



HAL
open science

Conditional Multi-Task learning for Plant Disease Identification

Sue Han Lee, Hervé Goëau, Pierre Bonnet, Alexis Joly

► **To cite this version:**

Sue Han Lee, Hervé Goëau, Pierre Bonnet, Alexis Joly. Conditional Multi-Task learning for Plant Disease Identification. ICPR 2020 - 25th International Conference on Pattern Recognition, The International Association for Pattern Recognition (IAPR); ICPR2020 Organizing committee, Jan 2021, Milan, Italy. pp.3320-3327, 10.1109/ICPR48806.2021.9412643 . hal-03415972

HAL Id: hal-03415972

<https://inria.hal.science/hal-03415972>

Submitted on 10 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Conditional Multi-Task learning for Plant Disease Identification

Sue Han Lee

Swinburne University of Technology
Sarawak Campus,
Jalan Simpang Tiga
93350 Kuching, Sarawak, Malaysia
Email: shlee@swinburne.edu.my

Hervé Goëau

CIRAD, UMR AMAP
Montpellier, France
Email: herve.goeau@cirad.fr

Pierre Bonnet

CIRAD, UMR AMAP
Montpellier, France
Email: pierre.bonnet@cirad.fr

Alexis Joly

INRIA
Montpellier, France
Email: alexis.joly@inria.fr

Abstract—Several recent studies have proposed an automatic plant disease identification system based on deep learning. However, these approaches are generally based on learned classification models with target classes of joint host species-disease pairs that may not allow optimal use of the available information. This is because these approaches require distinguishing between similar host species or diseases, and more importantly, they have limited scalability due to the size of a network gradually increases as new classes are added, despite the fact that information on host species or diseases is already available. This constraint is all the more important as it can be difficult to collect/establish a specific list of all diseases for each host plant species in an actual application. In this paper, we address the problems by proposing a new conditional multi-task learning (CMTL) approach which allows the distribution of host species and disease characteristics learned simultaneously with a conditional link between them. This conditioning is formed in such a way that the knowledge to infer the prediction of one concept (the diseases) depends on the other concept (the host species), which corresponds to the way plant pathologists used to infer the diseases of the host species. We show that our approach can improve the performance of plant disease identification compared to the usual species-disease pair modeling in the previous studies. Meanwhile, we also compose a new dataset on plant disease identification that could serve as an important benchmark in this field.

Index Terms—plant disease identification, multi-task learning

I. INTRODUCTION

Plant diseases are a major threat to agricultural production, causing a severe food recession and affecting crop quality [2]. To detect diseases in crops, plant pathologists generally use molecular and serological methods or measurements of various parameters such as morphological change, temperature change, change in transpiration rates and emission of volatile organic compounds from infected plants [3]. Expert consultation is one of the major way to control plant diseases, but it is time consuming and expensive, and it is not always easy to find an available pathologist, specialist of the target crop or disease. In recent years, automated plant disease identification has been introduced into plant taxonomy to potentially fill gaps in human disease identification capabilities. Researchers use deep learning techniques to automatically identify diseases of single or multiple host species. In this paper, the scope of research is not limited to diseases of a single host species, but

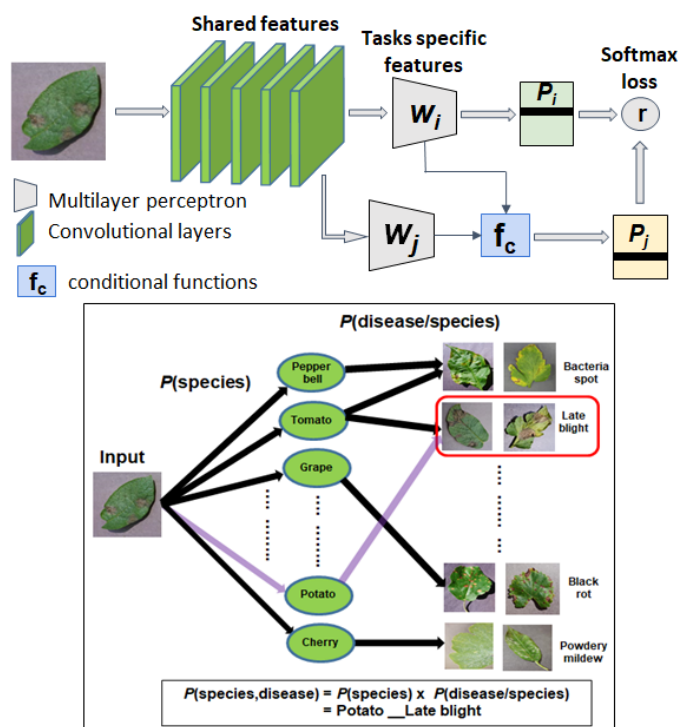


Fig. 1. The identification of plant diseases is based on two main concepts which are host species and diseases. These two concepts are closely related and can be structurally linked in such a way that knowledge of one concept (j) depends on the other (i). In this case, i is equivalent to the host species concept while j is the disease concept. Such modeling corresponds to the way plant pathologists use species information to infer associated diseases [1].

to a large number of diseases of several host species in order to increase the flexibility of applications in different fields.

