the calculation of NDVI, where ψIR and ψR are pixel values in the infrared and the red band respectively. The formula defines vegetation as areas that have a higher reflective index in the infrared than the red band
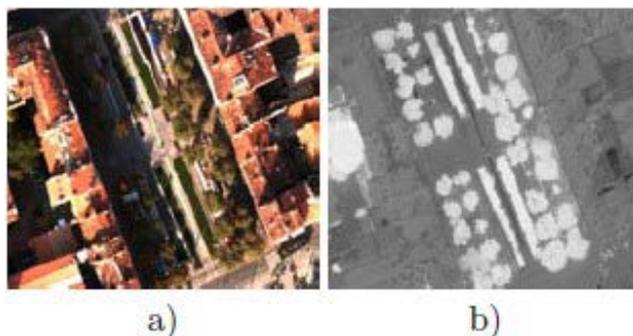
FIGURE 21. (a)RGB input image (b) NDVI image [69].

index. The formula was then refined to take into account the spectral index [70].

$$SI = \frac{\psi R - \psi B}{\psi R + \psi B} \quad (12)$$

Equation 5 illustrates the calculation of SI, where ψB is the pixel value in the blue band and ψR is the pixel value in the red band. An NDVI value and a SI value are binarized to create a vegetation mask. This is shown in Figure 22.
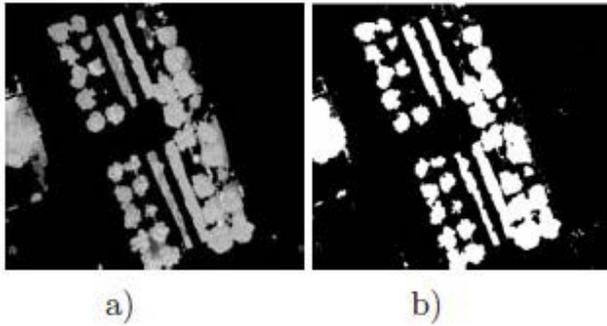
FIGURE 22. Image (a) SI image (b) vegetation mask obtained with the NDVI and SI indexes [69].

B. SUPERVISED CLASSIFICATION INDICES

Detection of vegetation by spectral indices is highly dependent on spectral characteristics. In other cases, supervised classification methods are primarily based on Support Vector Machines (SVMs). The feature vector that defines all pixels in the training data set contains four characteristics, namely: the reflectance value of each pixel in the infrared, red, green, and blue. Supervised methods do well in distinguishing between non-vegetation and vegetation areas through spectral indices. It necessitates the use of a SVM capable of determining the best linear separator. Random Forest (RF), k-Nearest Neighbour (kNN), SVM and sparse representations are among pixel wise classifiers that have been used for the last decade [71]. These traditional methods only consider spectral information as the basis of the classification process, disregarding spatial contextual information which contributes significantly to the classification performance [71]. Several researchers have proposed a hybrid of spectral-spatial classification that takes into account both the spatial context and spectral information, based on the assumption that pixels from a local region have similar spectral information. [71] proposed a hybrid model of kNN combined with guided filter for hyper-spectral image (HSI) classification of forest trees. Joint hybrid model of kNN and guided filter (PGF-kNN) was used to optimise hyper-spectral images produced by kNN. Optimised hyper-spectral images were taken in as input by the Joint kNN, and processed to produce the classification maps. Each class map was converted into a probability value and the class map with the highest probability value was chosen as the classification result. [72] conducted a study to determine the reliability of RF and SVM algorithms in the classification of very high resolution images (VHR), obtained from oak woodlands of a Mediterranean ecosystem. The first stage was data acquisition, where images were subjected to a Structure-Form-Motion (SFM) technique to identify common features in overlapping images. Each image was then orthorectified through the interpolated digital surface mode (DSM). Finally, all the images were combined into an orthomosaic. The workflow of the study followed 4 main steps, namely, pre-processing, segmentation, classification and accuracy assessment. Figure 23 shows the workflow of the proposed model. In the preprocessing stage each input layer was subjected
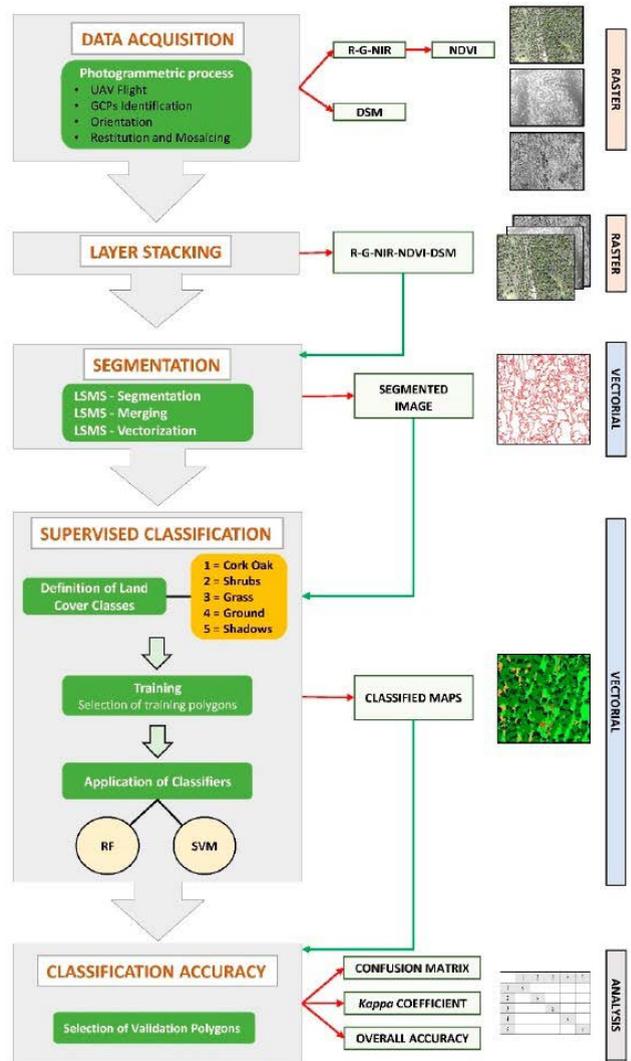
FIGURE 23. Workflow that presents the stages of preprocessing, segmentation, classification and accuracy assessment [72].

to a linear band covering a range of 8 bits, that is, from a minimum of 0 to a maximum of 255. The process was done to normalise each band, to suppress the effect of possible outliers on the segmentation. A layer stretching process was performed on images containing R-G-NIR (Red, Green, Near-Infrared) bands, obtained during spring and summer seasons through integrating NDVI and DSM data, to obtain the final 2 five band orthomosaics. Such a process was of significant importance because OTB segmentation requires only one rasta image as the input data. Spectral separability is of significant importance when it comes to image classification. The M-static defined in Equation 13 was employed [72] to measure the separability of NDVI and DSM layers of varying types of vegetation.

$$M = \frac{\mid \mu_1 - \mu_2 \mid}{\sigma_1 + \sigma_2} \qquad (13)$$

where, μ_1 is the mean value of class 1 and μ_2 is the mean value of class 2. σ_1 is the standard deviation of class 1 and

FIGURE 24. Flowchart of the model that harmonises RF and kNN [73].

σ_2 is the standard deviation of class 2. If M < 1 it signifies overlap of classes, if M > 1 it denotes that classes are well separable. The segmentation process considered both semantic properties and radiometric information. Large-scale mean shift (LSMS) segmentation was used in the stu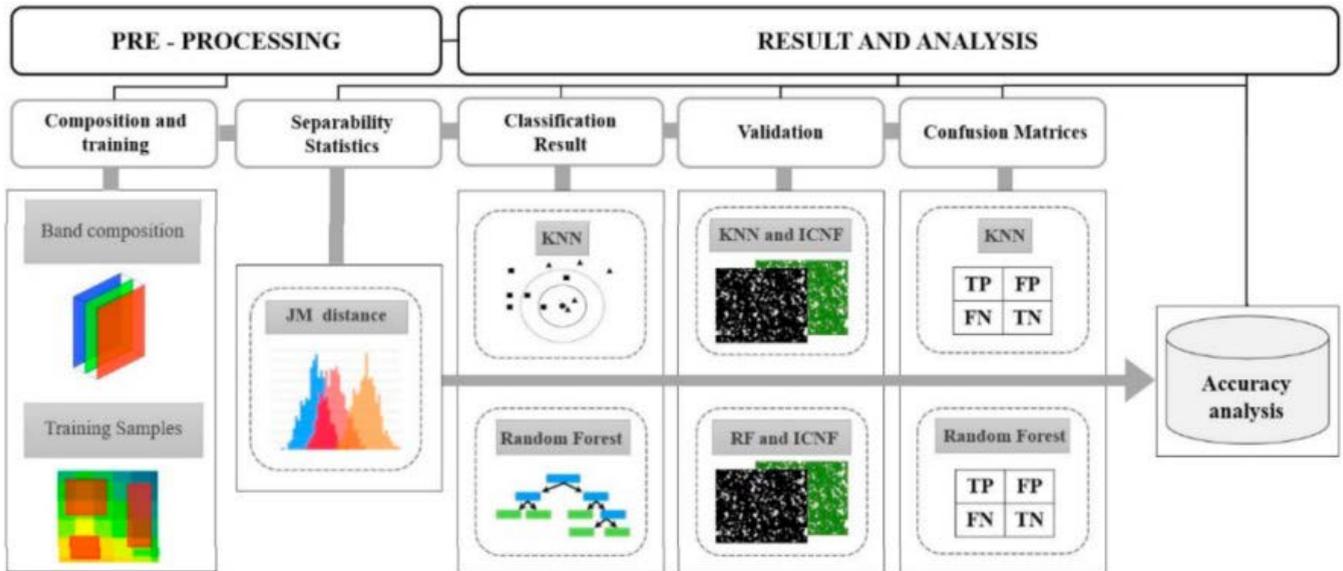dy because of its ability to perform tile-wise segmentation of large VHR imagery [72]. The OTB LSMS segmentation process followed the steps of LSMS smoothing, LSMS segmentation, LSMS merging and LSMS vectorisation. Classification was performed for five different land cover classes, namely, grass, cork oak, soil, shrubs and shadows. Two supervised learning algorithms including RF and SVM were used to perform the classification. SVM performs linear separation in a hyperspace using a $\mu(.)$ mapping function. In the case where objects are not linearly separable, the kernel method is used where it takes into account projections of feature space [72]. RF uses decision trees for bagging to produce different subsets of variety of trees. Every decision tree in the RF participates in the classification process and the classification label returned is the class with the most votes.

Another study [73] analysed the performance of kNN and RF classifiers for mapping forest fire areas. The authors [73] implemented kNN and RF to classify forest areas and explained the effects of different satellite images on both classifiers. Figure 24 shows the flow chart of the model. The model being a supervised approach was implemented by using multi-spectral images obtained from Landsat8, Landsat-2, and Terra sensors. The classification accuracy was determined by the confusion matrices. The machine learning classifier based on kNN and RF produced excellent results with k set to 5 for kNN and 400 trees for RF. The results from the hybrid model achieved a very high classification accuracy with an Overall accuracy (OA) > 89% and Dice coefficient (DC) > 0.8. Other studies [74], [75] have also implemented non-parametric algorithms such as kNN and RF in remote sensing applications.

