A SYSTEMATIC LITERATURE REVIEW ON INSECT DETECTION IN IMAGES

Lotfi Souifi[1], Afef Mdhaffar[1,2], Ismael Bouassida Rodriguez[1],
Mohamed Jmaiel[1,2] and Bernd Freisleben[3]

[1]ReDCAD Laboratory, ENIS, University of Sfax, B.P. 1173 Sfax, Tunisia
[2]Digital Research Center of Sfax, 3021 Sfax, Tunisia
[3]Dept. of Math. & Comp. Sci., Philipps-Universität Marburg, Germany

ABSTRACT

Due to the advancements of deep learning (DL), particularly in the areas of visual object detection and convolutional neural networks (CNN), insect detecetion in images has received a lot of attention from the research community in the last few years. This paper presents a systematic review of the literature on the topic of insect detection as a case of object detection in images. It covers 50 research papers on the subject and responds to three research questions: i) type of dataset used; ii) detection technique used; iii) insect location. The paper also provides a summary of existing methods used for insect detection.

KEYWORDS

Systematic Literature Review (SLR), Deep Learning (DL), Object Detection, Insect Detection.

1. INTRODUCTION

Agriculture is the first human activity that enabled mankind to progress and develop. Agriculture and the food industry are the most important activities in the world today, owing to the world's growing population and its increasing need for food [1], [2]. Insects have long been regarded as a serious crop threat. Insects have the primary effect of reducing the amount of food available to people by lowering agricultural productivity. This might decrease the quantity and the quality of food [3]. For agricultural pest forecasting, insect detection is critical. Agricultural professionals detect insect infestations based on daily observations. This manual process costs farmers a lot of time and money. To reduce the use of risky and expensive chemical products, early detection and monitoring of insects are necessary for taking the proper action and determining whether the insects are dangerous or not. As a result of the advancement of deep learning (DL), particularly in image processing, several methods for insect detection have been proposed. In the field of" insect detection", there is only one review of the semantic literature. Amarathunga et al. [4] wrote this SLR in 2021, where 2021 publications are not considered. The authors [4] mention that the publication window is only between 2010 and 2020. The article clarifies classification methods but does not discuss insect detection methods. Therefore, in addition to using the most recent detection and classification techniques, our paper also gives more attention to the gathering and preparation of the used dataset.

2. SYSTEMATIC LITERATURE REVIEW

We perform a systematic literature review following five steps:

David C. Wyld et al. (Eds): SIGEM, MLTEC, SEAPP, ITCON, NATL, FUZZY, CSEA - 2022
pp. 41-56, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.122003

- Define research questions,
- Identify papers that are related to this topic,
- Add papers to the list containing titles and abstracts,
- Remove non-relevant papers,
- Identify the capabilities of each studied paper to answer the research questions.

2.1. Research Questions

The research questions that need to be answered by our study are as follows:

- RQ1: What type of datasets of insects are used?

  This question focuses on the type of datasets used for insect detection (i.e., collected for the purpose of this paper or publicly available)
- RQ2: Which object detection method is used for insect detection?

  A response to this question presents the commonly used technique for insect detection and the modification added to it. We notice that in some cases, the studied article presents a combination of detection and classification of insects.
- RQ3: What is the location of the studied insect?

  RQ3's response typically includes a description of the study's location. It may not be mentioned in every instance. In this case, we consider the first author's research institute or university's location to be an insect's location.

2.2. Methodology

Our search methodology follows two phases. First, a group of keywords is defined based on the research questions. Second, the selected keywords are aggregated using AND and OR operators to formulate the results below:

("deep learning" AND "insect detection" AND "object detection" AND "insect detection" AND ("insect detection" OR "insect pest detection")).

2.3. Selection Criteria

This section presents the used selection criteria.

- Inclusion Criteria (IC):
  - publications that match one of the search items,
  - research studies from journals,
  - conferences studies that were published from January 2012 to March 2022.
- Exclusion Criteria (EC):
  - publications that were published before (or on) 31.12.2011,
  - publications that are not related to the research questions (i.e., insect detection), but appeared in the search,
  - language (only English is taken).

2.4. Data Collection

Figure 1 shows our search methodology. It follows three main steps. The first one (i.e., illustrated by the left side) describes the research findings for each of the search engines that were used. As a result, 407 articles were selected for our study. The 5 triangles illustrate the second step. They describe the elimination criteria and the number of articles that have been removed. The final step (i.e., illustrated by the right side) displays the 50 articles that met the exclusion criteria in total, along with how they were distributed among the search engines.

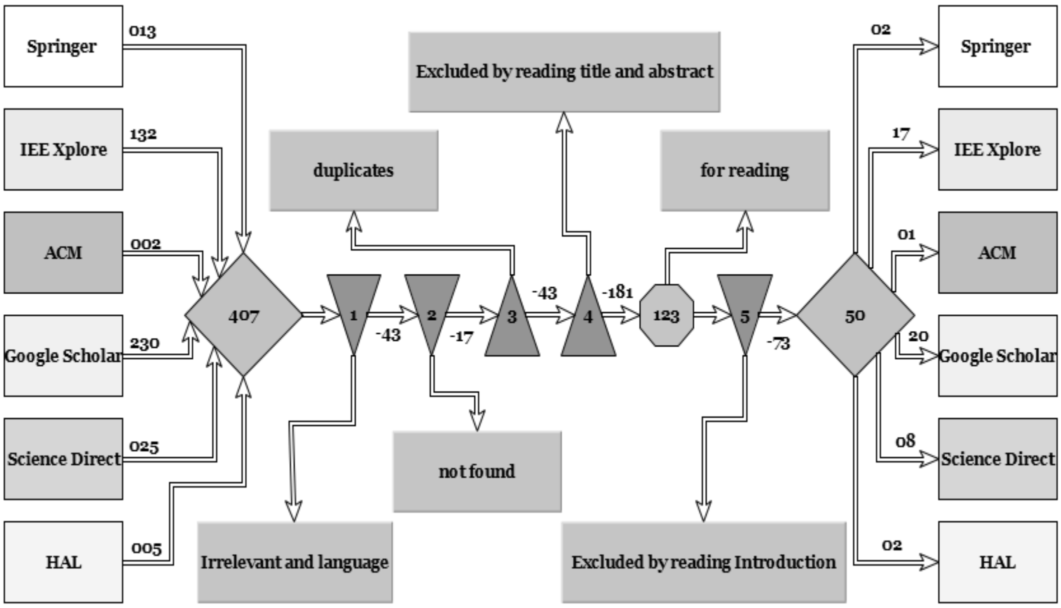

Figure 1. Data collection

3. RESULTS AND DISCUSSION

Our paper answers three different questions with respect to insect detection. Before we start discussing the questions, it is important to understand some basic concepts such as insect and insect pest.