To learn the representation of diseases of multiple host species [4], [5], [6], in all cases, researchers trained standard convolutional neural network (CNN) models to distinguish characteristics based on joint species-disease classes, for example: *Grape__Black_rot*, *Apple__Apple_scab* and *Apple__Black_rot*. Although successfully, it requires that the size of the network gradually increases when new classes are added, even though there is already information on crops

or diseases. In addition, we consider that it would be very difficult, if not impossible, in a real application to collect a complete set of images illustrating each species-disease pair. In fact, it may not be easy to establish a specific list of all diseases for each host species. We note that pathogens with a wide host range can infect many different host species. For example, as reported in [7], nearly 200 plant species can be infected by bacterial wilt (*Ralstonia solanacearum*). Therefore, a new strategy is needed to improve the efficiency of plant disease identification modelling in order to increase the flexibility to transfer knowledge from one disease context to another.

In this paper, we propose to use a conditional multi-task learning (CMTL) in the identification of plant diseases, replacing previous modelling approaches [4], [5], [6] with joint host species-disease pairs. More specifically, we propose to model two main concepts/tasks which are host species and diseases through multi-task learning, and each concept is represented by its own feature distribution but they are strongly linked by a specific conditional structure, as shown in Fig. 1. This structured process corresponds not only to the fundamental building block of human intelligence to reason for everyday visual input on the basis of related prior knowledge [8], but more importantly, corresponds to the way plant pathologists use species information to infer the associated diseases [1]. Therefore, unlike previous works on large-scale multitasking [9], [10], [11], in this paper, we aim to solve the problem of plant disease identification tasks where the tasks are interrelated of which the representation of one task is considered an obvious indicator for inferring the other task.

In short, the contributions of the paper are as follows:

- 1) This paper presents a new method for classifying plant diseases. It deploys a multi-task learning mechanism and proposes a conditional scheme strongly adapted to the context of plant disease identification.
- 2) The experimental results show that our proposed conditional multi-task learning can improve the performance of plant disease identification compared to the traditional joint species-disease pair modeling.
- 3) We compose a new dataset that is used for benchmarking the accuracy of plant disease identification. To our knowledge, this is the first plant disease identification dataset that covers the largest number of host species and diseases. In exact, we have 1145 joint species-disease classes, 311 host species and 289 diseases, which is larger than the well-known PlantVillage dataset [4] with only 38 joint species-disease, 14 host species and 20 number of diseases.

II. RELATED WORKS

A. Plant disease identification

Early studies on automated identification of plant diseases focused primarily on one single host species and relied on hand-crafted approaches, but as deep learning techniques have demonstrated extremely high recognition capabilities in many image recognition tasks, multi-host species and diseases based

models are receiving increasing attention [13], [14]. To form a multi-host species and diseases model, a straightforward approach is to consider target classes as *joint species-diseases* [15], [4], [5], [6], [16], as shown in Fig. 2c. Although successful, there remains a doubt as to whether the characteristics learned from such model are optimal for identifying plant diseases, especially when the model is trained to discriminate similar host species or diseases. A potentially undesirable effect that may occur is that the model learns more from the content related to the species than the visual content related to the disease [4], [17], [13]. This adverse effect can potentially be amplified if a species has a very pronounced characteristics while the disease's features are not very visible. In addition, it requires that the size of the network gradually increases when new classes are added, even though there is already information on crops or diseases. Although [13] addressed the problem by proposing a symptoms oriented feature modelling strategy by considering only diseases without considering host species, this strategy may not be consistent with the perception of expertise in distinguishing plant diseases, particularly when information on host species is often needed to delineate lists of diseases associated with host species [1]. This observation prompts us to examine methods that could use prior knowledge of host species to assist disease inference. Therefore, to overcome these limitations, in this work, we study an alternative approach to the usual modeling of joint species-disease pairs by proposing a new approach based on conditional multi-task learning (CMTL). We individually model the distribution of the host species and diseases, and train them simultaneously using our proposed CMTL method. The hypothesis is that such intuitive modeling will engage the model to learn better individual characteristics of the host species and the disease, while exploiting the relationship between them based on a conditional learning scheme that will be explained in Sec. III.

B. Multi-task learning (MTL)

Inspired by the ability of humans to generalize and transfer knowledge acquired in one type of activity to other related tasks, MTL has been widely explored in the field of artificial intelligence to accelerate up the learning process and improve the generalization performance of multiple tasks. It has been successfully applied from speech and natural language processing to computer vision, bioinformatics, health informatics, web applications and ubiquitous computing [18], [19]. There are multiple learning tasks in MTL, each of which can be a general learning task such as supervised tasks (classification and regression problems), unsupervised tasks (clustering problem), semi-supervised tasks, reinforcement learning tasks, etc. Among these learning tasks, it is assumed that all or at least a subset of them are related to each other. In fact, there is evidence that joint learning of these tasks can lead to a considerable improvement in performance compared to learning them individually [20], [18], [19]. General approaches based on deep learning can be categorised into *soft* and *hard* parameters sharing [21], [22]. In the soft parameter sharing, each task is modeled by its own parameters, while

