IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 29 February 2024, accepted 9 May 2024, date of publication 17 May 2024, date of current version 29 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3402213

RESEARCH ARTICLE

Efficient Application of Deep Neural Networks for Identifying Small and Multiple Weed Patches Using Drone Images

VYOMIKA SINGH[1], (Graduate Student Member, IEEE), DHARMENDRA SINGH[1], (Senior Member, IEEE), AND HARISH KUMAR[2]

[1]Department of Electronics and Communication Engineering, Indian Institute of Technology Roorkee, Roorkee 247667, India
[2]Department of Computer Science, College of Computer Science, King Khalid University, Abha 61413, Saudi Arabia

Corresponding author: Dharmendra Singh (dharm@ec.iitr.ac.in)

This work was supported in part by the Drone Research Centre, IIT Roorkee; and in part by the Deanship of Scientific Research at King Khalid University, Abha, Saudi Arabia, under Grant RGP.2/229/44.

ABSTRACT Deep learning (DL) based camouflage target localization, classification and detection using UAV data is currently an evolving and promising hotspot in the field of digital image processing and computer vision. Weed, an undesirable plant that hinders crop growth, acts as a camouflage target w.r.t sugarcane crop (*Saccharum Officinarum)*. Weed exhibits green-on-green color-based merging capability (background matching) and similar spectral behavior with the sugarcane crop, resulting in a visually complex camouflage habitat, making it difficult to distinguish and detect weed from crop. The concealed nature of weed in the sugarcane field, qualifies it as a camouflage target. In this scenario, weed is distributed in the form of multiple small patches across the UAV imagery, making the detection even more complicated, leading to a requirement of a pixel-based classification and subsequent target detection technique. The research problem is to successfully detect weed (camouflage target) at the smallest resolvable pixel size (2-4 cm/px), keeping in mind the similar behavior (merging color) and its difficulty to differentiate it from the sugarcane crop (background). To achieve this challenge, color and texture are exploited as important feature representations. They are extracted via UAV image(s), which can aid in small and multiple weed patch detection, by implementing a rich feature-based Deep Neural Network (DNN). In this revolutionary era of AI and modern drones, DNNs provide feature based elastic transformations, expedited processing, improved accuracy, and optimized mapping and detection. These representation-based networks traverse deeper into color and texture feature representations as a unified component, pixel by pixel, eventually detecting the camouflage targets by their self-learning capabilities in comparison to traditional classification approaches. DNNs exhibit good capabilities to solve camouflage target detection problems, therefore we have proposed a methodology based on deep learning feature-based modelling which generates classified and localized image maps that highlight multiple and small weed patches at the minimal pixel level possible from a merged green-on green background. We have explored and proposed a methodology that self learns exhaustive features from the UAV imagery via feature transformations which in turn reduces the feature bias and exhibits enhanced prospects w.r.t sensitivity of imagery background. This eventually results in improved localization and classification of camouflage targets with respect to traditional target detection and localization modelling techniques in natural background. It is observed that the proposed methodology is capable of detecting small patches of weed (similar color behavior) from its background with an accuracy of 90.5%.

The associate editor coordinating the review of this manuscript and approving it for publication was Jeon Gwanggil.

KEY HIGHLIGHTS

• Development of an adaptive and efficient algorithm based on color and texture features which classifies the study area into two classes (binary classification): weed and non-weed, eventually detecting small and multiple weed patches from RGB UAV images. • Proposing and formulating a self-learning framework trained to detect multiple weed patches of varying sizes, with promising visual detection in varying conditions such as: (a) saturation intensity, (b) contrast, (c) shadow, and (d) luminance conditions in the imagery w.r.t ground truth. • Critical comparison of the network with notable state of art algorithms from literature such as support vector machines (SVM), random forest classifier (RF), 3-layer multi-layer perceptron (MLP) and 2-D ConvNet on the basis of an array of post-classification metrics w.r.t ground truth.

INDEX TERMS UAV, deep neural networks, classification, target detection, computer vision.