3.1. Basic Concepts

Hill [5] defines pests as follows: any animal (or plant) that causes harm or damage to humans. Even if they are just causing annoyances, such animals or crops qualify as pests. Even if it does not belong to a pest species, an animal or plant taken out of the context is considered a pest (individually) in agricultural terms. The same author [6] later defines a pest as follows: in the broadest sense, a pest is an insect (or organism) that causes harm to humans, livestock, crops, or personal property. The key word is harm, which is usually translated as damage, which can be measured (often quantitatively) in many cases. Furthermore, damage is frequently equated to monetary losses. Nuisance and disturbance are examples of harm at the most basic level of interpretation. Thus, a buzzing mosquito at night can keep you awake, and face flies in tropical Africa can be very distracting and reduce your productivity. Insects are referred to as pests in our study.

3.2. RQ1: Type of Data Sets of Insects

Since the data set is very important for deep learning models, it is common to discuss the used data before the used model.

3.2.1. Web Sources

Articles follow this approach [7]– [10], and each of the previous articles uses different methods. However, the fact that they all focus on well-known insects is what unites them (e.g., spiders, mosquitoes, etc). The most well-known collection methods involve the use of search engines like Google [11] and Bing [12] or the use of websites like Flickr [13] as the standard gathering techniques. While Butera et al. [7] employ all three techniques, other authors [8], [9] use Google and Flickr, and some authors [10] do not mention the techniques employed.

Using web sources is relatively fast, but not suitable for all insects. Some insects may not be found (not the same name in all countries) or lack photos of the target insect. Combining web sources and on-site images could be the answer to this issue.

3.2.2. Combination of Web Sources and On-Site Images

Combining web sources and on-site images could be more effective. When the number of photographs is insufficient, this approach requires more time and resources than using web sources only [14]– [16]. According to Takimoto et al. [14], field collection took two years (2017 and 2018). While Hossain et al. [15] and Abeywardhana et al. [16] do not specify how much time has passed, Takimoto et al. [14] mentioned the use of a RICOH WG-4 digital camera alongside the Google search engine.

However, Hossain et al. [15] utilized a mobile phone to shoot images in the field along with the Google search engine. The web source is not mentioned by Abeywardhana et al. [16], although a digital camera and a mobile phone were used.

3.2.3. On-Site Images (Field or Laboratory)

Some researchers collect data in the field, even though it is the hardest method compared to the other approaches. Different techniques were used to collect data, such as different types of traps, including yellow traps [17], [18], pheromone traps [19]– [21], yellow boards [22], and smart traps [23], [24] with an integrated camera. Other researchers use cameras and mobile phones to take photos of insects directly in the laboratory or field. In Lyu et al. [25], the pictures of the desired insects were taken in the laboratory. While Yang et al. [26] used Zhongwei Kechuang industrial cameras installed in field, other authors [27]– [29] used digital mobile cameras. Currently, Du et al. [30] use high-resolution UAV, and Bjerge et al. [31] use previously constructed portable computer versions, while Ard et al. [32] use a Scoutbox along with a camera, and Rustia et al. [33] use a wireless image monitoring system. It took them two years and two months to collect all their data. Finally, Chen et al. [34] use a small smart car.

Collecting on-site images is still the most used approach for two reasons. First, a few open datasets are suitable for multi-class pest detection, and there is a lack of datasets that is suitable for a particular study goal. Menikdiwela et al. [9] explain why it is not a good idea to use an existing dataset. Because the spider’s scale is quite large in comparison to the background in those images, they do not use any of the spider images from the ImageNet dataset. Second, the trained model will be more robust if the dataset is as close as possible to a real scene in the natural environment.

3.2.4. Existing Image Datasets

The simplest method is to just take existing data, which in several cases is pre-processed. There are different types of existing data sets:

- Data collected by the farmer but not processed, e.g., Nam et al. [35],
- The use of well-known datasets, e.g., IP 102 proposed by Wu et al. [36], which was used by several authors [37]– [39], and the 10c and 5c datasets, which were used by Liu et al. [40],
- Datasets gathered by research institutions, such as the Moth Classification and Counting (MCC) dataset and European Moths dataset (EU-Moths), which were used by Korsch et al. [41], a dataset collected by the Laboratory of Medical Zoology, which was used by Luo et al. [42], and a dataset collected by the Hefei Institute of Physical Science (under the name of pest24), which was used by Liu et al. [43],
- datasets gathered by other researchers, e.g., Mamdouh et al. [44] used the Dacus Image Recognition Toolkit (DIRT) data set collected by Kalamatianos et al. [45], Tanjim et al. [46] used datasets gathered by Rajan et al. [47], and Deserno et al. [48] used the yellow stick trap data set gathered by Ard et al. [49],
- the Kaggle contest data set, used by SutHo et al. [50].

Since all images have already been processed, this is the most straightforward approach. Nevertheless, it is still insufficient for all use cases and cannot contain all insect species.

3.2.5. Combination of Existing Image Data Sets and Other Methods

Some articles combine more than one approach, such as Cabrera et al. [51] who use an existing data set called Yellow Sticky Trap data set along with a web data set. Also, Mazare et al. [52] use existing data photo libraries taken by the beneficiary institute and collected from traps. Xia et al. [53] utilize data already collected by Xie et al. [54] and data downloaded from the Internet. Huang et al. [55] use a combination of three methods collected from an e-trap and data from the Internet from iNaturalist [56], and an existing data set IP102 [36] and ImageNet (proposed by Deng et al. [57]).

There are two articles [58], [59] that present recent datasets that professionals have gathered and annotated. Wang et al. [58] present AgriPest, with a total of 49.000 images and a total of 264.000 objects. This dataset was collected in the field and labelled by 20 agricultural experts. Li et al. [59] mention that their dataset was collected in a warehouse under the name of RGBInsect, and they used smartphones with traps. The labelling was done by experts. RGBInsect has a total of 7.514 images and more than 159.000 insect instances.