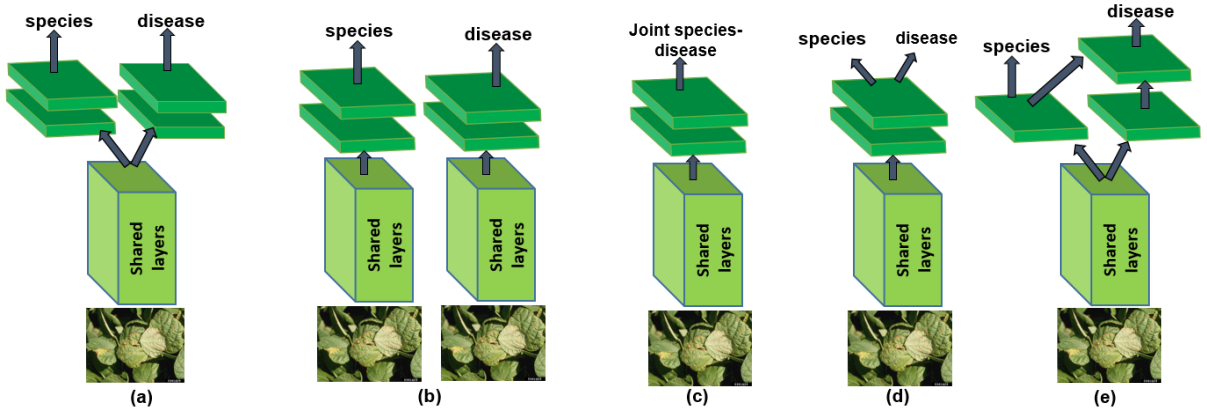


Fig. 2. Examples of CNN models used in the experiments. In order to obtain a fair comparison, all models were built on Inception-V3 with the same size of shared layers (depth). (a), (b) and (c) illustrate the baselines architectures of different configurations. (a) which is called the *generic MTL* is composed of shared layers and task-specific layers, (b) which is called the *single nets* is composed of individual networks for each concept and (c) which is called the *Joint net* illustrates the conventional classification of plant diseases with joint species-disease modelling. (d) is the classification model proposed by [12] which was designed to solve two closely related tasks of visual fashion analysis. We called it *Fashion net*. (e) is our proposed new terminology *CMTL* where predictive modeling of one concept depends on the other.

for the hard parameter sharing approach, the hidden layer parameters are shared between all tasks while keeping several output layers specific to the tasks. Since in the case of soft parameter sharing, the size of the network increases linearly with the number of tasks (which is not so effective in terms of parameters compared to the hard parameter sharing), and since hard parameter based approach has been recognised as the most commonly used approach for MTL [18], we use it to test our concept. Next, to adapt this context to the identification of plant diseases, we improvise the classical MTL approach (Fig. 2a) by introducing CMTL strategy (Fig. 2e), and by experimental analysis, we show that CMTL approaches can achieve better performance compared to MTL.

III. METHODOLOGY

A. Definitions and Preliminaries

In the general framework of MTL, given an entry x with several classification tasks t_i where $i = 1, 2, \dots, k$, the predictive modeling can be formulated by [23]:

$$P(t_1, \dots, t_k / f(\mathbf{x})) = P(t_1 / f(\mathbf{x})) \dots P(t_k / f(\mathbf{x})) \quad (1)$$

where $f(\mathbf{x})$ is neural network function that takes into account \mathbf{x} and is parameterized by a weight vector \mathbf{w} . The \mathbf{w} consist of the trainable shared weights \mathbf{w}^{sh} and the task-specific weights \mathbf{w}^{ts} .

In this paper, we aim to solve two related classification tasks in the identification of plant diseases by giving an input image \mathbf{x} . These two related classification tasks are the identification of diseases and host species. We present a new strategy, called CMTL, in which it introduces a conditional learning mechanism of task relationship where the learning of the characteristics of a task is conditioned by another task. Thus, unlike the traditional formulation in Eqn. (1), predictive modeling is formulated in such a way that:

$$P(t_s, t_d / f(\mathbf{x})) = P(t_s / f_s(\mathbf{x})) P(t_d / f_s(\mathbf{x}), f_d(\mathbf{x})) \quad (2)$$

where s and d represent the host species and disease respectively. To achieve the conditional probability $P(t_d / f_s(\mathbf{x}), f_d(\mathbf{x}))$, we learn the conditional function $f_c(\mathbf{x})$ by employing two different approaches which are the *conditional feature fusion* and *conditional feature-wise linear modulation*.

B. Conditional feature fusion

The feature fusion based approach is responsible to softly forward the disease-specific features conditioning on the species-specific features for the disease predictive modeling via feature fusion. We test with two common fusion strategies: sum fusion and cascade fusion.

1) *Sum fusion*: In sum fusion (Fig. 3a), it learns functions of H_σ and H_γ which output σ and γ as a function of $\mathbf{w}_s^{\text{ts}}(\mathbf{x})$ and $\mathbf{w}_d^{\text{ts}}(\mathbf{x})$ respectively, and both signals are modulated via the following functions:

$$\sigma = H_\sigma(\mathbf{w}_s^{\text{ts}}(\mathbf{x})) \quad \gamma = H_\gamma(\mathbf{w}_d^{\text{ts}}(\mathbf{x})) \quad (3)$$

$$f_c(\mathbf{x}) = \tanh(\sigma + \gamma) \quad (4)$$

\mathbf{w}_i^{ts} is the function of the first embedding layer of task i , producing hidden units $\mathbf{w}_i^{\text{ts}}(\mathbf{x}) \in \mathbb{R}^{1 \times N}$ and N is the number of hidden units. H_σ and H_γ can be arbitrary functions such as neural networks while in this case are formed using fully connected layers.