I. INTRODUCTION

Camouflage target identification and detection in natural background by localizing and classifying modern drone data, i.e. multi-spectral high-resolution sensor(s) equipped UAV (unmanned aerial vehicle) data, especially by machine learning based parameterized classifiers is an intricate and intuitive process which happens to be one of the active hotspots in the cumulative fields of drones, image processing and computer vision. In this study, UAV acquired aerial imagery of a sugarcane field with weed as a camouflage target, which is merged with green-on-green sugarcane crop (background) is the visually complex scenery which makes target localization and classification typical and difficult. Classification and localization of real-time captured, high-resolution UAV datasets requires good consideration of relevant and related factors. More complexity is automatically appended if the target is camouflage in nature, i.e. a type of target, which is concealed, by merging in the background. The camouflage nature of the target makes it hard to identify, detect and eventually classify. Relatively similar spectral behavior of the target (foreground) with the surroundings (background) creates false illusions such as disruptive camouflage (breaking/bending shape) [1], background matching (blending/merging with the background), masquerade (copying other target/object), and dazzle camouflage exhibited by animals, humans and man-made targets on purpose or naturally [2], [3], [4]. Indistinguishable visual nature of embedded and rapidly spreading weed (unwanted plant growth) in a sugarcane field, makes this imagery a background-matching camouflage target template [5]. The nature of weed patches is varying in sizes (smallest being in the range of 2-4 cm/px) and is multiple in number. Detecting and classifying these concealed targets into concise and accurate classes with distinguished class boundaries, independent from explicit programming but rather, trained on an exhaustive feature set is an emerging area of research these days. Finding relevant features in the imagery belonging to respective classes, labelling them and training the visual facets of imagery via deep learning algorithms is one of the active research domains, which assists such problems with automatic feature extraction and learning, robustness towards varying natural factors and flexible architecture which can be tuned according to the research problem statement. Deep learning has found commendable findings and applications in object detection and classification, precision agriculture, automatic disease detection, semantic segmentation, data clustering and target detection and monitoring.

II. RELATED WORK

Weed, in the sugarcane field, competes for nutrition, water, sunlight and growing space, making it tough for the original crop to flourish, resulting in a probable bad produce with respect to otherwise ideal conditions. Traditional weed management techniques constitute financial burden, man power and are eventually, time-based investments [6], [7]. An automated detection scheme via binary classification and localization can aid in swift and factual weed detection, control and management, finding its practical applications in small-object detection, precision agriculture (early weed detection) and eventually multiple camouflage target(s) detection. Discovering separability between two modalities (weed and non-weed areas), with relatively similar signature requires learning patterns of spatial association about target-background item-item co-occurrence via efficacious pixel-level feature extraction [8], which in turn aids in improved feature set and network training for deep learning-based classifiers. In addition to the factors discussed above, there is a need of a fault tolerant, feature rich algorithm which works on multi-target detection, effectively. The variation in appearance of the target, such as color, merges the camouflage target with the background. Pressing requirement is of an efficient target detection scheme that successfully segregates the target from the background with less bias in feature selection choice, and minimum background knowledge. The complexity of the images also makes target detection relatively difficult. Along with the fact that distinguishing and localizing camouflage from its natural background itself is a huge challenge, there also happens to be false alarm, misclassification, poor detection rate, that needs to be compensated by a thorough post-classification metric assessment against the ground truth.

Deep learning, a subclass of machine learning, allows complex and computational training of high dimensional data inputs with multiple levels of abstraction, by layering several self-learning algorithms [9]. The basic underlying concept of

deep learning lies in ''representation'', i.e. discovering and extracting meaningful pattern by feeding ''raw data'' via a feature extractor for data transformation(s) to finally represent a meaningful representation unit such as a feature vector, classifier or detector [10], [11], [12]. Deep learning classifiers are hence, representation-based learning algorithms with channeled levels of data sets using general-purpose learning procedure with substantial higher accuracy than traditional algorithms [9], [13], [14]. Modelling non-linear, varied representation units, on a high-resolution dataset is the beauty of deep neural networks, where the network increments the complexity and automates feature selection, as the network progresses. Deep Neural Networks (DNNs) have revealed exceptional performance metrics on image classification and localization of targets, provided computational needs are fulfilled, as these algorithms have relatively strenuous and tedious training periods, hence demanding specialized graphics processing units (GPU) for its protracted processing. The problem-solving approach is also more refined owing to step by step implementation and analysis. Hence, it has been observed that in a high-resolution UAV data that possesses a complex, concealed target problem with less information and separability about class-wise feature domain, deep learning plays a significant role in localization, classification and eventually camouflage target detection [15]. Popular machine learning techniques such as XGBoost [16] have been utilized for rich feature extraction such as spatial information, intensity and texture features to detect and discriminate in vegetation targets but lacks minimal pixel detection [17], whereas color and area based image filtering techniques for weed detection in vegetable crops using machine vision has been also been explored in recent works [18], regression based classification to map crop planting quality has also been widely explored and implemented giving meaningful insights in the application of computer vision in such complicated problem statements [19]. SegR-Net, a deep learning algorithm based on rich feature extraction and multiscale feature fusion for retinal vessel segmentation utilized deep feature magnification, feature precision and interference for accurate segmentation masks [20]. Various CNN architectures such as ResNet50, MobileNet were explored to classify weeds in soybean crops successfully and exhibited great potential in using deep learning as a tool for camouflage target detection [21]. Literature also indicates towards a greater potential in terms of accuracy, target detection and reduced false alarm rates when machine learning based techniques are used for such problem statements rather than traditional techniques [22]. Certain classifiers inspired by object detection techniques such as YOLO (You only look once) [23], which explores unified architecture in a single pipelined network, hereby, also gaining a lot of popularity in deep-learning based real-time weed detection from cotton and other crops [24]. It can be concluded that object/target detection using DNNs via classification and localization scheme, primarily depends on the depth of how information rich ''object/target representations'' are and the overall choice and design of the network implemented. It is preferred that though the scheme is tailor-made and optimized according to the problem statement, there should still be wise choice of feature representations and self-learning adaptability of feature extraction and feeding for the network. Practically, deep learning in this study, incrementally learns feature set extracted from UAV acquired dataset from low-level to high-level categories. Ability to traverse information pixel-wise and learn complex patterns to localize and classify targets is the beauty of deep learning algorithms. In comparison to traditional image recognition techniques, where feature set needs to be extracted manually and identified by an expert, deep learning automatically chooses rich feature set according to the necessary outcome.