Table 1. The percentage of each used type of dataset in all articles

Dataset type	**(%)**	**Data set sub-classes**	**(%)**
Collected datasets	56	Web source	8
		Combination of web sources and on-site images	8
		On-site (field or laboratory)	40
Existing datasets	44	Existing datasets	30
		Combination of existing data sets and other methods	14

Before discussing the next research question, some points need to be clarified such as data labelling techniques and data augmentation.

After collecting a data set, all images must be labelled. This task should be conducted by experts or some automated tools. Our paper describes the used labelling tools and the human-based labelling methods.

In the studied articles, only 11 of them mentioned the tool used to label the images. Nam et al. [35] mentioned that they used BBOX [60] while other authors [10], [14], [17], [25], [28], [33], [44] used LabelImg [61] and further authors [27], [30], [50] used LabelMe [62]. Some articles [35], [42], [44] mention that the labelling is done by experts, and Rustia et al. [33] state that labelling was done by experts and entomologists, while Ding et al. [19] mention technicians.

Data augmentation techniques (DA) are used in 54% of the studied papers. Insufficient data for model training and testing is the cause of DA, which can result in over-fitting (which happens when a statistical model matches its training data close to perfection. Unfortunately, when this occurs, the algorithm's goal is defeated because it cannot accurately perform against unobserved data.) or under-fitting (i.e., a data model has a high error rate on both the training set and unobserved data because it cannot accurately represent the relationship between the input and output variables. It happens when a model is overly simplistic, i.e., when there is insufficient regularization, training time, or input features). Rotation and zooming are the most used techniques for DA in general. Only Liu et al. [40] did not specify which method they employed. Rotation is used by Ding et al. [19], whereas reflection is used by Khalifa et al. [37]. Some articles use more than the first two, such as adding noise [8], scaling [10], and flipping and colour adjustments [18, 34].

3.3. RQ2: Object Detection Method Used for Insect Detection

The discussion of the detection methods presented in the chosen articles will now proceed, but first, let's define the widely employed method.

3.3.1. Convolutional Neural Networks CNNs

Before discussing the employed methods, it is necessary to first review some fundamentals of deep learning.

O'Shea et al. [63] state that Convolutional Neural Networks (CNNs) are a type of deep artificial neural network commonly used in image analysis. A CNN can learn spatially related features by treating an image as a volume. The volume of images in a CNN is transformed using a variety of specialized layers. Most of the computation for classifying an image is done by a convolutional layer. A convolutional layer contains a series of kernels that move or convolve over an image volume. The ability of CNNs to recognize textures, shapes, colours, and other image features as their training progresses is one of their most significant advantages.

3.3.2. Object Detection Techniques

This section presents commonly used object detection techniques. In 2015, Girshick et al. [64] presented the Faster Region-based Convolutional Network method (Faster R-CNN) that takes as input an entire image and a set of object proposals. To create a feature map, the network first processes the entire image with several convolutional networks. Then, it extracts a fixed-length feature vector from the feature map. This model improves training and testing speed with increasing detection accuracy.

Liu et al [65] presented the Single Shot multi-box Detector (SSD) algorithm as a one-stage representative detection algorithm, which has an obvious speed advantage compared to the two-stage algorithm. Because of its good accuracy, SSD has become one of the main algorithms studied at present.

For a unified detector, Redmon et al. [66] proposed YOLO, which casts object detection as a regression problem from image pixels to spatially separated bounding boxes with associated class probabilities. Unlike the other approaches, YOLO does not include a stage for generating region proposals. As a result, YOLO uses a small set of candidate regions to detect objects directly. YOLO generates C class probabilities, B bounding box locations, and confidence scores by dividing an image into an S*S grid. In the articles under consideration, the following versions were used: YOLO V3 proposed by Redmon et al. [67], YOLO V4 proposed by Bochkovskiy et al. [68], and YOLO V5 [69].

3.3.3. Classification Techniques

Finally, we define some of the commonly used classification methods that have been mentioned in several articles.

Support Vector Machines (SVM) [70] are used for classification in supervised machine learning. A SVM can deal with issues like high dimensionality, small sample sizes, and more. It is often used for forecasting [71] and for resolving nonlinear data modelling issues [72].

In 2012, Krizhevsky [73] designed a new CNN model called AlexNet. It achieved a top-5 error of 15.3 in the Large-Scale Visual Recognition Competition (ILSVRC). AlexNet is composed of 8 layers. The first 5 layers are convolutional layers. Some of them are followed by max-pooling layers. The last 3 layers are fully connected layers.

In 2014, Simonyan et al. [74] proposed VGG. By stacking a few tiny convolution kernels and max-pooling layers, it increased the representation depth of the network. The structure's simplicity offered the benefit that the network performance could be enhanced by adding more depth. However, VGG makes use of more parameters, which consumes more memory. In the articles under consideration, the following versions were used: VGG16 and VGG19.

In 2015, the new state-of-the-art for classification and detection in ILSVRC14 was achieved by Szegedy et al.'s CNN model [75], called GoogleNet. It is a transfer learning CNN with 22 layers, also known as Inception v1. The inception layer is the core concept of GoogleNet. Concatenating a 1 x 1 convolutional (Conv) layer, a 3 x 3 convolutional layer, and a 5 x 5 convolutional layer into a single output vector is what makes up the inception layer. In the articles under consideration, the following versions were used: Inception and InceptionV2.

In 2016, He et al. [76] proposed ResNet. ResNet is one of the most used networks for classification and object detection tasks. The underlying idea are the residual blocks, which aim to simplify the training of neural networks characterized by many layers. Based on ResNet, Xie et al. [77] proposed ResNeXt, which uses the idea of increasing the cardinality. Compared with ResNet, it has the same parameters but higher accuracy. In the articles under consideration, the following versions were used: ResNet, ResNet18, and ResNet50.

In 2016, Iandola et al. [78] designed a new CNN model called SqueezeNet. The authors intention with SqueezeNet was to develop a smaller CNN with fewer parameters that could more readily fit into computer memory. The authors use SqueezeNet to reduce model size by 50* (5 MB) compared to AlexNet (240 MB) of parameters while maintaining or improving Alex-Net's top-1 accuracy.