2) *Cascade fusion*: Different from the sum fusion, both the output neurons σ and γ are cascaded and then pass through a reduction layer before the logits layer, as illustrated in Fig. 3b. More formally, the function is formulated as follows:

$$\sigma = H_\sigma(\mathbf{w}_s^{\text{ts}}(\mathbf{x})) \quad \gamma = H_\gamma(\mathbf{w}_d^{\text{ts}}(\mathbf{x})) \quad (5)$$

$$f_c(\mathbf{x}) = \text{relu}(H_\beta([\sigma, \gamma])) \quad (6)$$

where the cascaded features have size $\mathbb{R}^{1 \times 2N}$ and H_β is the reduction function formed using fully connected layer.

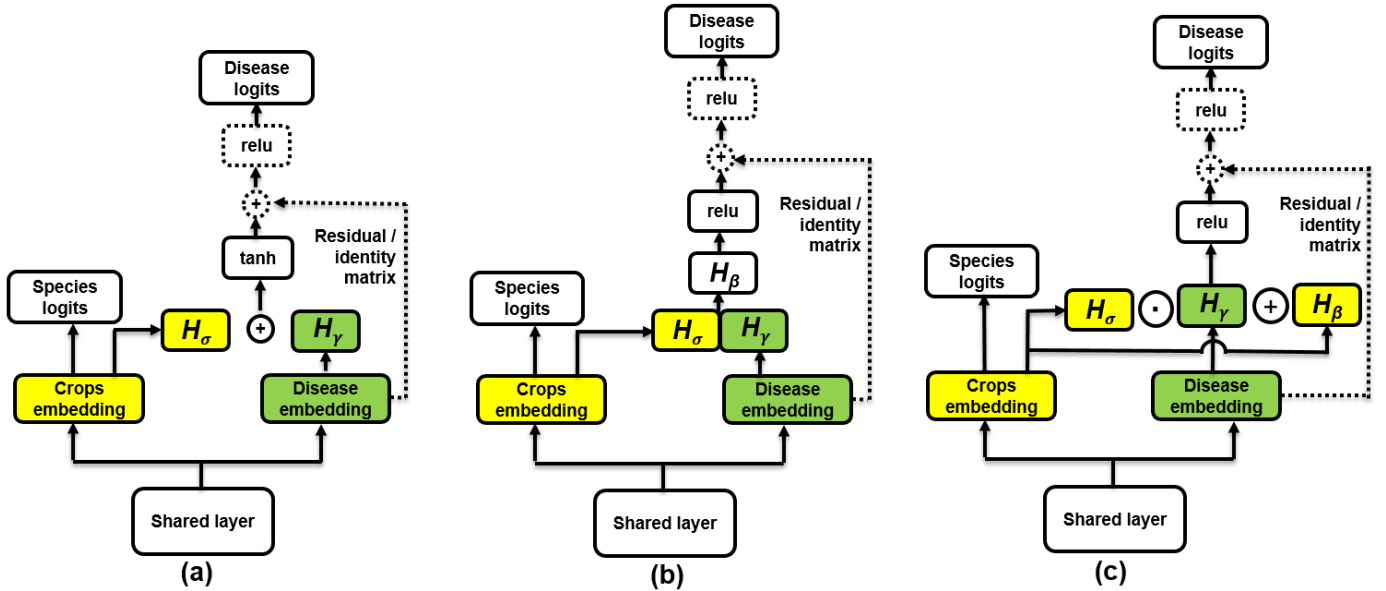


Fig. 3. Examples of our CMTL proposals: Conditional feature fusion scheme - (a) Sum (FFS) and (b) Cascade (FFC) fusion, and (c) Conditional feature-wise linear modulation (FiLM). Note that the identical mapping which is introduced to enrich the features of the disease layer is illustrated with dotted line.

C. Conditional feature-wise linear modulation

Feature-wise linear modulation (FiLM), considered as a conditional normalization method [24], has been widely used in several recent works proposing conditioning networks. For example, conditional image synthesis [25], image-based question answering [24] and conditional auto encoder for open set recognition [26]. In this work, we use the same approach to modulate the features of the disease at the fully connected layer, conditioned by information on the host species, as shown in Fig. 3c. As a function of input host species features $\mathbf{w}_s^{\text{ts}}(\mathbf{x})$, the FiLM conditioning layer on the disease features $\mathbf{w}_d^{\text{ts}}(\mathbf{x})$ is formulated as follows:

$$\sigma_j = H_\sigma(\mathbf{w}_s^{\text{ts}}(\mathbf{x})) \quad \beta_j = H_\beta(\mathbf{w}_s^{\text{ts}}(\mathbf{x})) \quad (7)$$

$$f_c(\mathbf{x}) = \sigma_j \odot H_\gamma(\mathbf{w}_d^{\text{ts}}(\mathbf{x})) + \beta_j \quad (8)$$

where H_σ and H_β are fully connected layers learned as a function of $\mathbf{w}_s^{\text{ts}}(\mathbf{x})$ while H_γ is fully connected learned as a function of $\mathbf{w}_d^{\text{ts}}(\mathbf{x})$, subscript j is refer to the j^{th} neurons and the \odot defines the Hadamard product. Fig. 3c illustrates the FiLM conditioning configuration.