VI. IMAGE SEGMENTATION

An input image is partitioned (or subdivided) into meaningful image objects (segments). Image segmentation can be classified into two categories: supervised (empirical discrepancy methods) and unsupervised (empirical goodness methods) [76]. Unsupervised approaches evaluate a segmentation result based on how well the image object matches a human perception of the desired set of segmented images, and they use quality criteria that are typically created in accordance with human perceptions of what constitutes a good segmentation. Supervised methods compare a result from segmentation with a ground truth [2]. If ground truth can be reliably established, supervised methods are preferred.

A. TYPES OF IMAGE SEGMENTATION

Pixel, edge, and region-based image segmentation methods are the three primary types of traditional image segmentation. [77].

(a)   Pixel Based Methods

This method involves two important processes: (1) image thresholding and (2) segmentation in feature space. For image thresholding, image pixels are divided according to their intensity level [78]. There are three types of thresholding [79], [80]:

(1)   Global thresholding - T being the appropriate threshold value. The output of an image q(x,y) based on T is obtained from an original image p(x,y) as

$$q(x, y) = \begin{cases} 1, & if p(x, y) > T \\ 0, & if p(x, y) \leq T \end{cases}$$

(2) Variable thresholding - This when the value of T varies varies over an image and it comes in two flavours:

- Local Threshold - T depends on the neighborhood of x and y.
- Adaptive Threshold - T's value is a function of x and y.

(3) Multiple thresholding - It has multiple values of T. The output image is computed as follows:

$$q(x, y) = \begin{cases} m, & if p(x, y) > T1 \\ n, & if p(x, y) \leq T1 \\ 0, & if p(x, y) \leq T0 \end{cases}$$

However, these methods suffer from incomplete segmentation, so the output results need to be clumped. Also, these methods are appropriate for images with lighter objects than the background.

(b) Edge Segmentation methods

Edge-detecting operators are employed to detect all possible edges that are found in an image. Adjacent edges are clearly separated by a gray sharp edge, but there could be a case where the gray value is not continuous [81]. The edges will be represented by discontinuity in gray level, color, texture, etc. This discontinuity is detected by using derivative operations such as differential operators [82]. The Prewitt, Roberts, and Sobel operators are the most frequently utilized first order differential operators [83]. There are a number of edge detection operators such as the template matching edge detectors. One challenge with edge-based segmentation is that sometimes it presents edges in locations where there is no border. Filtering, enhancement, and detection are the three processes in edge segmentation algorithms [77]. The purpose of the filtering process is to reduce the amount of noise present in the imagery. The enhancement uses high pass filtering to detect and reveal local changes in intensity. Finally, the edges detected (using threshold techniques) are combined or linked together to form the boundaries of the image object. One challenge with edge-based segmentation is that, sometimes it presents edges in locations where there is no border.

(c) Region Based Segmentation

Images are segmented into regions with similar properties using region-based approaches [81]. There are three types of region based segmentation, namely: (1) region-growing segmentation, (2) region-splitting and merging segmentation [82], [84].

(1) Region Growing Segmentation

It starts with the matrix's origin (seed point), which is then subjected to a rule that joins surrounding pixels to these starting regions, and the procedure is repeated until a particular threshold is met [81]. The method is repeated until there are no more pixels to ascribe. This process is repeated until the entire image is segmented. The algorithms, on the other hand, suffer from a lack of control over the region's growth break-off criterion [85].
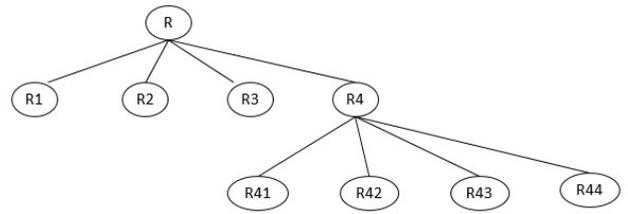

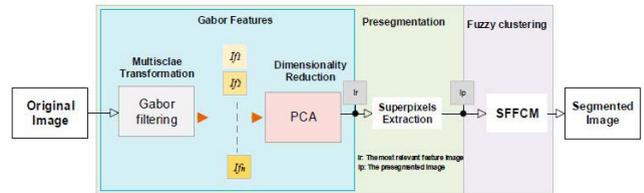**FIGURE 25.** Region splitting.

FIGURE 26. Image segmentation state of the art [76].

(2) Region splitting and region merging

The original image is split or subdivided into sub images. Each sub-image is recursively divided into its own sub-images based on the condition or predicate given. If the condition is not satisfied, further splitting ceases [82]. Figure 25 shows the splitting process.

B. IMAGE SEGMENTATION STATE OF THE ART

Reference [76] proposed a segmentation process that improves segmentation accuracy by modifying the super-pixel extraction methodology so as to increase robustness to added noise. The segmentation method is based on Gabor filtering and Principal Component Analysis (PCA). Figure 26 presents the state-of-the-art segmentation process. The method depends on two principal tasks: (1) pre-segmentation (super-pixel extraction), and (2) clustering of previously extracted pixels.

(a) Pre-segmentation

An input image is subdivided into a number of regions of interest. Each region is made up of pixels with similar features. The Watershed Transform (WT) clustering based super-pixel algorithm has previously been considered for super-pixel extraction [86], [87].

(b) Gabor filter

Gabor filters are used to extract spatially localized spectral features [76]. They have been advocated for because they are based on principles found in similar human visual systems and have key features that can be utilized to segment images.

Before the introduction of deep learning, machine learning techniques such as SVM, K-means clustering, Random Forest, etc., were the chief algorithms for image segmentation. Semantic segmentation using deep learning has proven to work better than the aforementioned techniques because they classify each pixel of an image rather than the entire
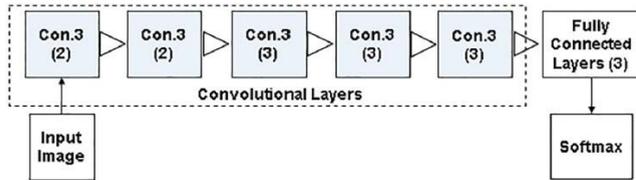
FIGURE 27. AlexNet [89].

FIGURE 28. VGGNet [89].

image object. The next chapter gives an overview of semantic segmentation techniques.

VII. SEMANTIC SEGMENTATION USING DEEP LEARNING
This section introduces fundamental ideas of CNNs and subsequent variants for semantic segmentation, as well as their network structures [88].

A. AlexNet, VGGNet AND GoogleNet
These are the three chief deep neural networks for image classification, which formed the major foundations of later developments. The networks support network architectures for semantic segmentation.

1) AlexNet
AlexNet is made up of five convolutional layers and three connected layers [89]. In between the convolutional layers is a pooling layer whose role is aimed at reducing dimensionality and computational complexity. AlexNet's pooling strategy is max pooling, and the strategy is to obtain the biggest value covered by the filter, which is used to remove noisy components [88]. Filters of sizes 11×11 and 5×5 are used in the first and second convolutional layers, respectively. The last three layers use small-sized filters of 3×3. The whole process is described in Figure 27. The primary purpose of such filters is to be solely used for feature extraction. Varying filters accommodate objects of different scales.

1) It supports the application of non-saturating Rectified Linear Unit (ReLU) whose output is defined by

$$F(x) = max(x, 0).$$

2) It employs the overlapping max pooling strategy (which means that each filtering operation's step size (stride) is smaller than the filter's overall size).
3) To reduce over-fitting, it uses the dropout approach in fully-connected layers.

2) VGGNet
The network is made up of three fully connected layers and a varying number of convolutional layers. This is shown in Figure 28. Unlike AlexNet, VGGNet has fixed small size filters of 3×3 in the convolutional layer [90]. The number of weights in the network is reduced by using small filters, which minimizes the training complexity. Just like AlexNet, VGGNet uses max pooling over a 2×2 window slide of 2 pixels. The advantage of simplifying convolutional layers
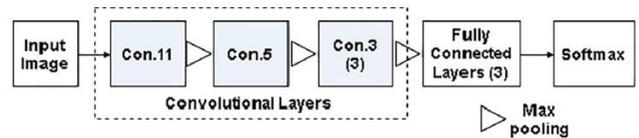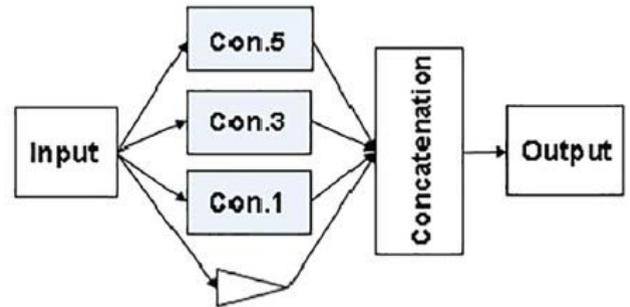
FIGURE 29. GoogLeNet [89].
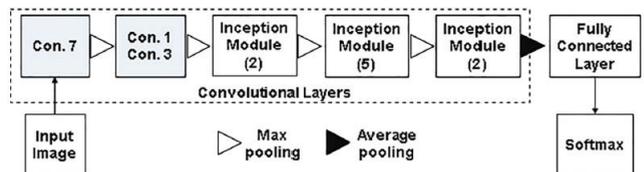
FIGURE 30. Inception Module [69].

to a greater extent is that it increases network depth, thereby improving the accuracy of the network. The network's performance in tasks like semantic segmentation and target detection is improved by using features extracted from CNN that are structured in a hierarchy of scales [91]. Other classifiers, such as SVMs, can use the features without fine-tuning [92].

3) GoogLeNet
The architecture is different from the other three in that it involves three aspects, namely the inception module, at the training stage, an auxiliary classifier is required, as well as one fully connected layer [93]. Output results from these filters are concatenated with the maximum pooling result. Between the inception modules, maximum pooling is employed, and after the last inception module, average pooling that employs dropout is used [94]. The flow chart diagram is shown in Figure 29. The network is so deep because it is made up of nine inception modules and up to three convolutional layers. Because of the profundity of the network, the smooth flow of gradient from layer to layer becomes an issue. Figure 30 shows the Inception Module. The issue is addressed by adding an auxiliary classifier in the middle of the convolutional layers, whose role is to process the outputs from the inception modules. The loss from these classifiers is added to the overall loss of the network during training. Auxiliary classifiers are prohibited from making decisions during the prediction phase.

FIGURE 31. FCN Network [95].