In 2018, Sandler et al. [79] presented MobileNet as a deep convolutional architecture for mobile phones. It has a much smaller architecture and less calculation complexity than popular object detector models like the R-CNN. In the articles under consideration, the following versions were used: MobileNet and MobileNetv2.

3.3.4. Existing Detection Techniques

YOLOv5 is used by several authors [44], [51], [80]. YOLOv4 is used by both Genaev et al. [27] and Chen et al. [34], where the former uses the regular version and the latter [30] uses the tiny version. Finally, several authors [10], [29], [31], [81], [82] use YOLOv3, also known by the use of Dark-Net backbone. We also notice that YOLO has been used in combination with other classification techniques, such as Takimoto et al. [14] used YOLOv4 and Efficient-Net, and Kuzuhara et al. [8] used YOLOv3 and Xception (proposed by Chollet et al. [83]), and Liu et al. [43] used YOLOv3 with Global Context (GC) Network (proposed by Cao et al. [84]). Finally, Rustia et al. [33] used a small version of YOLOv3 that is appropriate for small devices, as well as two different classifiers; the names of the two classifiers are not mentioned.

Faster R-CNN is the second most popular technique [39],[59]. Faster R-CNN is frequently combined with other classification techniques. It is used by Ramalingam et al. [85] with ResNet, [32], [86] with Inception ResNet [87], while Rong et al. [22] used Mask R-CNN (which is improved version of Faster R-CNN, proposed by He et al. [88]) with the ResNet backbone.

The last detection technique used is SSD. In some cases, it is used alone [25], [41], [89] while Nam et al. [35] use SSD with VGG16. Finally, Patel et al. [28] used a combination of three techniques: Faster R-CNN and SSD with two different feature extractors: Inception and MobileNet.

3.3.5. Existing Classification Techniques

In the case only for classification techniques, SutHo et al. [50] use MobileNetv2, Luo et al. [42] use Inception-Net v3, Abeywardhana et al. [16] use SqueezeNet, Porrello et al. [90] use ResNet, Roosjen et al. [24] use ResNet18, and Menikdiwela et al. [9] utilize VGG16 fin-tuned on spiders. In the case of multiple techniques combined, Xia et al. [53] use VGG19 with a Region Proposal Network (RPN), Rajan et al. [47] use SVM along with a prepossessing algorithm, Chen et al. [86] use an improved Retina-Net (proposed by Lin et al. [91]) and convolutional block attention module (CBAM). Wang et al. [38] use ResNet50 with Features Pyramid Network (FPN).

Several articles use transfer learning, such as Khalifa et al. [37] who use 3 classification techniques: AlexNet, GoogleNet, and SqueezeNet. Lie et al. [40] and Hong et al. [16] present 7 techniques that have been transferred, while Butera et al. [7] use 12 techniques. Wang et al. [58] use 6 techniques to validate the proposed data set.

3.3.6. Proposals of New Techniques

Six of the articles presented novel techniques. Ding et al. [19] demonstrated ConvNet, which is based on a standard CNN. Yang et al. [26] introduced the Multi-layer Convolutional Structure (MCSNet), a detection and classification model comprising three parts: a VGG16-based insect features subnet, a region proposal network (RPN), and a classification network. Region CNN (R-CNN) is used to build the model's overall architecture. MAMPNet, a Multi-Attention and Multi-Part convolutional neural network based on ResNet50, is presented by Huang et al. [55]. Du et al. [30] present Pest R-CNN, which is based on the classical object detection model, Faster R-CNN. Du et al. [30] also mention that their model is divided into three parts: the same as the previous model, but with improved feature extractors.

Liu et al. [21] propose Pest-Net, a region-based end-to-end approach that is also made up of three parts: the Channel Spatial Attention (CSA), a Region Proposal Network (RPN), and a Position-Sensitive Score Map (PSSM) used to replace the Fully Connected layer. The backbone used is based on CNN, but it does not mention which one exactly. The CSA consists of two parts: the 3D feature map and 1D feature vectors.

The use of existing techniques does not mean that no modifications have been made. Menikdiwela et al. [9] change the last layer from a 1000 output layer to two output layers (spider and non-spider). Lyu et al. [25] add a top-down module to the SSD used. SutHo et al. [50] add two fully connected layers to the existing model. Geneav et al. [27] change the backbone of the used YOLO. Yuan et al. [80] changes the case-based learning (CBL) of YOLOv5. Chen et al.[86] add an improved full convolutional network (FCN) with CBAM for detection and classification.

There are two articles that used existing techniques to propose an entire system for detection and classification based on traps along with monitoring systems (cameras). The articles are Junior et al. [17] and Bjerge et al. [23]. Junior et al. [17] propose InsectCv based on Mask R-CNN, while [23] propose an automated light trap to monitor moths and mentions the use of a CNN model.

Only two of the studied articles do not mention the name of the used models. Hossain et al. [15] mention CNN in general, and Mazare et al. [52] mention artificial neural networks.

Numerous types of insects, including spiders [9], wheat pests [25], and others, were found in the studied articles. There was an insect species called the moth that was present in numerous studies, and it has been a serious problem for farmers around the world. This insect has been detected in several publications [19], [23], [28], [30], [41], [50], [52]. whereas Patel et al. [28] detect three insects where two of them are moths, and Korsch et al. [41] detect this type of insect in 200 species.

Some articles add the count of the insect to their model [17], [19], [22], [44], [50], [55] to determine the effectiveness of their installed trap.

In the studied papers, the number of detected insects is variable:

- Some articles only detect one insect or one class: [9], [19], [23], [30], [41], [42], [44], [50], [52], [82], [90],
- Others detect two classes: [8], [14], [26], [46],
- Other papers detect 3 classes: [7], [20], [27], [28], [32], [48], [51],
- Other articles detect 4 classes: [17], and [89],
- Some of the studied papers detect 5 classes: [35], [10], and [39],
- There are some papers that detect 6 classes: [15], [16], [18], [25], [34],
- Other papers detect from 7 to 9 classes: [24] (7 classes), [31], [37], and [85] (8 classes), and [22] (9 classes),
- Some papers detect from 10 to 19 classes: [38] (10 classes), [86] (11 classes), [81] (12 classes), [55], and [40] (15 classes), [21] (16 classes), and [80] (17 classes),
- Only 3 articles detect more than 20 classes: [43], [53] (24 classes), and [29] (23 classes).