D. Identical mapping

With the features \mathbf{F} output from the conditional layer f_c , we introduce an identical function to allow the disease embedding features $\mathbf{w}_d^{\text{ts}}(\mathbf{x})$ and \mathbf{F} to be combined as input to the next layer as follows:

$$y = \text{relu}(\mathbf{F} + \mathbf{w}_d^{\text{ts}}(\mathbf{x})) \quad (9)$$

The function is illustrated by the dotted contents in Fig. 3a-c. This approach builds on the work on residual learning [27] to ensure that the performance of the model with conditional learning mechanism is not degraded compared to

a model without conditional learning. In addition, such a skip architecture could help to enrich the features of the final layer of the disease with the information that has been captured in the initial layers. We provide a detailed performance analysis in Sec. V.

E. Training objective

This paper aims to solve multi-task classification in the identification of plant diseases by giving an input image of x . The output of the last hidden layer corresponding to the i -th task is normalized by the softmax function

$$P_i(r|\mathbf{x}) = \frac{e^{s_{r,i}(\mathbf{x})}}{\sum_{m=1}^M e^{s_{m,i}(\mathbf{x})}} \quad (10)$$

where M and r represent the total number of classes and the target class respectively, while $s(\mathbf{x})$ represents the final activation of the x input image obtained in the last hidden layer/logits layer. After performing the softmax operation, we find the maximum likelihood of the sample by applying the objective function, $L_i = -\log P_i(r|\mathbf{x})$. To optimize both tasks t_s and t_d , we simply add up both losses $L = L_s + L_d$.

IV. PLANT DISEASE DATASET

In the area of identification of plant diseases for multiple host species, there are not many publicly available datasets with which researchers can work and, in many cases, researchers must develop their own datasets. The process, from the collection of samples to verification by the plant's experts, can take several hours or days of work. To date, only two main datasets with trusted labels are available online: PlantVillage (PV) [4] and Digipathos [28]. The majority of studies are based on the PV dataset [6], [4], which is currently the largest

TABLE I
DISTRIBUTION OF THE TRAINING AND VALIDATION SET FOR EACH DATASET: (1) PLANTVILLAGE, (2) DIGIPATHOS, (3) IPM AND (4) PL@NTNET.

Dataset	Number of training data	Number of validation data
PV	594	146
Digipathos	1668	345
IPM	4522	812
Pl@ntNet	3540	663
Total	10324	1966

dataset containing 54305 number of images. However, PV images are collected in a constrained environment with a homogeneous background, and the model trained using these images does not work well with images captured in a realistic scenario (such as in field conditions). This can be seen from the huge drop in model performance in [4] when it was used to predict images captured in environment with cluttered backgrounds, reflecting a real-world scenario. Although the Digipathos dataset addresses the above problem, the number of images is too small (less than 5%) compared to the number of PV images, which is 2326 images. Therefore, in this work, we create a new dataset on diseases of multiple host species that contains not only images captured in a homogeneous background, but also images captured in agricultural fields, in order to fit to a real-life scenario.

To build a new dataset, we downloaded the plant disease images from IPM website (<https://www.ipmimages.org/>) where the images were verified and contributed by the expertise in plant disease management, and also using a combination of PV and Digipathos to enlarge the database. From IPM website, we were able to download a total of 5334 images with complete information on host species and diseases. Since the PV dataset contains only images captured under a constrained background and the total number of images is larger than the samples that were captured in a real scenario, to avoid biasing the dataset into PV images constrained by a homogeneous background, we took only a subset of images (~ 20 number of images) for each PV class. Next, we acquired healthy images from the Pl@ntNet database (<http://identify.plantnet-project.org/>) through a visual verification step. Based on these processes, we obtained a total of 12,290 images with 1146 joint species-disease labels, 311 host species and 289 number of diseases (288 diseases + 1 healthy label). In addition, it covers a total of 7163 healthy images and 5122 unhealthy images. We then randomly separated the images into training and validation set with a number of 10324 ($\sim 80\%$) and 1966 ($\sim 20\%$) respectively. The distribution of the training and validation set for each dataset is shown in the Table I. Note that these images composed of different plant organs such as the leaf, branch, stem and fruit. Although the new database contains fewer images than the PV database, it is more complex due to a larger number of classes and more diverse images, making the task more realistic and difficult.

As for the testing set, we received images collected by

TABLE II
LISTS AND NUMBER OF IMAGES PROVIDED BY THE INRAE PHYTOPHATOLOGIST, WHICH IS USED AS A TESTING SET. NOTE THAT, THE NAMES ARE WRITTEN IN THE FORM '*species__disease*'.