B. FULLY CONVOLUTIONAL NETWORK

Fully convolutional networks for semantic image segmentation are an extension of AlexNet, VGGNet, and GoogLeNet [95]. Multi-convolutional, deconvolutional, and fusion are the three steps that define the network. The flow chart is shown in Figure 31. Convolutional layers have been substituted for fully linked layers, with the specification that each image's score be computed using a 1×1 convolution. Because of pooling, the outp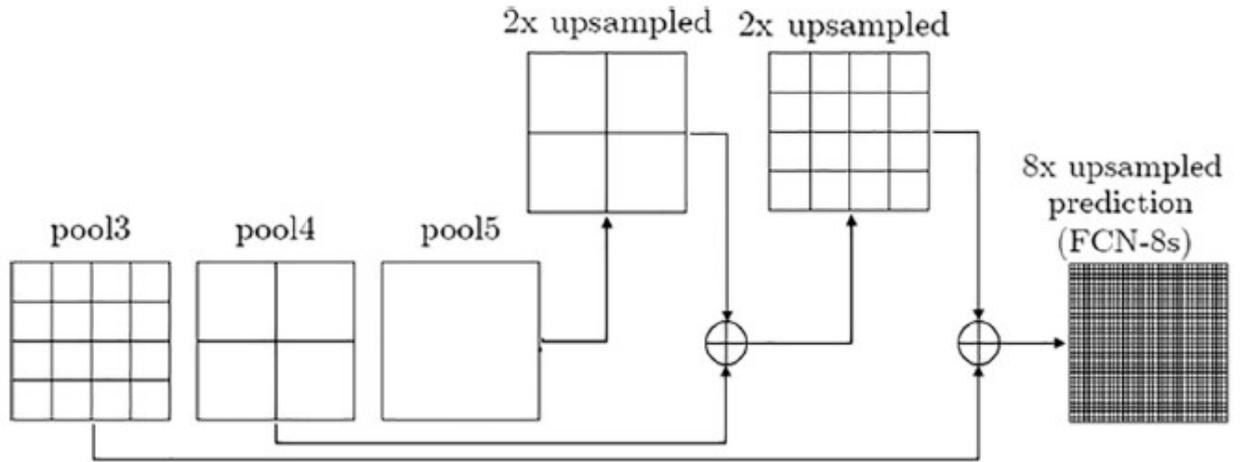ut image from convolutional layers is smaller than the input image. The deconvolutional process is used to restore the image. It uses the same methods as the convolutional process, but cushions the framework (by padding the matrix) and joins the elements inside a deconvolution filter to increase the input size. The process of recovering the original image through the deconvolution process has some side effects; for example, some details are lost as a result of the dilution of class scores. To circumvent the side effects, the skip architecture combines semantic information obtained from layers with location details obtained from previous layers. By element wise addition, the up-sampled deep layer is fused with the yield or output of a shallow layer.

C. UNET

The building blocks of Unet are the convolutional and deconvolutional layers. The network works well with small images, hence the paramount step is downsizing of input images [96]. Convolutional layers use filters of size 3×3 which produce output images that are subsequently subjected to Relu for processing, followed by maxpooling (which uses a stride of two). Maxpooling generates downsized outputs. Feature channels in the convolutional layers double at each and every step. The deconvolutional layer does upsampling, but a 2×2 convolution is used to limit the number of features to the required standard. The network generates the segmentation result by applying a 1×1 convolution on the feature map
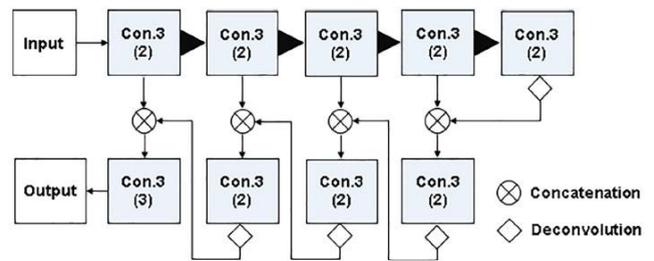
FIGURE 32. Unet [89].

and labeling pixels. The interconnection of layers in Unet is shown in Figure 32.

D. SegNet

The network is composed of two subnetworks, namely; the encoder and decoder networks. The encoder network's mandate is the downsizing of feature maps. It consists of a varying number of convolutions and subsequent maxpooling operations for feature extraction [97]. However, features produced have vague or ambiguous spatial information. The issue is solved by saving an element index that will be used later in the decoder network's up-sampling procedure. Convolutions map low-resolution features to high-resolution features in the decoder network. A 2×2 low-resolution feature, for example, is up-sampled to a 4×4 matrix. This process may result in the loss of spatial information; therefore, reusing the pooling index from the encoder network completely recovers the lost information. The SegNet network is depicted in Figure 33.

E. DeepNet

It is a variant of FCN that employs dilated convolution to broaden the scope of filters to include image context in a larger neighborhood while also allowing for flexibility over feature response resolution [17]. Deeplab uses Atrous
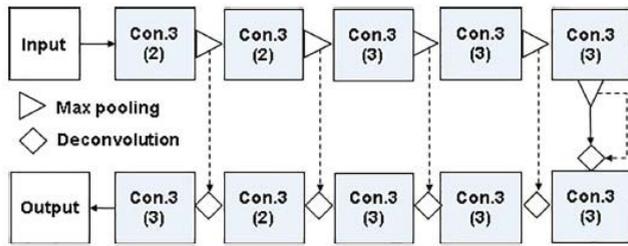
FIGURE 33. SegNe [95].
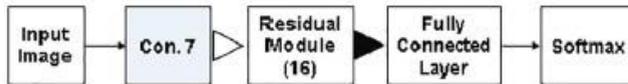
FIGURE 34. Residual Net [89].

FIGURE 35. Residual block [95].

Spatial Pyramid Pooling (ASPP) for up-sampling. Several atrous convolutions operated on the same kernel but with various sampling rates are used in the scheme. An additional operator combines the output from all convolutions. Downsampling processes and subsequent maxpooling operators make segmentation results lose some fine details. To solve the problem, conditional random filters (CRFs) are employed to improve the spatial localization of segmentation. CRF models contribute to the smooth segmentation process based on the underlying image intensities [98]. They boost the accuracy score by 1% to 2%.

F. ResNet

The residual network is well recognized for its 152 layer depth and residual block introduction [99]. The residual block is presented in Figure 35. As based on traditional neural networks, the greater the number of layers, the better the performance of the network. However, because of the vanishing gradient problem, first layer weights will not be updated correctly through the backpropagation algorithm [100]. As the error gradient is propagated to earlier layers it goes through a repeated multiplication process such that the gradient becomes very small hence the network performance gets saturated and will start to decrease. This problem is solved by using the identity function, whereby the gradient is multiplied by one so as to preserve the input and avoid any loss in the information. The network is made up of the following components; 3×3 filters, CNN downsampling layers with a stride of 2, global average pooling, and a 1000-way fully connected layer with softmax at the end. ResNet employs a skip relation, which means that an original input is also connected to the convolution block's output. This aids in the solution of the vanishing gradient problem by allowing the gradient to flow in a different direction. The network diagram of the residual network is shown in Figure 34.

G. APPLICATION OF DEEP LEARNING TECHNIQUES

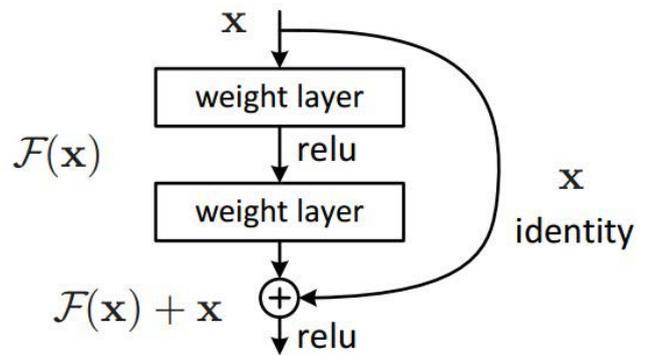New emerging technologies such as deep learning have gained ground in the remote sensing science fraternity because the automatic processing of images by these techniques chiefly depends on human expert knowledge, which has impacted the way land surveys are done [101]. The main advantage of deep learning approaches is the automatic computational extraction of features, unlike other machine learning algorithms where feature extraction is typically manual [102]. The strength of deep learning algorithms lies in learning from examples. The learning process consists of a number of steps: first, an architecture of a network of nodes is clearly defined. The nodes that form an Artificial Neural Network (ANN) are arranged into layers. An ANN with many layers is referred to as a Deep Neural Network (DNN). The behaviour of the DNN is determined by the type and number of nodes as well as the connection between the nodes [101]. If an existing DNN is to be customized for an new application context, its weights are recursively updated to achieve the new desired response. This process is referred to as "transfer learning". Deep learning was originally used for locating and classifying different tree species in a mosaic built from UAV-acquired images [103], [104]. [105] devised a deep learning technique to detect and identify tree species. The objective of the study was to classify patches corresponding to tree species. The authors developed a Deep Learning (DL) architecture, which is a hybrid of ResNet and UNet, to come up with a semantic segmentation algorithm for tree spices that is precise and efficient. Seven orthomosaic images were collected using UAV in the winter, and one orthomosaic image was collected using UAV in the summer. The algorithm pipeline is presented in Figure 36. The first step of the technique identified the classes corresponding to each mosaic patch. The focus was on classifying the pixels in each mosaic patch. The incorporation of the ResNet architecture into the DL network enhanced the accuracy and efficiency in classifying forest images [104], [106]. Images were divided into patches in response to the prescribed annotations, and each patch was assigned to a list corresponding to the classes that matched it. Patches could belong to more than one class, resulting in patches having to be labelled repeatedly. Because of the repeated labelling of patches, the algorithm is referred to as a Multi-label Patch (MLP) based classifier. The ResNet architecture went through the training
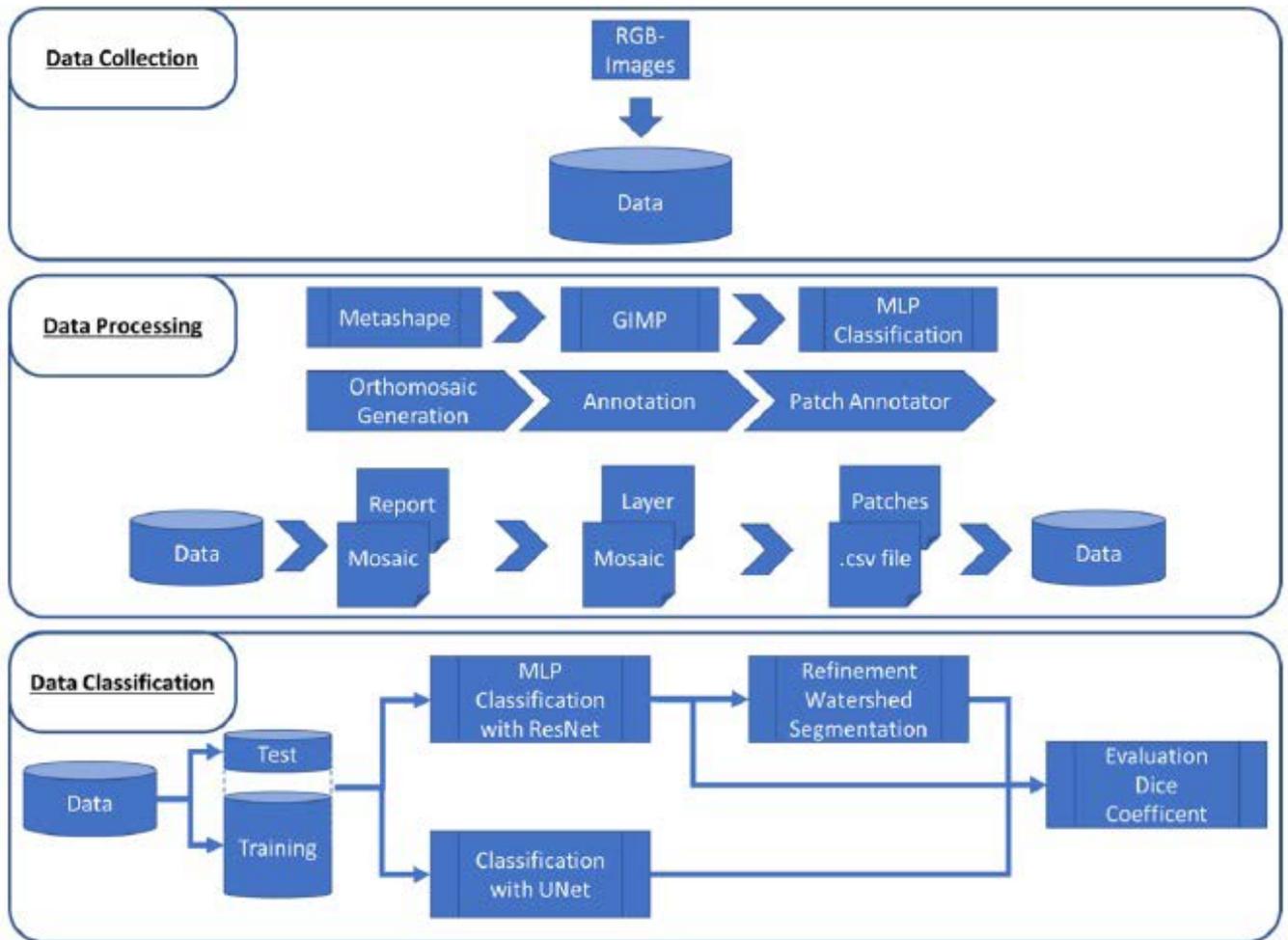
FIGURE 36. Algorithm pipeline [106].