Since there are fewer studies as the number of insects increases, 66% of all articles focus on insects with a population of one to six.

To summarize, this section presented the used detection techniques, which were divided into sub-classes: existing detection techniques; existing classification techniques; and proposals of novel techniques.

3.4. RQ3: The Location of the Studied Insects

In this section, we will present the locations of the studied articles in Africa, Europe, Asia, America, and Australia.

- Asia: China (16), Vietnam, Japan, Thailand, Sri-Lanka, India, Taiwan, Singapore, Korea, Bangladesh,
- America: Brazil, Canada, USA, Peru,
- Europe: Denmark, Hungary, Romania, The Netherlands, Germany, Russia,
- Africa: Egypt (2),
- Australia: Australia (2).

Table 2. The distribution of the chosen articles in continents

Continent	Number of Articles	(%)
Asia	27	054
Africa	02	004
Europe	14	028
America	05	010
Australia	02	004
Total	50	100

As shown in Table 2, Asia has the majority of studies done in the field of insect detection, for 54% of all articles (27 of 50). We notice that China has 16 out of the total of 27 articles in Asia. In Africa, only Egypt has two studies, of which Mamdouh et al. [44] detect olive insects and Khalifa et al. [37] detect 8 types of insects.

Two of the studied papers conduct their research in other countries: [32], [35]. Nam et al. [35] mention that the research centre is in Japan, but the studies are in Vietnam. They mention that their studies are based on customer requirements. The customer is facing a problem with a specific type of insect and asks for help. Ard et al. [32] mention the capture and annotation being done by experts from greenhouse research centres based in Belgium and Spain, since they are part of Europe.

The publication date of the selected articles should also be mentioned. The selection criteria only produced 3 articles prior to 2018 [19] in 2016; and [9], [46] in 2017; and the remaining articles were published within the previous 4 years, as shown in figure 2. A total of 22 articles, or 44% of all articles, were published in 2021. Additionally, only nine articles were published in the first three months of 2022. If there is any significance to this, it might mean that there is a lot of interest in this area.

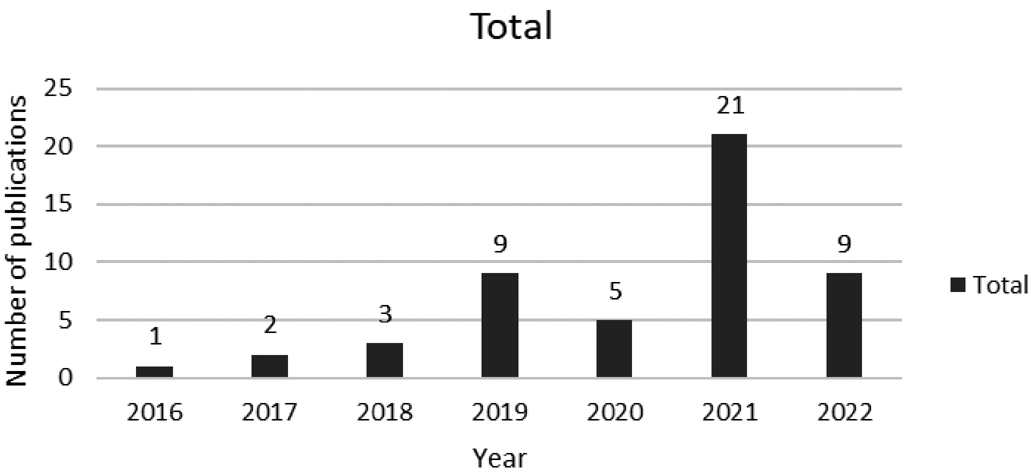

Figure 2. The distribution of the chosen articles

4. CONCLUSION

In our article, we provided a general overview of the field of insect detection by responding to three questions about it. The dataset that was used shows that on-site images that were taken in the field or a lab are more frequently used. We present the three types of detection techniques that are currently being used: existing detection or classification techniques; combinations of both; and new techniques. In addition, we discuss the use of transfer learning, which can shorten training periods. In response to the final query, which asks about the location of the detected insect, we note that China, which accounts for 56% of the articles chosen, is by far the leader in this area, while Africa still does not have a lot of concern in this field.

We now briefly outline our upcoming work. Initially, we will start with a small dataset that only shows one type of insect that is found in Tunisia before moving on to a larger data set that includes several types of them. Attacks on olive trees are commonplace for this insect. The next step is to develop a deep learning model capable of accurately identifying and counting this insect.

ACKNOWLEDGEMENTS

This work is supported by the German Academic Exchange Service (DAAD), Germany (Transformation Partnership: OLIVIA Project).

REFERENCES

[1] B. Ncube, W. Mupangwa, and A. French, (2018) “Precision agriculture and food security in africa,” in Systems analysis approach for complex global challenges. Springer, pp. 159–178.
[2] R. Gebbers and V. I. Adamchuk, (2010) “Precision agriculture and food security,” Science, vol. 327, no. 5967, pp. 828–831.
[3] R. N. Strange and P. R. Scott, (2005) “Plant disease: a threat to global food security,” Annual review of phytopathology, vol. 43, no. 1, pp. 83–116.
[4] D. C. K. Amarathunga, J. Grundy, H. Parry, and A. Dorin, (2021) “Methods of insect image capture and classification: A systematic literature review,” Smart Agricultural Technology, vol. 1, p. 100023.
[5] D. S. Hill, (1987) “Agricultural insect pests of the tropics and their control”. Cambridge University Press.

[6] D. S. Hill, (2012) “The economic importance of insects”. Springer Science & Business Media.

[7] L. Butera, A. Ferrante, M. Jermini, M. Prevostini, and C. Alippi, (2021) “Precise agriculture: Effective deep learning strategies to detect pest insects,” IEEE/CAA Journal of Automatica Sinica, vol. 9, no. 2, pp. 246–258.

[8] H. Kuzuhara, H. Takimoto, Y. Sato, and A. Kanagawa, (2020) “Insect pest detection and identification method based on deep learning for realizing a pest control system,” in 2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE). IEEE, pp. 709–714.