In this paper, we exploit the immense potential DNNs exhibit, by proposing and formulating a robust framework that (1) accurately classifies, localizes and detects large number of small and multiple weed patches of varying sizes in merged background (w.r.t sugarcane crop) with (2) promising visual detection on high dimensional multi-spectral UAV images with variations in (a) saturation, intensity, (b) contrast, (c) shadow, and (d) luminance conditions in the imagery w.r.t ground truth (3) validated by an array of post-classification performance metrics. This research explicitly classifies the problem area into two classes (binary classification), and composes a scheme which outputs a binary mask, exhibiting multiple weed patches successfully. This paper is divided into seven sections as follows: Section (3)provides an in-depth insight into the UAV data, study area of our research, and data acquisition approach (flight plan, testing and implementation, pre-processing and orthomosaic map generation). Section IV presents a systematic overview of technical concepts and current state of art, which acts as the theoretical anchor to conceptually support the paper. Section V describes the comprehensive methodology for weed detection (processed data, data labelling, network design and optimization, implementation along with accuracy assessment and information presentation). Results and detailed analysis of performance metrics are described and discussed in Section VI. Finally, Section VII concludes the paper with potential future scope.

III. STUDY AREA AND DATASET PREPARATION

This section describes the study area, data description and highlights the target of interest viz. weed w.r.t our problem statement, which will be used for developing and validating our proposed methodology.

A. STUDY AREA

Study area for our research problem composes of a weed-infested sugarcane field in the Dhanauri region (29°55'52.56''N, 77°57'58.85''E, ref system datum WGS84) of Roorkee, Uttarakhand, India, as depicted in **Figure 1(a)**. The topography of the study area is relatively flat. The imagery constitutes of multiple and varying sized green

patches as marked and depicted by red circles in **Figure 1(b)** that conceals with the background viz sugarcane field. Both the plants are green in color, hence making the unwanted weed exhibit merging properties with the background, therefore qualifying as a camouflaged target (background matching). Weed as a camouflage target mixes with the background (sugarcane crop), making target search and detection typical and challenging.

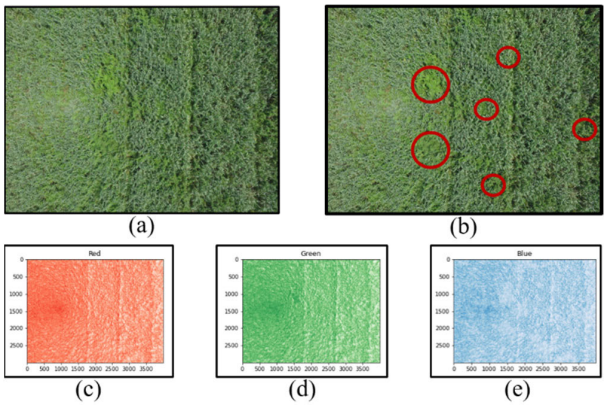

FIGURE 1. (a) Study area, (b) highlighted weed (target) patches in sugarcane field (background), (c) red band, (d) blue band and (e) green band of UAV image.

B. DESCRIPTION OF DATA

High resolution, airborne image dataset constituting of original tristimuli, i.e., red (R), green (G) and blue (B) bands was acquired using a professional DJI Phantom 3 drone mounted with a completely integrated 4K camera sensor, as displayed in **Figure 1 (c), (d), and (e)**. The DJI Phantom 3 Professional Edition quadcopter also has its own in-built imaging system. A dedicated flight plan for in-situ image and ancillary data acquisition along with its associated deliverables was planned and mapped with consideration for our area of interest, which is primarily a weed-infested sugarcane field. A UAV image's spatial resolution indicates the degree of detail that can be resolved in the image and is usually expressed in terms of the area of ground that a single pixel of an image can cover, also known as Ground Sampling Distance (GSD). The altitude of the UAV, the characteristics of the sensor, and the flight parameters are only a few of the variables that affect the spatial resolution, i.e., GSD of a UAV image. Keeping in mind our problem statement, we finalized the ideal pixel size range that may aid in resolving multiple and smallest weed patches. Finally, an acquisition of 261 images (median of 42807 key points per image) at a fixed altitude of 50 meters, with an effective pixel size (GSD) of 2.13 cm, which may be the optimum size to detect multiple weed patches according to our problem statement was conducted. UAV flight plan parameters for this dataset are shown in **Table 1**, which is a composition of a proper scheme to determine flight path, flight height (h) and image capture rate based on imagery needed, effective pixel size (GSD) and type of sensor used.