phase so that it would be able to classify the patches. The MLP classification algorithm produced coarsely segmented images. A watershed segmentation algorithm was applied to refine the segmentation process. The UNet architecture, originally used for medical image segmentation [96], is also very useful for remote sensing images. The UNet architecture was trained with data and pixel-wise annotation patches. The segmentation process follows a number of steps: (1) mosaic images were split into patches for processing, (2) a UNet model was trained to predict patch segmentation, and (3) patch joining was used to obtain semantic segmentation for the entire mosaic image. The model achieved an effective learning transfer with a 12.48% improvement over random weights. Overall, the model reached a higher accuracy of nearly 95%.

Another study [104] proposed a Residual Neural Network (ResNet) architecture for classifying tree species acquired using a camera mounted on a UAV platform. In temperate forests, UAV images have been successfully used to distinguish between living and dead forest species [107]. The motivation of the study was that, most of the existing methods

for tree species classification are cost-sensitive because they require very large data sets and are restricted to specific tree species [108]. The study proposed a model based on CNN to classify tree species at an individual level by analysing high resolution RGB images obtained from the UAV. A CNN was chosen in the study because of its ability to learn highly descriptive features from tree canopies. The study proposed a CNN model with 50 convolutional layers, referred to as ResNet50. Figure 38 shows the architecture of ResNet50. The procedure for performing tree crown delineation was based on the iterative local maxima filtering technique that was used to identify probable tree tops. Tree tops were designed as markers, hence a marker controlled watershed segmentation was performed as a means of complementing the DSM for segmenting the tree crowns. Figure 37 shows a tree crown segmented polygon. The tree crown delineation process enables tree crown identification labelling. In the training phase, images were shuffled in unison with their corresponding labels to randomise the input data so that the neural network becomes generalised. The model achieved an overall classification accuracy of 80%. The study concluded

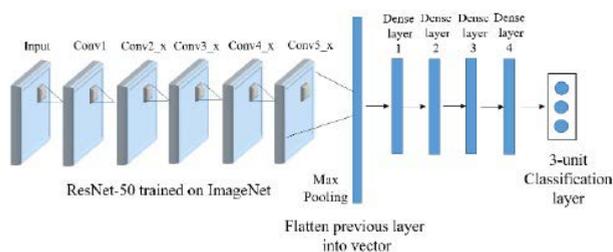**FIGURE 37.** Tree crown delineation [104].

FIGURE 38. CNN model architecture [104].

that classification accuracy increases with an increase in the number of training images.

The task of classifying and mapping vegetation images has been difficult because the conventional methods employed are highly labour intensive. Deep learning and CNN came as solutions to the problems posed by traditional methods, but they are still not efficient in detecting ambiguous objects [109]. There is a little research that employs CNN to detect and classify vegetation in remote sensing science images [109]. A study by Guirado [110] successfully used CNN to detect wild shrubs from Google Earth images. The author demonstrated that a CNN is much better than traditional object detection methods. Another study [109] used a deep learning model and the chopped picture method to detect vegetation from Google Earth images. The study was carried out against the backdrop that existing work still faces huge challenges in classifying vegetation that has ambiguous and amorphous shapes, such as clonal plants. The training data was prepared using the chopped picture method, and images were put into two sets; one set with images completely covered with bamboo trees and the other set without bamboo trees. Images were then chopped into small squares and subsequently used as training images. A classical deep learning model in the form of a LeNet network was employed by the study because it is efficient in processing small-sized images. The network is composed of two convolution layers,

two pooling layers, and one fully connected layer. The final layer was used to detect bamboo coverage in Google Earth images. Input images were randomly shuffled to alleviate overlapped training and validation data. 72% percent of the data was used for training and 25% of the data for testing. The model achieved an average classification accuracy of 97.52%.

VIII. FEATURE EXTRACTION TECHNIQUES
This section delves into the main techniques for feature extraction, and these include (1) Principal Component Analysis (PCA); (2) Independent Component Analysis (ICA); (3) Linear Discriminant Analysis (LDA); and (4) t-Distributed Stochastic Neighbor Embedding (t-SNE).

1) PRINCIPAL COMPONENT ANALYSIS (PCA)
PCA is popularly used as a dimensionality reduction technique [111]. It was first proposed by [54]. From the original data input, the PCA method tries combinations of input features in order to determine the best features that summarise the original data. This is accomplished by looking at pair-wise distances to maximize variances and minimize reconstruction error [112]. Since PCA is an unsupervised learning algorithm it leads to misclassification of data in some cases [111]. Distortion errors arise when data is reconstructed back because samples would have been projected onto a subspace [113].

2) INDEPENDENT COMPONENT ANALYSIS (ICA)
ICA, like PCA, is a linear dimensionality reduction method that combines discrete components to produce input data with the goal of correctly identifying each of them [111]. It is based on the principle that two features are deemed independent if their linear and nonlinear dependence are both zero [114]. Independent Component Analyses are extensively used in medical applications such as Electroencephalography (EEG) and Functional Magnetic Resonance Imaging (FMRI) analysis to differentiate useful from unhelpful signals [111].

3) LINEAR DISCRIMINANT ANALYSIS (LDA)
LDA is a supervised learning dimensionality reduction technique and a machine learning classifier [111]. The method is similar to PCA in the sense that it calculates the projection of data along a direction, but instead of maximising variation of data, the LDA uses label information to determine a projection by maximising the ratio of between class variance to within class variance [113]. The goal of LDA is formulated as the Fisher criterion [115].

$$J(u) := \frac{u^T S_B U}{u^T S_W U} \qquad (14)$$

Recently, this technique has been used for indoor positioning or localisation systems for the purpose of obtaining superior and higher accuracy [116]. The performance of LDA in the construction of data using independent variables is directly proportional to the number of data patterns [116]. However, its performance is yet to be confirmed in the context of non-linearity [117].

4) LOCALLY LINEAR EMBEDDING (LLE)

The LLE is built on a foundation of manifold learning. A manifold is a D-dimensional object that is embedded in a higher-dimensional space. A manifold is considered as an integration of small linear patches, which is done through piece-wise linear regression [118]. To do the integral operation, [119] proposed the construction of a kNN graph similar to an isomap. Then all the sample data is represented by a weighted summation of its k nearest neighbors. Considering w_i to be row i of the n x k weight matrix w, the solution to the goal is found by:

$$W_i = \frac{1}{1^T(G_i^{-1}T)}G_i^{-1}1 \tag{15}$$

$$G := (x_i 1^T - V_i)^T(x_i 1^T - V_i) \tag{16}$$

where G is called a Gram matrix and V is a n x k matrix. After the process of representing samples as a weighted summation of their neighbors, LLE represents samples in the lower dimensional space by their neighbors with the same obtained weight. The method has been successfully used in feature extraction of Motor Imagery Electroencephalography (MI-EEG) and it outperformed methods such as Discrete Wavelet Transform (DWT) in classification accuracy with fewer feature dimension [120].

5) T-DisTribuTed STOCHASTIC NEIGHBOR EMBEDDING (T-SNE)

tSNE is an improvement of Stochastic Neighbor Embedding (SNE) [121], which is used for data visualisation. The main goal is to preserve the joint distribution of data samples in the original and embedding spaces. Considering P_{ij} and Q_{ij} to donate the probability that x_i and x_j and are neighbors and y_i and y_j are neighbors, it follows that:

$$p_{ij} = \frac{P_j|i + P_j|j}{2n} \tag{17}$$

$$p_{ij} = \frac{exp(-||x_i - x_j||_2^2/2\sigma_i^2)}{\sum_{K \neq 1} exp(-||x_i - x_k||_2^2/2\sigma_i^2)} \tag{18}$$

$$q_{ij} = (\frac{1 + ||y_i - y_j||_2^2)^{-1}}{\sum_{k \neq 1}(1 + ||y_k - y_l||_2^{2)^{-1}}} \tag{19}$$

Embedded samples are then obtained by adopting the gradient descent method over minimizing Keullback-Leibler divergence [122] of p and q distributions. The main advantage of t-SNE is the ability to deal with the problem of visualising "crowded" high dimensional data in a low dimensional space (e.g., 2D or 3D) [122], [123].

A. FEATURE EXTRACTION STATE OF THE ART

In image retrieval, calibration, classification, and clustering, it is critical to extract useful features or characteristics from the image [124]. Color histogram is the most significant method to represent color features [125]. [126] provided a state-of-the-art feature extraction model that consists of

two parts: (a) adaptive color region extraction via the definition circle (DC) model, and (b) corner feature extraction via the edge detection model, which includes a suppression mechanism.

The purpose of the algorithm was to produce a clear and precise forest saliency map. The algorithm is broken down into three parts, and those are: (a) the color feature extraction part; (b) the determination of the center of the DC model; and (c) an accurate description of color. The algorithm is expressed in figure 36.