[9] M. Menikdiwela, C. Nguyen, H. Li, and M. Shaw, (2017) “Cnn-based small object detection and visualization with feature activation mapping,” in 2017 International Conference on Image and Vision Computing New Zealand (IVCNZ). IEEE, 2017, pp. 1–5.

[10] K. Li, J. Zhu, and N. Li, (2021) “Insect detection and counting based on yolov3 model,” in 2021 IEEE 4th International Conference on Electronics Technology (ICET). IEEE, pp. 1229–1233.

[11] “Google images,” https://images.google.com/, (Accessed on 08/05/2022).

[12] “Bing image inspiration feed,” https://www.bing.com/images/feed?form=Z9L, (Accessed on 06/01/2022).

[13] “flicker photos on flickr — flickr,” https://www.flickr.com/photos/tags/flicker/ (Accessed on 6/01/2022).

[14] H. Takimoto, Y. Sato, A. J. Nagano, K. K. Shimizu, and A. Kanagawa, (2021) “Using a two-stage convolutional neural network to rapidly identify tiny herbivorous beetles in the field,” Ecological Informatics, vol. 66, p. 101466.

[15] M. I. Hossain, B. Paul, A. Sattar, and M. M. Islam, (2019) “A convolutional neural network approach to recognize the insect: A perspective in bangladesh,” in 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART). IEEE, pp. 384–389.

[16] D. Abeywardhana, C. Dangalle, A. Nugaliyadde, and Y. Mallawarachchi, (2021) “Deep learning approach to classify tiger beetles of sri lanka,” Ecological Informatics, vol. 62, p. 101286.

[17] T. D. C. Junior, R. Rieder, J. R. Di Domenico, and D. Lau, (2022) “Insectcv: A system for insect detection in the lab from trap images,” Ecological Informatics, vol. 67, p. 101516.

[18] Y. Zhong, J. Gao, Q. Lei, and Y. Zhou, (2018) “A vision-based counting and recognition system for flying insects in intelligent agriculture,” Sensors, vol. 18, no. 5, p. 1489.

[19] W. Ding and G. Taylor, (2016) “Automatic moth detection from trap images for pest management,” Computers and Electronics in Agriculture, vol. 123, pp. 17–28, 2016.

[20] S.-J. Hong, S.-Y. Kim, E. Kim, C.-H. Lee, J.-S. Lee, D.-S. Lee, J. Bang, and G. Kim, (2020) “Moth detection from pheromone trap images using deep learning object detectors,” Agriculture, vol. 10, no. 5, p. 170.

[21] L. Liu, R. Wang, C. Xie, P. Yang, F. Wang, S. Sudirman, and W. Liu, (2019) “Pestnet: An end-to-end deep learning approach for large-scale multiclass pest detection and classification,” IEEE Access, vol. 7, pp. 45 301– 45 312.

[22] M. Rong, Z. Wang, B. Ban, and X. Guo, (2022) “Pest identification and counting of yellow plate in field based on improved mask r-cnn,” Discrete Dynamics in Nature and Society, vol. 2022.

[23] K. Bjerge, J. B. Nielsen, M. V. Sepstrup, F. Helsing-Nielsen, and T. T. Hoye, (2021) “An automated light trap to monitor moths (lepidoptera) using computer vision-based tracking and deep learning,” Sensors, vol. 21, no. 2, p. 343.

[24] P. P. Roosjen, B. Kellenberger, L. Kooistra, D. R. Green, and J. Fahrentrapp, (2020) “Deep learning for automated detection of drosophila suzukii: potential for uav-based monitoring,” Pest Management science, vol. 76, no. 9, pp. 2994–3002.

[25] Z. Lyu, H. Jin, T. Zhen, F. Sun, and H. Xu, (2021) “Small object recognition algorithm of grain pests based on ssd feature fusion,” IEEE Access, vol. 9, pp. 43 202–43 213.

[26] H. Yang, H. Zhao, D. Zhang, Y. Cao, S. W. Teng, S. Pang, X. Zhou, and Y. Li, (2022) “Auto-identification of two sitophilus sibling species on stored wheat using deep convolutional neural network,” Pest Management Science, vol. 78, no. 5, pp. 1925–1937.

[27] M. A. Genaev, E. G. Komyshev, O. D. Shishkina, N. V. Adonyeva, E. K. Karpova, N. E. Gruntenko, L. P. Zakharenko, V. S. Koval, and D. A. Afonnikov, (2022) “Classification of fruit flies by gender in images using smartphones and the yolov4-tiny neural network,” Mathematics, vol. 10, no. 3, p. 295.

[28] D. J. Patel and N. Bhatt, (2019) “Insect identification among deep learning’s meta-architectures using tensorflow,” Int. J. Eng. Adv. Technol, vol. 9, no. 1, pp. 1910–1914.

[29] P. Tresson, P. Tixier, W. Puech, and D. Carval, (2019) “Insect interaction analysis based on object detection and cnn,” in 2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP). IEEE, pp. 1–6.

[30] L. Du, Y. Sun, S. Chen, J. Feng, Y. Zhao, Z. Yan, X. Zhang, and Y. Bian, (2022) “A novel object detection model based on faster r-cnn for spodoptera frugiperda according to feeding trace of corn leaves,” Agriculture, vol. 12, no. 2, p. 248.

[31] K. Bjerge, H. M. Mann, and T. T. Hoye, (2021) “Real-time insect tracking and monitoring with computer vision and deep learning,” Remote Sensing in Ecology and Conservation.

[32] N. Ard and H. S. J. Hemming, (2018) “Detection and classification of insects on stick-traps in a tomato crop using faster r-cnn,” in Proceedings of the Netherlands Conference on Computer Vision, pp. 1–4.

[33] D. J. A. Rustia, J.-J. Chao, L.-Y. Chiu, Y.-F. Wu, J.-Y. Chung, J.-C. Hsu, and T.-T. Lin, (2021) “Automatic greenhouse insect pest detection and recognition based on a cascaded deep learning classification method,” Journal of Applied Entomology, vol. 145, no. 3, pp. 206–222.

[34] C. Chen, Y. Liang, X. Tang, M. Dai, and K. Zhou, (2022) “The research of pest detection in granary based on yolov4,” Social Science Research Network.