TABLE 1. Flight plan for UAV acquisition.

Flight plan parameters	Description/Values
Type of Project	Camouflage target weed detection
Type of Terrain	Agricultural field with green weed merged/concealed in sugarcane crop
Camera	DJI Phantom 3 Professional
Sensor specifications	Multispectral camera with R, G and B channels, 1/2.3" CMOS
Image Dimensions ($I_w \times I_h$)	(4000 × 3000) pixel(s) cross-sections
Focal length of sensor (f)	3.61 mm
Ground sampling distance (GSD)	2.13 cm/px
Height of the flight (h)	50 m
Number of pixels	12.4 M pixels effective (12.76M total)

Multiple green on green (weed on sugarcane) patches of varying size were observed in the orthophotos (4000 3000 pixels distribution) collected, with the effective pixel size of the orthomosaic image being *2.13 cm/pixel*.

C. DATA PREPROCESSING

A high resolution, low value GSD orthomosaic that covers the sugarcane field to capture multiple weed patches at a very fine spatial resolution is the dataset requirement. For the purpose of effective and precise data acquisition and data quality, a detailed flight plan and data preprocessing scheme is needed. Transforming the raw data (obtained from DJI Phantom 3 professional quadcopter) into meaningful and integrated data points is known as data pre-processing. It describes the steps taken to prepare data for visual analysis, such as cleaning, converting, and transforming it. Enhancing the quality of the data and tailoring it to the problem statement are the objectives of data preprocessing. In our case, high-resolution UAV data set acquisition and preprocessing is performed using DroneDeploy [25] and Pix4D [26] tools.

Step 1: DJI Phantom 3 drone is calibrated and pre-requisite planning on DroneDeploy [25] is initiated.

Step 2: Keeping and assuming flight height (H = 50m) to be constant during flight, and finalizing study area's geographic coordinates, starting and end point, data acquisition is launched.

Step 3: Post-acquisition, parameters such as calibration (camera model optimization), geolocation information status (availability of GPS/GCP), KMZ band order (blue, green and red), and co-ordinate system (set to WGS84) are optimized.

Step 4: Using Pix4D, data pre-processing steps are initiated, such as, a.) initial processing b.) match grid + point cloud and c.) DSM resolution.

Step 5: Noise reduction, lens distortion correction, band registration, radiometric correction are the pre-processing steps performed to acquire the georeferenced high resolution RGB orthomosaic images.

Step 6: UAV orthomosaic processed and effective pixel size of drone images is calculated by computing Ground Sampling Distance (GSD), which describes the distance between two consecutive pixel centers [39]. High resolution drone dataset is ready and the processed image is up for interpretation, visually and/or digitally or electronically, to extract information about the target which was illuminated

Higher spatial resolution implies that smaller objects and features can be detected and identified in the image accurately [40]. To acquire and process these high-resolution images, DroneDeploy, Pix4D and a dedicated GPU setup (NVIDIA Quadro P2200) is used in this study.

IV. THEORETICAL BACKGROUND

Our problem statement comprises of weed (target) merged with sugarcane crop (background), in different shapes and sizes, making localization and classification difficult. Detecting multiple and variating weed patchesof different intensity variations of color, especially in early stage is quite typical too. Imaging modality for our problem statement is *unmanned aerial vehicle* (UAV) dataset, popularly known as "Drone" [27]. High resolution UAV digital images are acquired without actually being in direct contact with the target; by sensing or recording reflected energy and subsequently surveying, processing, analyzing and utilizing the acquired digital information. Exploiting this imaging modality by using problem-specific computative designed machine-learning based algorithm(s) to detect merged targets is feasible due to the technical advantages drone surveying exhibits, such as reduced field time and survey costs, increased accessibility along with detailed and comprehensive high-resolution data collection. Data acquisition by modern drones is to-the-point accurate w.r.t the flight plan and has deliverables such as orthomosaic photo maps, point clouds, digital surface/terrain models (DSM/DTM), contour lines, etc. The concerned deliverable for our study are *UAV frame raster* and *orthomosaic maps*, which is a digitally stitched, high resolution raster made by combining multiple captured images (also called ortho-rectification), which are free from lens distortion or anomalies such as camera tilt or perspective [28]. Orthomosaic images accurately depict enormous geographical areas and, in our case, exhibits our targets accurately. Accurate image acquisition and organized flight plan scheme for UAV facilitates cleaner and accurate dataset. Pixel, defined as "a two-dimensional picture element that is the smallest non-divisible element of a digital image" is the most informative source w.r.t problem statement as it provides necessary information needed to know about the difference between weed and non-weed pixels in the imagery [29]. In case of UAV imagery, ground sampling distance (GSD) is calculated as the overall effective pixel of the imagery. Higher spatial resolution, and in turn lower GSD or effective pixel size aids in resolving minute objects/targets and navigate through the intricacies of high dimensional data. Image acquisition parameters (sensor's shutter speed, aperture, ISO, image side and front overlap conditions, flight altitude), choice of GCPs, pre-processing and having access to high-performance computing plays a crucial and decisive role in orthomosaicking drone images. GSD, is henceforth, calculated by the following formulation [30]:

$$GSD_h = \frac{H \times S_h}{f \times I_h} \tag{1}$$

$$GSD_w = \frac{H \times S_w}{f \times I_h} \tag{2}$$

where:

GSD_h = GSD w.r.t height
GSD_w = GSD w.r.t width
H = flight height
f = focal length
S_h/S_w = sensor height/sensor width
I_h/I_w = image height/image width

Greater computed value, out of GSD_h or GSD_w will be the final GSD or effective pixel size of the study area. The final computed GSD for our study is 2.13 cm/px. GSD calculation is important for both aerial photography and photogrammetry, and in our case an effective pixel size of *2.13 cm/px* implies that the image represents linearly 2.13 cm 2.13 cm on the ground, which is a rich and high-resolution dataset for camouflage classification and localization. Metadata is the data or information about data which includes file name, level of quantization, number of bands, geo-referencing and pixel size, and is the primary and most contributory information source with regards to the image-based dataset acquired. The preliminary image statistics such as total pixels, image dimension and central moments such as mean, standard deviation, histogram give a glimpse into the dataset and image quality, as summarized in **Table 2.** These preliminary data information points provide an insight into the raw data and helps decide the image processing operations needed to be implemented on the dataset for further processing (training dataset generation, feature extraction and eventually feeding to the network).

Spanning through individual pixels, is the most useful preliminary method for accessing the quality and information content of the data. Visual inspection and traversing through the basic image statistics such as min, max, mean, standard deviation, image histogram values deduce that color and texture are significant and important clues for target localization and identification in our problem. Total pixels for our study are 4000 3000 pixels, with each pixel having effective size of 2.13 2.13 cm. Image band values (min and max) while navigating through both weed and non-weed areas give us a range-based value set to work on and whether the image is affected by atmospheric effects or not, while the range of histogram (max-mean) gives insight into the contrast of the image, which is quite good, in our case. Brightness values (BV) of pixels in an image depict the relation of the amount

TABLE 2. Preliminary image statistics.

UAV image statistics	Description
Total pixels (points)	Displays the total number of pixels in the image [31]
Dims	Displays the pixel dimensions of the image (Width x Height) [32]
Min/Max pixel values	Displays the minimum and maximum pixel values in every band within the range [0,255]
Mean	Displays the mean value (average) of all the pixel values within the image, in every band
Stdev	Displays the value of standard deviation of the image, in every band
Histogram	Displays the histogram in each band
Covariance	Calculates the covariance matrix, the correlation matrix, eigenvalues, and eigenvectors [33]

of radiance in watts per sq. cm striking an image detector, and lie in the range of 0-255. For an image pixel P, represented by row (i), column (j) in an array of matrix, BV_{ij} represent the brightness values of the pixel. The mathematical average of the BVs of an image is known as mean [32] represented as:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} BV_i \tag{3}$$

where:
BV = digital number of the pixel
N = total no. of pixels
i = pixel descriptor

The measure of the statistical dispersion is formulated by Standard Deviation (SD) [32], which is the root mean square of values from their arithmetic mean and is represented as:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(BV_i - \mu)^2} \tag{4}$$

where:
BV = digital number of the pixel
N = total no. of pixels
μ = arithmetic mean
i = pixel descriptor

SD values indicate the nature of clustering of data points around the mean. Role of statistical evaluation of image datasets help to curate effective inter-class relations and associations, eventually aiding in accurate feature identifiers [6]. Texture, on the other hand is an innate property which expresses the local spatial structure in the pixel, in turn describing the variation in the tone of an image. It depends on the variation of gray levels and also depends on the BV. Texture measures can be 1.) First order statistics and 2.) Second order statistics. The first set comprises of mean, entropy and variance; while the second set is constituted by co-occurrence metrics, such as mean, contrast, second moment, variance, dissimilarity, homogeneity, entropy and correlation [34]. Gray Level Co-occurrence Matrix (GLCM) is a decision-based gray-tone spatial dependency matrix, which provides deeper insight into the pixel-pixel co-occurrence relationship in terms of smoothness, coarseness and regularity [21], [22].