(A) Colour feature extraction

Model appropriate for the extraction of color features is the DC model, which is comprised of the following steps: (1) using the RGB picture G histogram to calculate the DC model's center; (2) mapping the image to the HIS color space or lab color; (3) using the k-means procedure to find the DC model's radius. The flow chart of the DC model is shown in Figure 41.

(1) Determine the center of the of the DC model

While the DC model can describe color fluctuations under specific gradients, the forest region's dominating hue is generally green, implying that the 'greenish' pixels in the forest area must be filtered off. As a result, the G channel (green) in the RGB three-channel system will be the focal point for filtering out pixels that fall within a given range and calculating the mean value within the range. That value will be regarded as the center of the circle.

(2) Color description

It is critical to note that the purity of the green is determined by the circle's center, thus the radius must be adjusted to account for a variety of color variations and fault tolerance. The RGB channel, on the other hand, does not function well for color adjustments. The RGB color system is converted to Hue, Saturation, and Intensity (HSI) or lab color space to fix the problem. The color can be defined more correctly using only two channels, namely hue and saturation, rather than the RGB color space.

(3) Adjustment of DC Model radius

To improve the accuracy and adaptability of forest region extraction, the center and entire remote sensing picture acquired in the first phase is mapped or converted to HIS color space. Each pixel's Euclidean distance to the RSI center is calculated. The k-means clustering algorithm subdivide the forest into clusters and determines the Euclidean distance between the cluster center and the DC model's center, which is then used as the DC model's radius.

$$h = [h, s, i] \tag{20}$$

$$R = (h - h_0)^2 + (s - s_0)^2 + (i - i_0)^2 \tag{21}$$

$$\delta(i) = \sum_{k=1}^{k} |U_k^i - U_k^{i-1}| \tag{22}$$

P denotes the center of the DC model and the value would have been obtained by the histogram model in the RGB to HIS color scheme. R is the Euclidean distance and $\delta(i)$ represents
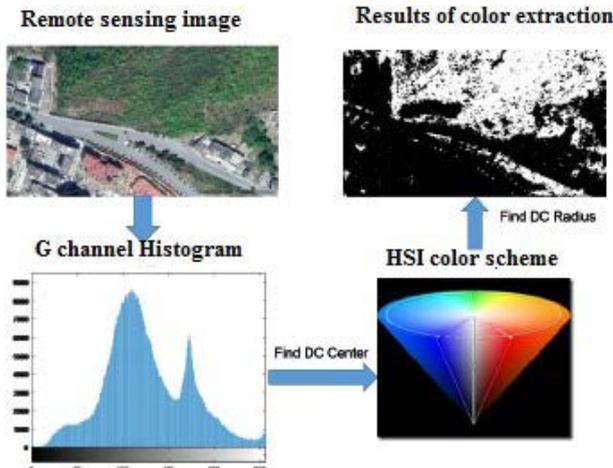
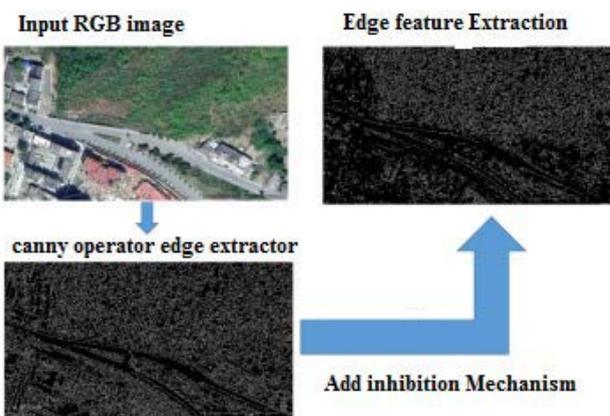**FIGURE 39.** DC model in color extraction feature [126].

FIGURE 40. Edge feature extractor.

an is the iterations of the class algorithm. Figure 39 shows the color extraction feature of the DC model

(B) Edge Feature extraction

The goal of this procedure is to successfully eliminate non-forest areas. [126] proposed the canny operator as the edge detection operator because of its better performance than other operators in terms of edge feature detection. In particular, denoising is key for image processing, and in this particular instance, a Gaussian filter was employed to smoothen the image, thereby preserving the edges. The amplitude and direction of the gradient are then calculated using the finite difference of the step-wise derivative. The canny edge detector operator returns only the maximum value and uses the non-maximum suppression operation to suppress the field's conspicuous points, resulting in a corner point with high precision and clear vision. Finally, by using a dual threshold setting, discrete edges are linked together to form a continuous edge. Figure 35 shows the stages of an edge feature extractor.

IX. PERFORMANCE EVALUATION MATRIX
The major matrices to measure the performance of the model in forest image classification are: False Positive Rate (FPR), Accuracy (Acc), F1 score, Precision-Recall Curve, and
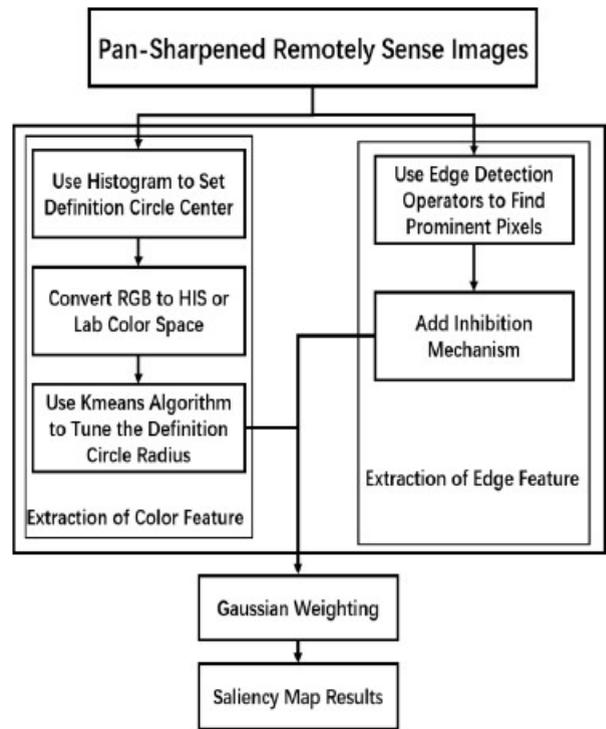
FIGURE 41. Flow diagram of the algorithm [126].

Average Precision (AP).

$$F1 = 2\frac{Precision * Recall}{Precision + Recall} \tag{23}$$

$$FRP = \frac{Numberofmisclassifiedforestimages}{Numberofimages} * 100\% \tag{24}$$

$$Acc = \frac{Numberofcorrectlyclassifiedimages}{Numberofimages} * 100\% \tag{25}$$

$$AP = areaundertheprecision\text{-}recallcurvePRC \tag{26}$$

Measurements for image segmentation area evaluation are presented in Table 2. The Area Fitness Index (AFI) was proposed by [63] and the remaining measurements by [2]

The average distance between the reference object and its matching image object is described by the Position Discrepancy Index (PDI). The Overall PDI is the average of the PDI.

$$let\ a = (x(k) - X_r)^2 + (Y(k) - Y_r)^2 \tag{27}$$

$$let\ b = \sum_{i=1}^{M} \sqrt{(X(I) - X_r)^2 + (Y(I) - Y_r)^2)} \tag{28}$$

$$PDI = \frac{1}{N + M}(\sum_{k=1}^{N} \sqrt{(a + b)} \tag{29}$$

$$PDI_{Overall} = \frac{1}{n}\sum_{i=n}^{n} PDI(i) \tag{30}$$

X. PERFORMANCE ANALYSIS OF THE STATE OF THE ART
Results based on the CNN with hyperparameter settings of patch size L = 15, regulation strength $\alpha = 0.001$, and

TABLE 2. Evaluation Matrix.

Evaluation Matrix		
Measurement	Definition	Description
Area Fitness Index(AFI)	$\frac{A_r - A_{LargestImageObject}}{A_r}$	When AFI$>$ 0,over segmentation, When AFI $<$ 0,under segmentation
$OE_{Overall}$	$\frac{\sum_{i=1}^{n}(OE(i)*A_r(i))}{\sum_{i=1}^{n} A_r(i)}$	The weighted average of OE
Omission error (OE)	$\frac{\sum_{k=1}^{nexp} A_e(k)-(A_e(k)\in A_r)}{A_r}$	Describes the under segmentation. An OE closer to zero means less under segmentation.
$CE_{Overall}$	$\frac{\sum_{i=1}^{n}(CE(i)*A_r(i))}{\sum_{i=1}^{n} A_r(i)}$	The weighted average of OE
Overall Area Discrepancy Index $ADI_{Overall}$	$\sqrt{OE_{Overall}^2 + CE_{Overall}^2}$	The overall of area discrepancy index of over and under segmentation. When ADI is zero, segmentation quality is optimal

TABLE 3. Segmentation results based on PDI and ADI.

Species	Species Code	Hyperspectral Precision	Recall	F-score	RGB Precision	Recall	F-Score
White fir	0	0.76	0.81	0.78	0.46	0.53	0.49
Red fir	1	0.76	0.72	0.74	0.41	0.30	0.35
Incense cedar	2	0.90	0.85	0.88	0.50	0.44	0.47
Jeffrey pine	3	0.93	0.96	0.95	0.65	0.73	0.69
Sugar pine	4	0.90	0.96	0.93	0.67	0.68	0.67
Black oak	5	0.73	0.61	0.67	0.69	0.61	0.65
Lodgepole pine	6	0.84	0.87	0.86	0.54	0.47	0.50
Dead	7	0.90	0.85	0.88	0.88	0.86	0.87
Ave/Total		0.87	0.87	0.87	0.64	0.64	0.64

TABLE 4. Edge feature extractor.

Shape	Compactness	Scale	OE$_{overall}$ (%)	CE$_{overall}$ (%)	ADI$_{overall}$ (%)	PDI$_{overall}$ (m)
0.1	0.1	60	6.34	5.25	8.22	5.06
0.1	0.1	70	7.75	4.99	9.22	4.83
0.1	0.1	80	8.09	5.47	9.77	4.73
0.1	0.1	90	8.99	6.83	11.30	4.71
0.1	0.1	100	9.02	8.27	12.23	5.08
0.1	0.1	110	10.25	8.84	13.54	4.77
0.1	0.1	120	11.46	8.68	14.37	4.71
0.1	0.3	60	5.22	4.55	6.93	5.48
0.1	0.3	70	6.01	4.45	7.48	4.89
...	...	...	...	...	...	...
0.3	0.7	60	4.11	4.05	5.77	3.81
0.3	0.7	70	4.75	4.28	6.39	4.06
0.3	0.7	80	5.43	4.26	6.90	4.17
...	...	...	...	...	...	...
0.9	0.9	60	22.41	12.71	25.76	5.25
0.9	0.9	70	24.02	16.75	29.28	5.41
0.9	0.9	80	31.88	14.78	35.14	5.68
0.9	0.9	90	36.01	15.14	39.07	6.88
0.9	0.9	100	40.85	12.74	42.80	7.65
0.9	0.9	110	45.56	10.32	46.71	8.23
0.9	0.9	120	48.05	9.00	48.89	9.33

TABLE 5. Performance of CNN.