[35] N. T. Nam and P. D. Hung, (2018) “Pest detection on traps using deep convolutional neural networks,” in Proceedings of the 2018 International Conference on Control and Computer Vision, pp. 33–38.

[36] X. Wu, C. Zhan, Y.-K. Lai, M.-M. Cheng, and J. Yang, (2019) “Ip102: A large scale benchmark dataset for insect pest recognition,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 8787–8796.

[37] N. E. M. Khalifa, M. Loey, and M. H. N. Taha, (2020) “Insect pests recognition based on deep transfer learning models,” J. Theor. Appl. Inf. Technol, vol. 98, no. 1, pp. 60–68.

[38] Z. Wang, L. Qiao, and M. Wang, (2022) “Agricultural pest detection algorithm based on improved faster rcnn,” in International Conference on Computer Vision and Pattern Analysis (ICCPA 2021), vol. 12158. SPIE, pp. 104–109.

[39] M. E. Karar, F. Alsunaydi, S. Albusaymi, and S. Alotaibi, (2021) “A new mobile application of agricultural pests recognition using deep learning in cloud computing system,” Alexandria Engineering Journal, vol. 60, no. 5, pp. 4423–4432.

[40] H. Liu and J. S. Chahl, (2021) “Proximal detecting invertebrate pests on crops using a deep residual convolutional neural network trained by virtual images,” Artificial Intelligence in Agriculture, vol. 5, pp. 13–23.

[41] D. Korsch, P. Bodesheim, and J. Denzler, (2021) “Deep learning pipeline for automated visual moth monitoring: Insect localization and species classification,” INFORMATIK 2021.

[42] C. Luo, P. Pearson, G. Xu, and S. M. Rich, (2022) “A computer vision-based approach for tick identification using deep learning models,” Insects, vol. 13, no. 2, p. 116.

[43] Q. Liu, Z. Yan, F. Wang, and C. Ding, (2021) “Research on object detection algorithm for small object of pests based on yolov3,” in 2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI). IEEE, pp. 14–18.

[44] N. Mamdouh and A. Khattab, (2021) “Yolo-based deep learning framework for olive fruit fly detection and counting,” IEEE Access, vol. 9, pp. 84 252–84 262.

[45] R. Kalamatianos, I. Karydis, D. Doukakis, and M. Avlonitis, (2018) “Dirt: The dacus image recognition toolkit,” Journal of Imaging, vol. 4, no. 11, p. 129.

[46] M. Tanjim-Al-Akib and N. Islam, (2017) “Feature optimization of whitefly detection algorithm using image segmentation and feature analysis,”.

[47] P. Rajan, B. Radhakrishnan, and L. P. Suresh, (2016) “Detection and classification of pests from crop images using support vector machine,” in 2016international conference on emerging technological trends (ICETT).IEEE, pp. 1–6.

[48] M. Deserno and A. Briassouli, (2021) “Faster r-cnn and efficient net for accurate insect identification in a relabeled yellow sticky traps dataset,” in 2021 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor). IEEE, pp. 209–214.

[49] “Raw data from yellow sticky traps with insects for training of deep learning convolutional neural network for object detection — research@wur,” https://research.wur.nl/en/datasets/raw-data-from-yellowsticky-traps-with-insects-for-training-of-de, (Accessed on 08/05/2022).

[50] J. Suto, (2021) “Embedded system-based sticky paper trap with deep learning based insect-counting algorithm,” Electronics, vol. 10, no. 15, p. 1754.

[51] J. Cabrera and E. Villanueva, (2022) "Investigating generative neural-network models for building pest insect detectors in sticky trap images for the Peruvian horticulture," in Annual International Conference on Information Management and Big Data. Springer, pp. 356–369.
[52] A. Mazare, L. Ionescu, D. Visan, N. Belu, and A. Lita, (2019) "Pests detection system for agricultural crops using intelligent image analysis," in 2019 IEEE 25th International Symposium for Design and Technology in Electronic Packaging (SIITME). IEEE, pp. 228–231.
[53] D. Xia, P. Chen, B. Wang, J. Zhang, and C. Xie, (2018) "Insect detection and classification based on an improved convolutional neural network," Sensors, vol. 18, no. 12, p. 4169.
[54] C. Xie, J. Zhang, R. Li, J. Li, P. Hong, J. Xia, and P. Chen, (2015) "Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning," Computers and Electronics in Agriculture, vol. 119, pp. 123–132.
[55] R. Huang, T. Yao, C. Zhan, G. Zhang, and Y. Zheng, (2021) "A motor-driven and computer vision-based intelligent e-trap for monitoring citrus flies," Agriculture, vol. 11, no. 5, p. 460.
[56] "A community for naturalists inaturalist," https://www.inaturalist.org/, (Accessed on 06/01/2022).
[57] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, (2009) "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition. Ieee, pp. 248–255.
[58] R. Wang, L. Liu, C. Xie, P. Yang, R. Li, and M. Zhou, (2021) "Agripest: A large-scale domain-specific benchmark dataset for practical agricultural pest detection in the wild," Sensors, vol. 21, no. 5, p. 1601.
[59] J. Li, H. Zhou, D. S. Jayas, and Q. Jia, (2019) "Construction of a dataset of stored-grain insects images for intelligent monitoring," Applied Engineering in Agriculture, vol. 35, no. 4, pp. 647–655.
[60] "puzzledqs/bbox-label-tool: A simple tool for labeling object bounding boxes in images," https://github.com/puzzledqs/BBox-Label-Tool, (Accessed on 06/01/2022).
[61] "heartexlabs/labelimg: labelimg is a graphical image annotation tool and label object bounding boxes in images," https://github.com/heartexlabs/labelImg, (Accessed on 06/01/2022).
[62] "wkentaro/labelme: Image polygonal annotation with python (polygon, rectangle, circle, line, point and image-level flag annotation)." https://github.com/wkentaro/labelme, (Accessed on 06/01/2022).
[63] K. O'Shea and R. Nash, (2015) "An introduction to convolutional neural networks," arXiv preprint arXiv:1511.08458.
[64] R. Girshick, (2015) "Fast r-cnn," in Proceedings of the IEEE International Conference on Computer Vision (ICCV).
[65] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, (2016) "Ssd: Single shot multibox detector," in European conference on computer vision. Springer, pp. 21–37.
[66] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, (2016) "You only look once: Unified, real-time object detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[67] J. Redmon and A. Farhadi, (2018) "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767,.
[68] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, (2020) "Yolov4: Optimal speed and accuracy of object detection," ArXiv, vol. abs/2004.10934.
[69] "ultralytics/yolov5: Yolov5 in pytorch onnx coreml . tflite," https://github.com/ultralytics/yolov5, (Accessed on 06/01/2022).
[70] C. Cortes and V. Vapnik, (1995) "Support-vector networks," Machine learning, vol. 20, no. 3, pp. 273–297.
[71] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu et al., (2008) "Top 10 algorithms in data mining," Knowledge and information systems, vol. 14, no. 1, pp. 1–37.
[72] H.-x. Zhao and F. Magoules, (2012) "A review on the prediction of building energy consumption," Renewable and Sustainable Energy Reviews, vol. 16, no. 6, pp. 3586–3592.
[73] A. Krizhevsky, I. Sutskever, and G. E. Hinton, (2012) "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, vol. 25.
[74] K. Simonyan and A. Zisserman, (2014) "Very deep convolutional networks forvlarge-scale image recognition," arXiv preprint arXiv:1409.1556.
[75] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, (2015) "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9.