For an image pixel P, represented by row (i) and column (j) in an array of matrix, and (i, j)th entry in a normalized gray-tone spatial dependence matrix [34], textural features (t1-t5) are computed as follows:

Angular Second Moment (*t1*)

$$t1 = \sum_i \sum_j \{p(i,j)\}^2 \tag{5}$$

Contrast (*t2*)

$$t2 = \sum_{i,j=0}^{N-1} P(i-j)^2 \tag{6}$$

Correlation (*t3*)

$$t3 = \frac{\sum_i \sum_j (i,j)\, P\,(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{7}$$

Entropy (*t4*)

$$t4 = -\sum_i \sum_j p\,(i,j) \log\,(p\,(i,j)) \tag{8}$$

Dissimilarity (*t5*)

$$t5 = \sum_{i,j=0}^{N-1} P\,(i,j)\,|i-j| \tag{9}$$

Both color and texture can be discriminating factors to distinguish between weed and non-weed, in turn being utilized as features to train the network, though the similar color of the target (weed) and background (crop) happens to be a major challenge. Our goal is to detect small and multiple patches of weed, especially multiple patch detection to the smallest pixel size. When it comes to data modelling, different applications utilize different kinds of architectures, such as, for predicting probabilities a predictive model comes to use, while to show relationship or dependencies, a descriptive model is chosen, whereas for in-depth statistical analysis, a diagnostic model is preferred while for a yes/no or 0/1 answer, classification model is best suited. Classification is a robust process that categorizes clear, accurate class boundaries which, ideally should be scale and source independent. It should be practically oriented and comprehensive in nature. Image classification can be expanded into different forms on the basis of the research problem statement/scientific application.

A pure classification scheme constitutes of assigning a particular label/class to a part of an image that is supposed to depict varied label/class information, whereas classification with localization scheme involves particular object/target

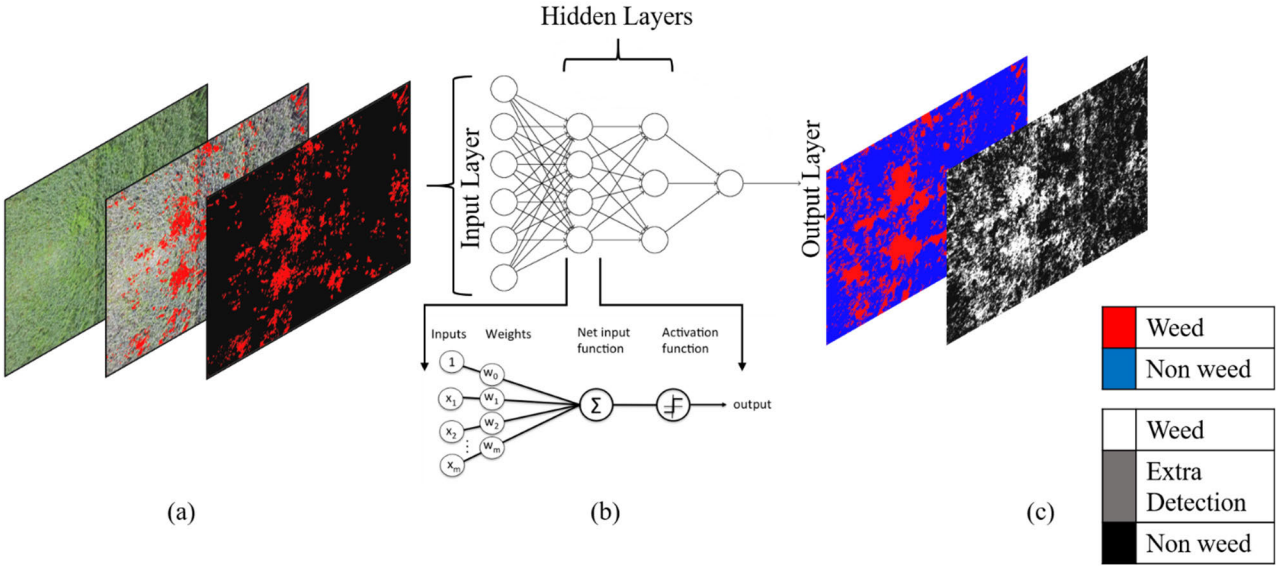

FIGURE 2. Flowchart of the proposed methodology: (a) input UAV rasters at input layer, (b) hidden layers and nodes, and (c) classified binary map(s) at output layer.