Name	Total images
<i>Zucchini__Podosphaera_xanthii</i>	90
<i>Tomato__Cladosporiose</i>	82
<i>Vine__Plasmopara_viticola</i>	150
<i>Tomato__ToMV</i>	66

INRAE plant pathologists in the field, and we named this dataset *INRAEdi*. The collection consists of 4 well-known plant diseases which are listed in the Table II. By comparing *INRAEdi* with the aforementioned database listed in Table I, we note that the class *Zucchini__Podosphaera_xanthii* is new but its individual species and disease classes exist in the database listed in Table I. In fact, this condition is somewhat consistent with the problems encountered in the actual scenario where complete species and disease data would not always be available when training a model. Although the *INRAEdi* data is not large, it allows us to assess the effectiveness of a model in transferring knowledge from one disease context to another.

V. EXPERIMENTS AND RESULTS

To test the concept of the CMTL mechanism, we built our network upon Inception-V3. The final activation layer of Inception-V3 were modified according to the different strategies proposed as explained in Sec. III. We used random cropping of images during pre-processing for training as well as mirroring and color data augmentation. During model evaluations, each test image was cropped in the center. All images were resized to 299×299 . We trained our models using the Tensorflow library [29]. The networks were trained with back-propagation, using stochastic gradient descent. For the training parameter setting, we employed the fixed learning policy. We set the learning rate to 0.01, and then decreased it by a factor of 10 until the validation set accuracy stops improving. The momentum was set to 0.9 and the weight decay to 0.0001. In all experiments, we used a mini-batch size of 45. We run the experiments using an NVIDIA GTX1080 graphics card.

Baseline models In this section, we analyzed the classification performance of the proposed models and compared them to three reference models: (1) Generic MTL (Fig. 2a), (2) single nets (Fig. 2b) and (3) joint net (Fig. 2c). To model generic MTL, we replaced the final activation layer of Inception-V3 with two fully connected branched layers for the classification of diseases and host species. Specifically, having the same shared layers as our proposed models, each task-specific layer is composed of an individual task embedding layer and a logit layer. On the other hand, single nets are two distinct models formed with different tasks, the classification of host species and diseases. In this model, we replaced the final activation layer of the Inception-V3 with a new logit

layer corresponding to the respective task. Next, joint net is a network formed using conventional modeling of joint species-diseases, which is widely used by [15], [4], [5], [6], [16]. For this model, we replaced the final activation layer of the Inception-V3 with a new logit layer to classify the joint species-disease classes. Particularly, in order to obtain a fair comparison, all trained models were built on Inception-V3 to obtain the same model depth. Next, we also compare our proposed models to the classification model [12] (Fig. 2d) that was designed to solve two closely related tasks of visual fashion analysis. In spite of the tasks are not in the field of plant diseases but the characteristic of the targeted problem aligned with the problem we intended to solve. The details of this architecture will be explained later in the sub-section ‘Compared to the state-of-the-art’. We called this architecture Fashion net.

Evaluation methods To measure the model classification performance on the validation set, we used the top-1 accuracy in our experiments and also the average accuracy per class (Ave) to assess the performance of every individual class. Besides evaluating the classification performance of each task (the identification of infectious diseases and hosts species) itself, we also measured the extent to which the models can correctly predict the conjoint species-disease classes. It should be noted that with the exception of the joint net with immediate probability output for each class of joint species-disease, other modeling techniques require post-calculation to obtain this conjoint prediction. To do this, we multiply the probability outputs of the host species and disease, and measure the ranking based on the 1146 joint species-disease labels. Table III shows the comparison results.

As for the evaluation of the testing set, the prediction of models takes into account all possible combinations, which corresponds to the *total number of diseases* \times the *total number of host species*. However, it should be noted that the joint net prediction outputs are limited to 1146 joint species-disease classes. Hence, to test the ability of the joint net to identify all possible combinations of host species and diseases, we first extract prediction for each host species and disease by calculating the average probability output of the conjoint classes associated with them, then multiply the probability outputs of the host species and disease.

Performance analysis on validation set From Table III, it can be seen that the conditional FFS-res achieved the highest accuracy for all predictions. In particular the conjoint prediction which reached 69.43%, surpassing the joint net [15], [4], [5], [6], [16], [28] that was widely used before, which reached only 68.82%. In addition, compared to all the baseline approaches, the conditional FFS-res showed a highest accuracy for both diseases and host species. This indicates that characteristics learned using the conditional FFS-res are more robust, contributing to the highest optimal prediction for each individual classification task. Next, we find that identical mapping (-res) could help improve the

TABLE III
CLASSIFICATION RESULTS OF DIFFERENT MODEL CONFIGURATIONS. WE DENOTE THE NEW METHODS OF CONDITIONAL FEATURE FUSION: SUM AND CASCADE FUSION BY CONDITIONAL FFS AND FFC RESPECTIVELY, AND IDENTITY MAPPING IN THE FORM OF ‘RES’.