Metrics	Deviation percentage interval	5% sampling proportion ^{1}T	^{2}CP (%)	10% sampling proportion ^{1}T	^{2}CP (%)	15% sampling proportion ^{1}T	^{2}CP (%)	20% sampling proportion ^{1}T	^{2}CP (%)	25% sampling proportion ^{1}T	^{2}CP (%)	30% sampling proportion ^{1}T	^{2}CP (%)
ADI$_{overall}$	<2%	16	32	23	46	29	58	37	74	39	78	46	92
	2–4%	11	54	13	72	9	76	13	100	11	100	4	100
	4–6%	9	72	10	92	11	98	0	100	0	100	0	100
	6–8%	6	84	2	96	1	100	0	100	0	100	0	100
	8–10%	5	94	2	100	0	100	0	100	0	100	0	100
	>10%	3	100	0	100	0	100	0	100	0	100	0	100
PDI$_{overall}$	<2%	13	26	17	34	27	54	33	66	41	82	44	88
	2–4%	14	54	14	62	8	70	17	100	9	100	6	100
	4–6%	5	64	6	74	9	88	0	100	0	100	0	100
	6–8%	9	82	8	90	3	94	0	100	0	100	0	100
	8–10%	8	98	5	100	3	100	0	100	0	100	0	100
	>10%	1	100	0	100	0	100	0	100	0	100	0	100

C = 32 filter kernels in the first convolutional layer up to a maximum of C' = 128 kernels. Using Tensorflow and Keras mechanisms, the final CNN classifiers used the hyperspectral imagery to outperform the RGB subset image as indicated by precision, recall, or F-score. Results are presented in Table 3.

Table 4 displays state-of-the-art segmentation results obtained using a supervised segmentation method and the following matrix measurements: AFI (index), OE, $OE_OVERALL$, CE, $CE_OVERALL$, ADI, PDI.

Object fate analysis and the method proposed by [63] do not objectively express segmentation quality results. Table 5 indicates that AFI ranges from 0.561 to −0.280 when shape and compactness are both at 0.1 and the scale parameter is changed from 60 to 120.

XI. RECOMMENDATION

Pixel-based techniques have been commonly used for image analysis and classification for a very long time. However, due to the massive growth of high spatial resolution images and the fact that pixel based methods only work with spectral information, the technique could not be fully utilized because it does not incorporate spatial, texture, and shape information, [127]. Previous studies have also shown that such approaches cause noise in the output message, otherwise known as the "salt and pepper effect." [128]. Due to the limitations of traditional pixel-based methods to cope with high-resolution imagery, OBIA methods have become increasingly popular because they have a high degree of information utilization, strong anti-interference, a high degree of data integration, and high classification accuracy [129], [130]. However, GEOBIA techniques are made up of knowledge and rules purely from domain expert knowledge, such that they enhance the subjectivity of image interpretation processes. Given the evolution of remote sensing science as a result of artificial intelligence, this study suggests that we pay more attention to Good Old-Fashioned Artificial Intelligence (GOFAI), which is based on sound mathematics and logic to construct symbolic representations of abstract notions [1]. This research highly recommends a shift towards remote sensing image analysis with ontologies because such technology allows management, aggregation, and sharing of the knowledge of remote sensing and domain experts. Formal ontologies explicitly define expert knowledge that is used to interpret remote sensing images. This improves the sharing and reuse of formalized remote sensing expert knowledge.

XII. CONCLUSION

This paper is a critical and analytical survey of the methods for forest image detection and classification. It is a comprehensive review of the techniques used to detect objects of interest in an image that will be analysed for classification of forests. These techniques cover semantic segmentation techniques, feature extraction methods and finally classification techniques. Exploration of knowledge based approaches in form of GEOBIA were analysed and how their shortcoming in terms of dual mode of defining geographic concept, vagueness and ambiguity of geographic concepts, and semantic gaps were addressed by ontology knowledge based approaches. Performance of the state of the art Tensorflow and Keras for image classification were analysed. Formal ontologies knowledge representation was recommended for state of the art approach for detecting objects of interest. CNN methods for semantic segmentation were critically analysed and these were; AlexNet, VGGNet, GoogLeNet, FCN, UNet, SegNet, DeepNet and ResNet.

REFERENCES

[1] D. Arvor, M. Belgiu, Z. Falomir, I. Mougenot, and L. Durieux, "Ontologies to interpret remote sensing images: Why do we need them?" *GIScience Remote Sens.*, vol. 56, no. 6, pp. 911–939, Aug. 2019.

[2] J. Cheng, Y. Bo, Y. Zhu, and X. Ji, "A novel method for assessing the segmentation quality of high-spatial resolution remote-sensing images," *Int. J. Remote Sens.*, vol. 35, no. 10, pp. 3816–3839, May 2014.

[3] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 177, pp. 11–28, Jul. 2016.

[4] R. M. Dufour, E. L. Miller, and N. P. Galatsanos, "Template matching based object recognition with unknown geometric parameters," *IEEE Trans. Image Process.*, vol. 11, no. 12, pp. 1385–1396, Dec. 2002.

[5] K. Bahareh, M. Shattri, S. Helmi, and H. Alfian, "Integration of template matching and object-based image analysis for semi-automatic oil palm tree counting in UAV images," in *Proc. 37th Asian Conf. Remote Sens. (ACRS)*, vol. 3, 2016, pp. 2333–2340.

[6] I. A. Aljarrah and A. S. Ghorab, "Object recognition system using template matching based on signature and principal component analysis," *Int. J. Digit. Inf. Wireless Commun.*, vol. 2, no. 2, pp. 156–163, 2012.

[7] I. Jordi, V. Aurthur, M. Arias, B. Tardy, D. Morin, and I. Rodes, "Operational high resolution land cover map production at the country scale using satellite image time series," *Remote Sens.*, vol. 9, no. 1, p. 95, 2017.

[8] M. Cristina, A. Picoli, G. Camara, I. Sanches, R. Simões, A. Carvalho, A. Maciel, A. Coutinho, J. Esquerdo, J. Antunes, R. Anzolin, D. Arvor, and C. Almeida, "Big earth observation time series analysis for monitoring Brazilian agriculture," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 328–339, Nov. 2018.

[9] M. Papadomanolaki, M. Vakalopoulou, S. Zagoruyko, and K. Karantzalos, "Benchmarking deep learning frameworks for the classification of very high resolution satellite multispectral data," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 7, pp. 83–88, Jun. 2016.

[10] J. Campbell, *Introduction to Remote Sensing*, 3rd ed. New York, NY, USA: Guilford Press, 2002.

[11] G. Marcus, "Deep learning: A critical appraisal," 2018, *1801.00631*.

[12] A. L. Ali, Z. Falomir, F. Schmid, and C. Freksa, "Rule-guided human classification of volunteered geographic information," *ISPRS J. Photogramm. Remote Sens.*, vol. 127, pp. 3–15, May 2017, doi: 10.1016/j.isprsjprs.2016.06.003.

[13] T. R. Martha, N. Kerle, C. J. V. Westen, V. Jetten, and K. V. Kumar, "Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4928–4943, Dec. 2011.

[14] D. Chaudhuri, N. K. Kushwaha, and A. Samal, "Semi-automated road detection from high resolution satellite images by directional morphological enhancement and segmentation techniques," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1538–1544, Oct. 2012.

[15] A. H. S. Solberg, "Contextual data fusion applied to forest map revision," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1234–1243, May 1999.

[16] A. Hanif, A. B. Mansoor, and A. S. Imran, *Performance Analysis of Vehicle Detection Techniques: A Concise Survey*, vol. 746. Cham, Switzerland: Springer, 2018, doi: 10.1007/978-3-319-77712-2_46.

[17] G. Chen, X. Zhang, Q. Wang, F. Dai, Y. Gong, and K. Zhu, "Symmetrical dense-shortcut deep fully convolutional networks for semantic segmentation of very-high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1633–1644, May 2018.

[18] R. Mohan and R. Nevatia, "Using perceptual organization to extract 3D structures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 11, pp. 1121–1139, Nov. 1989.

[19] H. G. Akcay and S. Aksoy, "Building detection using directional spatial constraints," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2010, pp. 1932–1935.

[20] J. Peng and Y. Liu, "Model and context-driven building extraction in dense urban aerial images," *Int. J. Remote Sens.*, vol. 26, no. 7, pp. 1289–1307, 2005.

[21] A. O. Ok, C. Senaras, and B. Yuksel, "Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1701–1717, Mar. 2013.

[22] Y.-T. Liow and T. Pavlidis, "Use of shadows for extracting buildings in aerial images," *Comput. Vis., Graph., Image Process.*, vol. 49, no. 2, pp. 242–277, Feb. 1990.

[23] X. Zhang, X. Feng, P. Xiao, G. He, and L. Zhu, "Segmentation quality evaluation using region-based precision and recall measures for remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 102, pp. 73–84, Apr. 2015.

[24] T. Su and S. Zhang, "Object-based crop classification in hetao plain using random forest," *Earth Sci. Informat.*, vol. 14, no. 1, pp. 119–131, Mar. 2021.

[25] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward objective evaluation of image segmentation algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 929–944, Jun. 2007.

[26] Y. Wang, Q. Qi, and Y. Liu, "Unsupervised segmentation evaluation using area-weighted variance and jeffries-matusita distance for remote sensing images," *Remote Sens.*, vol. 10, no. 8, p. 1193, Jul. 2018.

[27] H. Gao, Y. Tang, L. Jing, H. Li, and H. Ding, "A novel unsupervised segmentation quality evaluation method for remote sensing images," *Sensors*, vol. 17, no. 10, p. 2427, Oct. 2017.

[28] G. Modica, G. De Luca, G. Messina, and S. Praticò, "Comparison and assessment of different object-based classifications using machine learning algorithms and UAVs multispectral imagery: A case study in a citrus orchard and an onion crop," *Eur. J. Remote Sens.*, vol. 54, no. 1, pp. 431–460, Jan. 2021.

[29] G. J. Hay and G. Castilla, "Geographic object-based image analysis (GEOBIA): A new name for a new discipline," in *Object-Based Image Analysis*. Berlin, Germany: Springer, 2008, pp. 75–89.

[30] J. Krishnaswamy, M. C. Kiran, and K. N. Ganeshaiah, "Tree model based eco-climatic vegetation classification and fuzzy mapping in diverse tropical deciduous ecosystems using multi-season NDVI," *Int. J. Remote Sens.*, vol. 25, no. 6, pp. 1185–1205, Mar. 2004.

[31] K. J. Feeley, T. W. Gillespie, and J. W. Terborgh, "The utility of spectral indices from Landsat ETM+ for measuring the structure and composition of tropical dry forests," *Biotropica: J. Biol. Conservation*, vol. 37, no. 4, pp. 508–519, 2005.

[32] G. A. Sanchez-Azofeifa, K. L. Castro, B. Rivard, M. R. Kalascka, and R. C. Harriss, "Remote sensing research priorities in tropical dry forest environments," *Biotropica*, vol. 35, no. 2, pp. 134–142, Jun. 2003.