where the localization part creates a bounding box/shape mask around the object/target. Classification with object detection involves a bounding box or a polygon region of interest (ROI) which assigns specific classes to the different objects in the image. Detecting a target is hence a combination of both image classification and localization. Marking a polygon-based ROI around the target(s) shape in the input image at a pixel image level not only segments down the input image into areas containing the pixels of objects and background, separately, but this type of pixel-wise segmentation also presents us with the object's shape instead of just a rectangular bounding box enclosing the object, making our detection more convenient. A basic classification scheme constitutes of steps such as digital UAV dataset acquisition, primary visual inspection and statistical analysis, followed by necessary image enhancement techniques and operations on the dataset. Without proper pre-processing or planning, there could be data ambiguity which can lead to a low classification scheme. Our problem statement deals with the type of object(s)/target(s) that hide their signatures and merge into the natural background making it a complex and tricky task for the camouflaged target to be detected, classified and monitored. Deep learning-based target detection extracts group of rich features that group similar pixels into one class, while detecting boundaries of objects in varying shapes and sizes, eventually localizing the camouflaged target and detecting weed patches. Ability to self-learn with a labelled training set, without being explicitly programmed with clear class definition is one of the goals of this research. In the process of detecting targets, there is a need to use multi-features such as spectral, spatial & texture-based characteristics which can be efficiently used together along with training datasets [19], [23], [24]. There is an abundant layer of sensitivity with respect to background to detect weed patches such as color-based sensitivity and textural variations which may be useful to address this type of problem. There is a need for a fault tolerant and robust classifier that works on multi-target identification and localization, i.e. small objects/targets (patches) and multiple objects/targets (patches) detection with a wise choice of feature set. There also is a need of an establishment of a metric-based quality and post-classification assessment scheme for validation of the detection. Machine learning's target-background association approach can be very fruitful to accurately classify and localize small and multiple weed patches in UAV imagery.

It can overall be concluded, that there is a need of a fault-tolerant algorithm that works on multi-target identification viz, small and multiple target (weed) detection in a spatially rich dataset (UAV), keeping in consideration low feature bias and high feature extraction scheme to successfully and accurately detect camouflage targets. For validation purposes of the same, establishment of a metric based quality assessment scheme is highly encouraged. Incorporating the scientific concepts aforementioned in this section, research gaps such as occlusion, similarity in texture and color, inadequate data sets, cluttered background, error-free technique without much background knowledge are attempted to be resolved to a great extent, while simultaneously classifying and localizing multiple and varying size weed patches in the form of a binary masked image output.

V. METHODOLOGY

Proposed methodology (**Figure 2**) aims at efficient camouflage target detection using densely connected deep neural network which is trained by mapping complex data associations and features (feature engineering) via various data

transformations and feedback mechanisms. The proposed methodology of our study is divided into two sub-sections, A) dataset generation, and, B) training and implementation of designed DNN as explained below.

A. DATA SET GENERATION

1) REFERENCE DATA GENERATION

Ground truth or reference dataset in machine learning refers to the accuracy of the training set being classified, basically the ''reality'' one wants their classier/model to predict that checks the results against the real world. Data set generation via labelling is also critical as it aids in making targets identifiable and easier for the network to recognize, category/class wise. Labels/tags are of many types such as ''in-house data labeling'', ''synthetic labeling'', ''outsourcing'', ''crowdsourcing'', etc. Labels adds meaning to the dataset which aids in precise predictions and better data usability. For our problem statement, we will be using both in-house dataset generation and synthetic data generation.

Our main focus will be on curating manual polygon-based ROIs(Region of Interest) of both weed and non-weed areas. After successful ground truth labeling, weed masks were generated (to be also used for accuracy assessment later). A mask is simply an image where some of the pixel intensity values are zero, and rest are non-zero. Wherever the pixel intensity value is zero in the mask image, the pixel intensity of the resulting masked image will be set to the background value (normally zero) [41]. Mask creation on basis of visual inspection of weed and non-weed patches is done by using a set of polygon ROIs for visualization and evaluation purposes. ENVI, a geospatial tool industry standard image processing tool, is used to make manual ROIs. Data labelling via ROI masking, is cumulated in the steps below:

Step 1: Launch ENVI. Navigate visually through the image. On the basis of visual navigation, confirm weed camouflage target as ''weed'' and mark weed precisely and steadily in the image, in freeform polygon shapes using the ROI Tool [42], ENVI. Mark the vertices of the polygon or polyline, or click and drag to draw the shape. Also, assign a $Label \in \{0,1\}$ to each pixel i, where $Label_i$ represents the value of each binary pixel i. A tag value of 0 means that the pixel belongs to non-weed class (sugarcane crop), while a tag value of 1 means that the pixel belongs to the weed class.

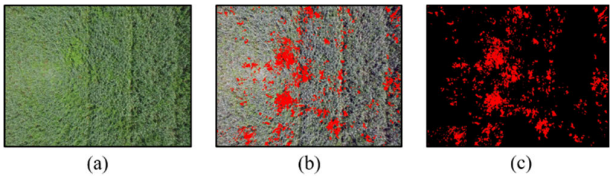

FIGURE 3. (a) UAV raster, (b) polygon ROI of weed and (c) weed mask of UAV raster.

Step 2: After labelling ROIs for both the classes, mask file and shapefiles are created, which can also be exported in ASCII format, as shown in **Figure 3**.