[76] K. He, X. Zhang, S. Ren, and J. Sun, (2016) “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
[77] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, (2017) “Aggregated residual transformations for deep neural networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[78] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, (2016) “Squeezenet: Alexnet-level accuracy with 50x fewer parameters and! 0.5 mb model size,” arXiv preprint arXiv:1602.07360.
[79] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, (2018) “Mobilenetv2: Inverted residuals and linear bottlenecks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510–4520.
[80] Z. Yuan, W. Fang, Y. Zhao, and V. S. Sheng, (2021) “Research of insect recognition based on improved yolov5,” Journal of Artificial Intelligence, vol. 3,no. 4, p. 145.
[81] F. A. G. d. Silveira, E. C. Tetila, G. Astolfi, A. B. d. Costa, and W. P. Amorim, (2021) “Performance analysis of yolov3 for real-time detection of pests in soybeans,” in Brazilian Conference on Intelligent Systems. Springer, pp. 265–279.
[82] A. M. Hubalde, D. A. Padilla, and D. A. C. Santos, (2021) “A yolo-based approach for aedes aegypti larvae classification and detection,” in 2021 6th International Conference on Image, Vision and Computing (ICIVC). IEEE, pp. 161–167.
[83] F. Chollet, (2017) “Xception: Deep learning with depthwise separable convolutions,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1251–1258.
[84] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, (2019) “Gcnet: Non-local networks meet squeeze-excitation networks and beyond,” in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops.
[85] B. Ramalingam, R. E. Mohan, S. Pookkuttath, B. F. G´omez, C. S. C. Sairam Borusu, T. Wee Teng, and Y. K. Tamilselvam, (2020) “Remote insects trap monitoring system using deep learning framework and iot,” Sensors, vol. 20, no. 18, p. 5280.
[86] Y. Chen, X. Zhang, W. Chen, Y. Li, and J. Wang, (2020) “Research onrecognition of fly species based on improved retinanet and cbam,” IEEE Access, vol. 8, pp. 102 907–102 919.
[87] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, (2017) “Inception-v4, inception-resnet and the impact of residual connections on learning,” in AAAI.
[88] K. He, G. Gkioxari, P. Dollar, and R. Girshick, (2017) “Mask r-cnn,” in 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2980–2988.
[89] R. Trufelea, M. Dimoiu, L. Ichim, and D. Popescu, (2018) “Detection of harmful insects for orchard using convolutional neural networks,” U.P.B. Sci. Bull., Series C.
[90] A. Porrello, S. Vincenzi, P. Buzzega, S. Calderara, A. Conte, C. Ippoliti, L. Candeloro, A. Di Lorenzo, and A. C. Dondona, (2019) “Spotting insects from satellites: modeling the presence of culicoides imicola through deep cnns,” in 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). IEEE, pp. 159– 166.
[91] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, (2017) “Focal Loss for Dense Object Detection,” arXiv e-prints, p. arXiv:1708.02002.

AUTHORS

Lotfi Souifi received the licence in computer science from the faculty of science of Monastir, university of Monastir, Tunisia, in 2016. He is currently a student in the second year of master in cyber physical systems in the high institute of computer science and multimedia of Sfax, university of Sfax, Tunisia. His current research areas include computer vision, deep learning, and object detection.

Afef Mdhaffar received the Engineering and M.Sc. degrees from the National Engineering School of Sfax, Tunisia, in 2007 and 2009, respectively, and the Ph.D. degree in computer science from the University of Marburg, Germany, in 2014. She is currently an Assistant Professor at the University of Sfax, Tunisia, and an Associate Researcher at the Digital Research Center of Sfax, Tunisia. She is involved in self-* architectures applied to cloud computing environments, data analytics methods, e-health, and precision agriculture.

Ismael Bouassida Rodriguez received the engineering and M.Sc. degrees in computer science from the National School of Computer Sciences of Manouba (Tunisia) and his Ph.D. from the University of Toulouse (France) and the University of Sfax (Tunisia) in 2011. He joined the National School of Engineers in Sfax as Assistant Professor of computer science in 2010 and he became an Associate Professor in 2012 at the Higher Institute of Computer Science and Multimedia of Sfax. His current research areas include software engineering of distributed systems, graphs grammars, self-adaptive and pervasive systems, and autonomic middleware.

Mohamed Jmaiel obtained his diploma of engineer in computer science from Kiel (Germany) University in 1992 and his Ph.D. from the Technical University of Berlin in 1996. He is, since 2009, full professor in computer science at the National School of Engineers of Sfax (Tunisia). His current research areas include software engineering of distributed systems, formal methods in model-driven architecture, self-adaptive and pervasive systems, autonomic middleware. Currently, he is the general director of the Digital Research Center at the Technopark of Sfax.

Bernd Freisleben received his M.Sc. in computer science from Pennsylvania State University, USA, in 1981, and his PhD and Habilitation in computer science from Technical University of Darmstadt, Germany, in 1985 and 1993, respectively. He is a full professor of computer science in the Department of Mathematics and Computer Science, University of Marburg, Marburg, Germany. His current research interests include data analytics, machine learning, and biomedical applications.