[33] S. Martinuzzi, W. A. Gould, O. M. Ramos González, A. Martínez Robles, P. Calle Maldonado, N. Pérez Buitrago, and J. J. Fumero Cabán, "Mapping tropical dry forest habitats integrating landsat NDVI, ikonos imagery, and topographic information in the Caribbean island of Mona," *Revista de Biología Tropical*, pp. 625–639, Nov. 2006.

[34] B. Bennett, "What is a forest? On the vagueness of certain geographic concepts," in *Proc. TOPOI*, 2002, pp. 2–17.

[35] B. Bennett, "Foundations for an ontology of environment and habitat," in *Proc. FOIS*, 2010, pp. 31–44.

[36] A. Mayamba, R. M. Byamungu, B. V. Broecke, H. Leirs, P. Hieronimo, A. Nakiyemba, M. Isabirye, D. Kifumba, D. N. Kimaro, M. E. Mdangi, and L. S. Mulungu, "Factors influencing the distribution and abundance of small rodent pest species in agricultural landscapes in eastern Uganda," *J. Vertebrate Biol.*, vol. 69, no. 2, Oct. 2020, Art. no. 020002.

[37] H. G. Lund, "When is a forest not a forest?" *J. Forestry*, vol. 100, no. 8, pp. 21–28, 2002.

[38] E. Romijn, J. H. Ainembabazi, A. Wijaya, M. Herold, A. Angelsen, L. Verchot, and D. Murdiyarso, "Exploring different forest definitions and their impact on developing REDD+ reference emission levels: A case study for Indonesia," *Environ. Sci. Policy*, vol. 33, pp. 246–259, Nov. 2013.

[39] C. E. Woodcock and A. H. Strahler, "The factor of scale in remote sensing," *Remote Sens. Environ.*, vol. 21, no. 3, pp. 311–332, Apr. 1987.

[40] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[41] C. Unger and P. Cimiano, "Pythia: Compositional meaning construction for ontology-based question answering on the semantic web," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Lecture Notes in Computer Science), vol. 6716. Germany: Bielefeld Univ., 2011, pp. 153–160.

[42] R. H. Kilmann and K. W. Thomas, "Developing a forced-choice measure of conflict-handling behavior: The 'MODE' instrument," *Educ. Psychol. Meas.*, vol. 37, no. 2, pp. 309–325, 1977.

[43] B. Bachimont, A. Isaac, and R. Troncy, "Semantic commitment for designing ontologies: A proposal," in *Proc. Int. Conf. Knowl. Eng. Knowl. Manage.* Berlin, Germany: Springer, 2002, pp. 114–121.

[44] E. F. Fama and M. C. Jensen, "Separation of ownership and control," *J. Law Econ.*, vol. 26, no. 2, pp. 301–325, 1983.

[45] K. Satoh, *Nonmonotonic Reasoning by Minimal Belief Revision*. Tokyo, Japan: ICOT Research Center (Institute for New Generation Computer Technology), 1988.

[46] T. Gruber, "What is an ontology," Stanford Univ., Stanford, CA, USA, Tech. Rep. KSL92-71, 1993.

[47] S. Andrés, D. Arvor, I. Mougenot, T. Libourel, and L. Durieux, "Ontology-based classification of remote sensing images using spectral rules," *Comput. Geosci.*, vol. 102, pp. 158–166, May 2017.

[48] D. Mallenby, "Handling vagueness in ontologies of geographical information," Ph.D. dissertation, School Comput., Univ. Leeds, Leeds, U.K., 2008. [Online]. Available: http://etheses.whiterose.ac.U.K./1373/

[49] N. Eric Maillot and M. Thonnat, "Ontology based complex object recognition," *Image Vis. Comput.*, vol. 26, no. 1, pp. 102–113, Jan. 2008.

[50] C. Eschenbach and M. Grüninger, "Formal ontology in information systems," in *Proc. 5th Int. Conf. (FOIS)*, vol. 110, 2008, pp. 68–71.

[51] M. Davis, S. King, N. Good, and R. Sarvas, "From context to content: Leveraging context to infer media metadata," in *Proc. 12th Annu. ACM Int. Conf. Multimedia*, 2004, pp. 188–195.

[52] F. Nack, C. Dorai, and S. Venkatesh, "Computational media aesthetics: Finding meaning beautiful," *IEEE MultimediaMag.*, vol. 8, no. 4, pp. 10–12, Oct. 2001.

[53] H. Gu, H. Li, L. Yan, Z. Liu, T. Blaschke, and U. Soergel, "An object-based semantic classification method for high resolution remote sensing imagery using ontology," *Remote Sens.*, vol. 9, no. 4, p. 329, 2017.

[54] A. Baraldi, V. Puzzolo, P. Blonda, L. Bruzzone, and C. Tarantino, "Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2563–2586, Sep. 2006.

[55] A. M. Arifjanov, S. B. Akmalov, T. U. Apakhodjaeva, and D. S. Tojikhodjaeva, "Comparison of pixel to pixel and object based image analysis using worldview-2 satellite images of vangiobod village of syndria province," *Remote Methods Earth Res.* vol. 26, no. 2, pp. 313–321, 2020.

[56] N. Durand, S. Derivaux, G. Forestier, C. Wemmert, P. Gançarski, O. Boussaid, and A. Puissant, "Ontology-based object recognition for remote sensing image interpretation," in *Proc. 19th IEEE Int. Conf. Tools Artif. Intell. (ICTAI)*, vol. 1, Oct. 2007, pp. 472–479.

[57] S. R. Phinn, C. M. Roelfsema, and P. J. Mumby, "Multi-scale, object-based image analysis for mapping geomorphic and ecological zones on coral reefs," *Int. J. Remote Sens.*, vol. 33, no. 12, pp. 3768–3797, Jun. 2012.

[58] B. Gajderowicz, "Using decision trees for inductively driven semantic integration and ontology matching," M.S. thesis, Dept. Comput. Sci., Ryerson Univ., Toronto, ON, Canada, 2011.

[59] B. Gajderowicz and A. Sadeghian, "Ontology granulation through inductive decision trees," in *Proc. URSW*, 2009, pp. 39–50.

[60] N. Kartha and A. Novstrup, "Ontology and rule based knowledge representation for situation management and decision support," *Proc. SPIE*, vol. 7352, May 2009, Art. no. 73520P.

[61] J. C. Giarratano and G. D. Riley, *Expert Systems: Principles and Programming*. Pacific Grove, CA, USA: Brooks/Cole, 2005.

[62] D. A. Waterman, D. B. Lenat, and F. Hayes-Roth, *Building Expert Systems*. Reading, MA, USA: Addison-Wesley, 1983.

[63] P. He, "Counter cyber attacks by semantic networks," in *Emerging Trends in ICT Security*. Amsterdam, The Netherlands: Elsevier, 2014, pp. 455–467.

[64] G. De Giacomo, D. Lembo, M. Lenzerini, A. Poggi, and R. Rosati, "Using ontologies for semantic data integration," in *A Comprehensive Guide Through Italian Database Res. Over Last 25 Years*. Cham, Switzerland: Springer, 2018, pp. 187–202.

[65] F. Baader, *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[66] B. C. Grau, I. Horrocks, B. Motik, B. Parsia, P. Patel-Schneider, and U. Sattler, "OWL 2: The next step for OWL," *J. Web Semantics*, vol. 6, no. 4, pp. 309–322, Nov. 2008.

[67] B. C. Grau, I. Horrocks, Y. Kazakov, and U. Sattler, "A logical framework for modularity of ontologies," in *Proc. IJCAI*, 2007, pp. 298–303.

[68] S. Ghilardi, C. Lutz, and F. Wolter, "Did I damage my ontology," in *Proc. KR*, 2006, pp. 187–197.

[69] S. Roy and I. J. Cox, "A maximum-flow formulation of the N-camera stereo correspondence problem," in *Proc. 6th Int. Conf. Comput. Vis.*, Jan. 1998, pp. 492–499.

[70] R. Geerken, B. Zaitchik, and J. P. Evans, "Classifying rangeland vegetation type and coverage from NDVI time series using Fourier filtered cycle similarity," *Int. J. Remote Sens.*, vol. 26, no. 24, pp. 5535–5554, Dec. 2005.

[71] Y. Guo, S. Han, Y. Li, C. Zhang, and Y. Bai, "K-nearest neighbor combined with guided filter for hyperspectral image classification," *Proc. Comput. Sci.*, vol. 129, pp. 159–165, Jan. 2018.

[72] G. De Luca, J. M. N. Silva, S. Cerasoli, J. Araújo, J. Campos, S. Di Fazio, and G. Modica, "Object-based land cover classification of cork oak woodlands using UAV imagery and orfeo toolbox," *Remote Sens.*, vol. 11, no. 10, p. 1238, May 2019.

[73] A. D. P. Pacheco, J. A. D. S. Junior, A. M. Ruiz-Armenteros, and R. F. F. Henriques, "Assessment of K-nearest neighbor and random forest classifiers for mapping forest fire areas in central Portugal using Landsat-8, Sentinel-2, and Terra imagery," *Remote Sens.*, vol. 13, no. 7, p. 1345, Apr. 2021.

[74] P. T. Noi and M. Kappas, "Comparison of random forest, K-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery," *Sensors*, vol. 18, no. 1, p. 18, 2018.

[75] E. Tomppo, M. Haakana, M. Katila, and J. Peräsaari, *Multi-Source National Forest Inventory: Methods and Applications*, vol. 18. Springer, 2008.

[76] L. Tlig, M. Bouchouicha, M. Tlig, M. Sayadi, and E. Moreau, "A fast segmentation method for fire forest images based on multiscale transform and PCA," *Sensors*, vol. 20, no. 22, p. 6429, Nov. 2020.

[77] S. M. De Jong and F. D. Van der Meer, *Remote Sensing Image Analysis: Including the Spatial Domain*, vol. 5. Springer, 2007.

[78] D. Kaur and Y. Kaur, "Various image segmentation techniques: A review," *Int. J. Comput. Sci. Mobile Comput.*, vol. 3, no. 5, pp. 809–814, 2014.

[79] Y.-J. Zhang, "An overview of image and video segmentation in the last 40 years," in *Advances in Image and Video Segmentation*. Dordrecht, The Netherlands: 2006, pp. 1–16.

[80] T. Lindeberg and M.-X. Li, "Segmentation and classification of edges using minimum description length approximation and complementary junction cues," *Comput. Vis. Image Understand.*, vol. 67, no. 1, pp. 88–98, Jul. 1997.

[81] S. Yuheng and Y. Hao, "Image segmentation algorithms overview," 2017, *arXiv:1707.02051*.

[82] N. Senthilkumaran and R. Rajesh, "Image segmentation—A survey of soft computing approaches," in *Proc. Int. Conf. Adv. Recent Technol. Commun. Comput.* Stockholm, Sweden: KTH (Roy. Inst. Technol.), Oct. 2009, pp. 844–846.