Model	Conjoint prediction		Diseases	Host species
	Top-1	Ave	Top-1	Top-1
conditional FFS-res	69.43	64.79	82.96	78.64
conditional FFS	67.34	61.96	81.59	77.52
conditional FFC-res	68.06	62.58	81.84	77.67
conditional FFC	67.19	62.55	81.38	77.11
conditional FiLM-res	68.01	62.97	82.15	76.91
conditional FiLM	68.62	64.39	82.86	77.16
Generic MTL	68.16	62.91	82.76	76.96
Single nets	68.67	63.10	82.25	78.48
Joint net	68.82	64.27	82.15	76.70
Fashion net [12]	66.48	61.53	81.49	76.09

performance of approaches based on conditional FFS and FFC. For example, in the case of conditional FFS-res and FFC-res, each has an increase of 2.09% and 0.87% in conjoint prediction when compared to the conditional FFS and conditional FFC respectively. However, this is not the case of conditional FiLM. This may be due to the fact that FiLM has introduced a certain amount of conditioning by means of affine transformation of features. The introduction of identity mapping after this transformation to deviate the features again may not produce a better response. More conditional hidden layers are probably needed as depicted in [30] before introducing the identity mapping.

Compared to the state-of-the-art Many areas of research use MTL to improve the application concerned. Among these areas, the problem addressed in [12] on the classification of clothing categories is more relevant to the problem we intend to solve. Specifically, they addressed two classification tasks that consist of classifying clothing categories and clothing attributes, and these two tasks are strongly related to each other. Unlike our proposed CMTL and the baseline generic MTL, the MTL proposed in [12] does not have an individual task embedding layer before the logits layer. To study the impact of this architecture, we imply the same concept in modeling the identification of plant diseases, without involving the attention mechanism in order to have a fair comparison to our proposed model, we called it Fashion net. The Table III shows that with the proposed CMTL methods and even with the generic MTL, the performances are better than the Fashion net. This therefore indicates the importance of task specific embedding layers for each task.

Prior knowledge assessment In the current CMTL approach, we assume a unidirectional distribution of prior knowledge, i.e. information about the species is injected into the disease layer or we can say that the disease characteristics are species-dependent. However, it is possible that from the disease information we can link to possible crops. In order to determine whether different ways of linking knowledge

TABLE IV

COMPARISON RESULTS OF THE CONDITIONAL FFS-RES APPROACH FOR DIFFERENT KNOWLEDGE DISTRIBUTION MODELS. NOTE THAT $S \rightarrow D$ STANDS FOR DISEASE LAYER CONDITIONED ON SPECIES AND $D \rightarrow S$ FOR SPECIES LAYER CONDITIONED ON DISEASE.

Model	Conjoint prediction		Diseases	Host species
	Top-1	Ave	Top-1	Top-1
$S \rightarrow D$	69.43	64.79	82.96	78.64
$D \rightarrow S$	68.51	63.02	82.81	77.01

TABLE V

THE TOP-1 CONJOINT PREDICTION FROM DIFFERENT MODEL CONFIGURATIONS ON EACH DATASET: (1) PV, (2) DIGIPATHOS, (3) IPM, (4) PL@NTNET AND (5) INRAEDI

Model	PV	Digipathos	IPM	Pl@ntNet	INRAEdi
conditional FFS-res	93.84	85.80	64.16	61.99	8.76
conditional FFS	91.10	86.09	61.70	59.28	16.49
conditional FFC-res	94.52	86.38	61.82	60.33	16.24
conditional FFC	94.52	84.64	60.47	60.33	13.40
conditional FiLM-res	91.78	85.22	63.67	59.13	7.99
conditional FiLM	93.15	83.77	65.27	59.43	9.28
Generic MTL	93.15	85.51	62.68	60.33	9.54
Single nets	93.84	84.35	63.18	61.69	2.06
Joint net	91.10	85.51	65.02	59.88	6.96
Fashion net [12]	90.41	84.64	61.70	57.62	13.40

introduce a difference in performance, we therefore tested the knowledge transfer of the other cycle, i.e. the injection of disease information into the host species. Table IV presents the results of the comparison. It can be seen that the performance of $S \rightarrow D$ modeling is still better compared to the $D \rightarrow S$. This could be due to the fact that the number of classes for species is higher, resulting in a greater diversity of learned features in the species layer, which could provide more information to the disease concept. In addition, it is also possible that having an initial knowledge of the species may be easier to infer disease than the other way around.

Performance of individual dataset We compare the performance of each dataset detailed in the Sec. IV and show in Table V. We can see that, for the validation sets, in general, all models were able to obtain a better classification result (greater than 80%) in Digipathos and PV compared to IPM and Pl@ntNet. This is due to the fact that the images captured in Digipathos and PV come mainly from an easy and less noisy background than in IPM and Pl@ntNet where the images are mainly captured in fields cluttered by a complicated and noisy background.

As for the INRAEdi testing set, we can see that the overall performance of the models has decreased. We found that most of the wrongly classified samples have the prediction results for the disease classification biased in favour of the “healthy” class because the training images for the healthy class are comparatively larger compared to the other disease classes. This inevitably shows that the class imbalance in the training data had an impact on the generalization of the models. Besides that, the conjoint host species and disease classes that has not been seen by the model during training was found

also another factor that affects the classification performance. However, we notice that the overall performance of the host species is generally much higher than that of the diseases. This is believed to be due to the fact that the host species classifier is more generalized and that the characteristics of the model to differentiate host species are more distinctive in relation to those of diseases, or conversely, that the visual appearance of diseases is more variable. Apart from that, it can be seen that the single net and the joint net are clearly lagging behind compared to those MTL based models, showing the importance of MTL in modelling plant disease identification to capture not only individual characteristics but also the mutual characteristics between each concept/task.