[83] M. K. Kundu and S. K. Pal, "Thresholding for edge detection using human psychovisual phenomena," *Pattern Recognit. Lett.*, vol. 4, no. 6, pp. 433–441, 1986.

[84] M. R. Khokher, A. Ghafoor, and A. M. Siddiqui, "Image segmentation using multilevel graph cuts and graph development using fuzzy rule-based system," *IET Image Process.*, vol. 7, no. 3, pp. 201–211, 2013.

[85] T. Blaschke, C. Burnett, and A. Pekkarinen, "Image segmentation methods for object-based analysis and classification," in *Remote Sensing Image Analysis: Including The Spatial Domain*. Dordrecht, The Netherlands: Springer, 2004, pp. 211–236.

[86] T. Lei, X. Jia, Y. Zhang, S. Liu, H. Meng, and A. K. Nandi, "Superpixel-based fast fuzzy C-means clustering for color image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 9, pp. 1753–1766, Sep. 2019.

[87] P. Neubert and P. Protzel, "Compact watershed and preemptive SLIC: On improving trade-offs of superpixel segmentation algorithms," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 996–1001.

[88] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," *Expert Syst. Appl.*, vol. 169, May 2021, Art. no. 114417.

[89] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1097–1105.

[90] B. Liu, X. Yu, A. Yu, and G. Wan, "Deep convolutional recurrent neural network with transfer learning for hyperspectral image classification," *J. Appl. Remote Sens.*, vol. 12, no. 2, 2018, Art. no. 026028.

[91] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2414–2423.

[92] O. A. B. Penatti, K. Nogueira, and J. A. dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 44–51.

[93] M. Volpi and V. Ferrari, "Semantic segmentation of urban scenes by learning local class interactions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 1–9.

[94] M. Lin and Q. Yan, "Network in network," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2014, pp. 1–4.

[95] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, 2016.

[96] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2015, pp. 234–241.

[97] J. E. Ball, D. T. Anderson, and C. S. Chan, "Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community," *J. Appl. Remote Sens.*, vol. 11, no. 4, 2017, Art. no. 042609.

[98] S. Chilamkurthy. (2017). *A 2017 Guide to Semantic Segmentation with Deep Learning*. [Online]. Available: https://blog.qure.ai/notes/semantic-segmentation-deep-learning-review

[99] J. Le. (2017). *How to do Semantic Segmentation Using Deep Learning*. [Online]. Available: https://nanonets.com/blog/how-to-do-semantic-segmentation-using-deep-learning/

[100] A. Mittal. (2019). *Introduction to U-Net and Res-Net for Image Segmentation*. [Online]. Available: https://aditi-mittal.medium.com/introduction-to-u-net-and-res-net-for-image-segmentation-9afcb432ee2f

[101] S. Kentsch, M. L. Lopez Caceres, D. Serrano, F. Roure, and Y. Diez, "Computer vision and deep learning techniques for the analysis of drone-acquired forest images, a transfer learning study," *Remote Sens.*, vol. 12, no. 8, p. 1287, Apr. 2020.

[102] M. Šulc, D. Mishkin, and J. Matas, "Very deep residual networks with maxout for plant identification in the wild," in *Proc. Work. Notes CLEF*, 2016, pp. 1–8.

[103] M. Onishi and T. Ise, "Automatic classification of trees using a UAV onboard camera and deep learning," 2018, *arXiv:1804.10390*.

[104] S. Natesan, C. Armenakis, and U. Vepakomma, "ResNet-based tree species classification using UAV images," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 475–481, Jun. 2019.

[105] M. Dyrmann, H. Karstoft, and H. S. Midtiby, "Plant species classification using deep convolutional neural network," *Biosyst. Eng.*, vol. 151, pp. 72–80, Nov. 2016.

[106] M. Onishi and T. Ise, "Automatic classification of trees using a UAV onboard camera and deep learning," 2018, *arXiv:1804.10390*.

[107] O. Brovkina, E. Cienciala, P. Surový, and P. Janata, "Unmanned aerial vehicles (UAV) for assessment of qualitative classification of Norway spruce in temperate forest stands," *Geo-spatial Inf. Sci.*, vol. 21, no. 1, pp. 12–20, Jan. 2018.

[108] M. J. Zimmer-Gembeck and M. Helfand, "Ten years of longitudinal research on U.S. adolescent sexual behavior: Developmental correlates of sexual intercourse, and the importance of age, gender and ethnic background," *Developmental Rev.*, vol. 28, no. 2, pp. 153–224, 2008.

[109] S. Watanabe, K. Sumi, and T. Ise, "Identifying the vegetation type in Google Earth images using a convolutional neural network: A case study for Japanese bamboo forests," *BMC Ecol.*, vol. 20, no. 1, pp. 1–14, Dec. 2020.

[110] E. Guirado, S. Tabik, D. Alcaraz-Segura, J. Cabello, and F. Herrera, "Deep-learning versus OBIA for scattered shrub detection with Google Earth imagery: Ziziphus lotus as case study," *Remote Sens.*, vol. 9, no. 12, p. 1220, Nov. 2017.

[111] P. P. Ippolito. (2019). *Feature Extraction Techniques*. Accessed: Apr. 29, 2020. [Online]. Available: https://towardsdatascience.com/feature-extraction-techniques-d619b56e31be

[112] A. Ghodsi, "Dimensionality reduction a short tutorial," Ph.D. dissertation, Dept. Statist. Actuarial Sci., Univ. Waterloo, Waterloo, ON, Canada, 2006, vol. 37, no. 38.

[113] B. Ghojogh, M. N. Samad, S. Asif Mashhadi, T. Kapoor, W. Ali, F. Karray, and M. Crowley, "Feature selection and feature extraction in pattern analysis: A literature review," 2019, *arXiv:1905.02845*.

[114] C. Citro, "rules. In Proc. 20th Int. Conf. very large data bases, VLDB, volume 1215, pages 487–499, 1994.[5] Alfred V. Aho, Ravi Sethi and Jeffrey D. Ullman. Compilers: Principles, techniques, and tools. Boston, MA: Addison-Wesley. 1986.[6] Adrian Akmajian, Ann K. Farmer, Lee Bickmore, Richard A. Demers and," *Learning*, vol. 5, no. 1, pp. 71–99, 1990.

[115] C. A. Brooks and K. Iagnemma, "Vibration-based Terrain classification for planetary exploration rovers," *IEEE Trans. Robot.*, vol. 21, no. 6, pp. 1185–1191, Dec. 2005.

[116] F. Subhan, S. Saleem, H. Bari, W. Z. Khan, S. Hakak, S. Ahmad, and A. M. El-Sherbeeny, "Linear discriminant analysis-based dynamic indoor localization using Bluetooth low energy (BLE)," *Sustainability*, vol. 12, no. 24, p. 10627, Dec. 2020.

[117] Y. Mo, Z. Zhang, Y. Lu, W. Meng, and G. Agha, "Random forest based coarse locating and KPCA feature extraction for indoor positioning system," *Math. Problems Eng.*, vol. 2014, Oct. 2014, Art. no. 850926.

[118] L. C. Marsh and D. R. Cormier, *Spline Regression Models*, no. 137. Newbury Park, CA, USA: Sage, 2001.

[119] L. K. Saul and S. T. Roweis, "Think globally, fit locally: Unsupervised learning of low dimensional manifolds," *J. Machine Learn. Res.*, vol. 4, pp. 119–155, Jun. 2003.

[120] M. Li, X. Luo, J. Yang, and Y. Sun, "Applying a locally linear embedding algorithm for feature extraction and visualization of MI-EEG," *J. Sensors*, vol. 2016, Aug. 2016, Art. no. 7481946.

[121] G. Hinton and S. T. Roweis, "Stochastic neighbor embedding," in *Proc. NIPS*, vol. 15, 2002, pp. 833–840.

[122] S. Kullback, *Information Theory and Statistics*. Chelmsford, MA, USA: Courier Corporation, 1997.

[123] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne," *J. Mach. Learn. Res.*, vol. 9, no. 11, 2008.

[124] D. Gu, Z. Han, and Q. Wu, "Feature extraction to polar image," *J. Comput. Commun.*, vol. 5, no. 11, pp. 16–26, 2017.

[125] F. Alamdar and M. Keyvanpour, "A new color feature extraction method based on QuadHistogram," *Proc. Environ. Sci.*, vol. 10, pp. 777–783, Jan. 2011.

[126] H. Du and Y. Zhuang, "Optical remote sensing images feature extraction of forest regions," in *Proc. IEEE Int. Conf. Signal, Inf. Data Process. (ICSIDP)*, Dec. 2019, pp. 1–5.

[127] H. Luo, L. Li, H. Zhu, X. Kuai, Z. Zhang, and Y. Liu, "Land cover extraction from high resolution ZY-3 satellite imagery using ontology-based method," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 3, p. 31, 2016.

[128] O. Oke Alice, O. Omidiora Elijah, A. Fakolujo Olaosebikan, S. Falohun Adeleye, and S. Olabiyisi, "Effect of modified Wiener algorithm on noise models," *Int. J. Eng. Technol.*, vol. 2, no. 8, pp. 1024–1033, 2012.

[129] G. Hay and G. Castilla, "Object-based image analysis: Strengths, weaknesses, opportunities and threats (SWOT)," in *Proc. 1st Int. Conf. (OBIA)*, 2006, pp. 4–5.

[130] D. C. Duro, S. E. Franklin, and M. G. Dubé, "A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery," *Remote Sens. Environ.*, vol. 118, pp. 259–272, Mar. 2012.

CLOPAS KWENDA received the B.Sc. degree (Hons.) in computer science from the Bindura University of Science Education (BUSE), Zimbabwe, and the M.Sc. degree in computer science from the University of Zimbabwe (UZ), Zimbabwe. He is currently pursuing the Ph.D. degree with the University of KwaZulu-Natal (UKZN), South Africa. His research interests include image processing, artificial intelligence, machine learning, deep learning, and ontology building.

MANDLENKOSI GWETU received the Ph.D. degree in computer science (CS), specializing in medical image processing, from University of KwaZulu-Natal (UKZN), South Africa. He is a Senior Lecturer with UKZN. He is currently the Academic Leader of CS with UKZN. He is the Principal Investigator of the UKZN node in the Erasmus+ funded the Living Laboratories for Climate Change Multi-National Project and is an Alumni of the Heidelberg Laureate Forum. His research interests include deep learning, pattern recognition, and computer vision.

JEAN VINCENT FONOU DOMBEU received the B.Sc. degree (Hons.) in computer science from the University of Yaoundé I, Cameroonis, the M.Sc. degree in computer science from the University of KwaZulu-Natal, South Africa, and the Ph.D. degree in computer science from the North-West University, South Africa. He is a Senior Lecturer with the Department of Computer Science, University of KwaZulu-Natal (UKZN). His research interests include ontology engineering, semantic web, and machine learning—specifically, in ontology building, learning, modularization, ranking, summarization and visualization, artificial intelligence, machine learning and data mining methods for the semantic web, knowledge representation and reasoning on the web, and knowledge graphs and deep semantics.

● ● ●