From Table V, although we could not conclude which CMTL based approaches perform best because each approach achieved the highest accuracy in different datasets, we show that CMTL-based approaches generally outperform baselines, showing a significant improvement in the identification of plant diseases.

VI. CONCLUSION

This paper has shown that a multi-task learning approach is more efficient than the usual CNN modeling method, the joint net, in plant disease identification. Furthermore, with the CMTL conditional learning mechanism, we could achieve better performance, which could serve as a reference for researchers working in the same field. The future direction of our research is to address the problem of class imbalance that resides in the current dataset and to explore possible alternatives to improve CMTL methods to deal with it. In addition, further research will be carried out to explore new alternatives to solve problems where knowledge of host species and diseases is limited during training and where unknown classes may appear during testing. Finally, our new dataset covering more host species and diseases than the PV [4] and Digipathos [28] is obviously more challenging, which encourages researchers in this field to continue to venture further.

ACKNOWLEDGMENT

This research project was supported by Agropolis Foundation, Numev, Cemeb, #DigitAG, under the reference ID 1604-019 through the “Investissements d’avenir” programme (Labex Agro:ANR-10-LABX-0001-01). This work was also supported by the French National Research Agency under the Investments for the Future Program, referred as ANR-16-CONV-0004.

REFERENCES

- [1] M. B. Riley, M. R. Williamson, and O. Maloy, “Plant disease diagnosis,” *The Plant Health Instructor*, vol. 10, 2002.
- [2] M. Bhang and H. Hingoliwala, “Smart farming: Pomegranate disease detection using image processing,” *Procedia Computer Science*, vol. 58, pp. 280–288, 2015.
- [3] Y. Fang, R. P. Ramasamy, and G. C. Rains, “Current and prospective methods for plant disease detection,” in *Biosensors*, vol. 5, no. 3, 2015, pp. 537–561.

- [4] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in plant science*, vol. 7, p. 1419, 2016.
- [5] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture*, 2018.
- [6] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018.
- [7] S. Prospero and M. Cleary, "Effects of host variability on the spread of invasive forest diseases," *Forests*, vol. 8, no. 3, p. 80, 2017.
- [8] I. J. Deary, L. Penke, and W. Johnson, "The neuroscience of human intelligence differences," *Nature reviews neuroscience*, vol. 11, no. 3, p. 201, 2010.
- [9] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [10] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in *CVPR*, 2014.
- [11] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 2066–2073.
- [12] W. Wang, Y. Xu, J. Shen, and S.-C. Zhu, "Attentive fashion grammar network for fashion landmark detection and clothing category classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4271–4280.
- [13] S. H. Lee, H. Goëau, P. Bonnet, and A. Joly, "New perspectives on plant disease characterization based on deep learning," *Computers and Electronics in Agriculture*, vol. 170, p. 105220, 2020.
- [14] J. Boulent, S. Foucher, J. Théau, and P.-L. St-Charles, "Convolutional neural networks for the automatic identification of plant diseases," *Frontiers in plant science*, vol. 10, 2019.
- [15] M. Arsenovic, M. Karanovic, S. Sladojevic, A. Anderla, and D. Stefanovic, "Solving current limitations of deep learning based approaches for plant disease detection," *Symmetry*, vol. 11, no. 7, p. 939, 2019.
- [16] M. TÜRKÖĞLU and D. Hanbay, "Plant disease and pest detection using deep learning-based features," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 27, no. 3, pp. 1636–1651, 2019.
- [17] Y. Toda, F. Okura *et al.*, "How convolutional neural networks diagnose plant disease," *Plant Phenomics*, vol. 2019, p. 9237136, 2019.
- [18] S. Ruder, "An overview of multi-task learning in deep neural networks," *arXiv preprint arXiv:1706.05098*, 2017.
- [19] Y. Zhang and Q. Yang, "An overview of multi-task learning," *National Science Review*, vol. 5, no. 1, pp. 30–43, 2018.
- [20] —, "A survey on multi-task learning," *arXiv preprint arXiv:1707.08114*, 2017.
- [21] R. Caruana, "Multitask learning: A knowledge-based source of inductive," in *Proceedings of the Tenth International Conference on Machine Learning*, pp. 41–48, 1993.
- [22] —, "Multitask learning," *Machine learning*, vol. 28, no. 1, pp. 41–75, 1997.
- [23] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7482–7491.
- [24] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville, "Film: Visual reasoning with a general conditioning layer," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [25] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2337–2346.
- [26] P. Oza and V. M. Patel, "C2ae: Class conditioned auto-encoder for open-set recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2307–2316.
- [27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [28] J. G. A. Barbedo, L. V. Koenigkan, B. A. Halfeld-Vieira, R. V. Costa, K. L. Nechet, C. V. Godoy, M. L. Junior, F. R. A. Patricio, V. Talamini, L. G. Chitarra *et al.*, "Annotated plant pathology databases for image-based detection and recognition of diseases," *IEEE Latin America Transactions*, vol. 16, no. 6, pp. 1749–1757, 2018.
- [29] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 2016, pp. 265–283.
- [30] E. Perez, H. De Vries, F. Strub, V. Dumoulin, and A. Courville, "Learning visual reasoning without strong priors," *arXiv preprint arXiv:1707.03017*, 2017.