Step 3: To re-use these accurate yet time-consuming ground truth, images are augmented synthetically on the basis of their saturation and contrast factor, to increase the training raster dataset which shares the same ROIs and masks.

Steps 1-3, from Reference data generation, are also implemented on images with variations in (a) saturation intensity (b) contrast (c) shadow (d) luminance (illumination) conditions, as depicted in **the next section.**

2) SYNTHETIC DATA GENERATION

Synthetic dataset is annotated information that is generated artificially. A synthetically augmented dataset will enable us to re-use the time-consuming labor of manually designed ROIs again on those images, hence avoiding further data annotation. Synthetic data has been regarded as one of the most promising general techniques on the rise in modern DL, especially computer vision-based target detection techniques [43], [44], [45]. In our work, we have attempted to augment data using saturation stretch and variating contrast/intensity of the whole image by multiplying it with contrast and intensity factors respectively.

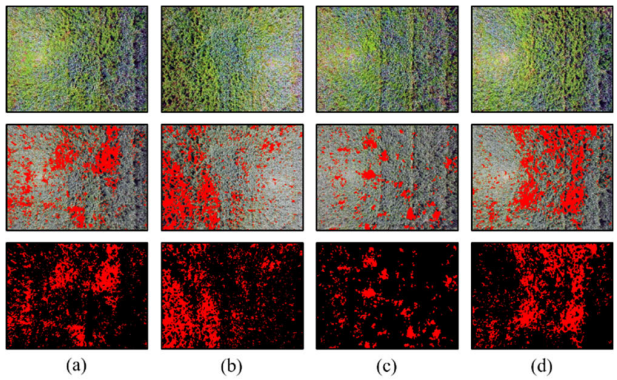

FIGURE 4. Saturation stretch of varying UAV rasters, with their weed polygons marked and masked, (a), (b), (c) and (d).

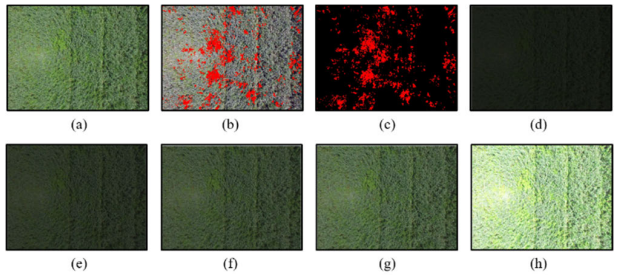

FIGURE 5. (a) UAV raster with its (b) marked weed, (c) weed mask, and augmented rasters with contrast factor (d) 0.2, (e) 0.4, (f) 0.6, (g) 0.8 and (h) 2.0.

Saturation stretch intensifies the color of a 3-band input image by performing a gaussian stretch by producing output bands that have more saturated colors [46]. The input data are transformed from red, green, and blue (RGB) space to hue, saturation, and value (HSV) space. The HSV data are

then automatically transformed back to RGB space [46], [47], as shown in **Figure 3** for various UAV rasters, along-with their ROIs with weed marked and masked respectively.

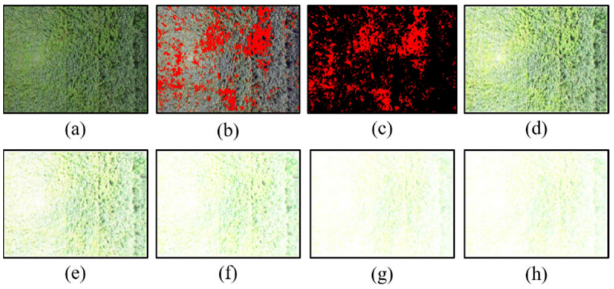

FIGURE 6. (a) UAV raster with its (b) marked weed, (c) weed mask and augmented rasters with contrast factor (d) 2, (e) 4, (f) 6, (g) 8 and (h) 10.

On the basis of varying contrast and intensity factors of contrast and saturation, augmented images can be created which will be good training data sets having same masks, as shown in **Figure 5 and Figure 6** for contrast and **Figure 7** for saturation factor.

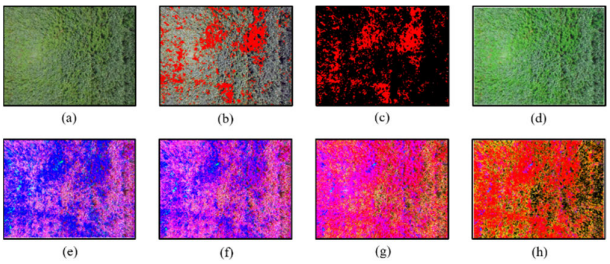

FIGURE 7. (a) UAV raster with its (b) marked weed, (c) weed mask and augmented rasters with saturation factor (d) 2, (e) 2.5, (f) 3, (g) 3.5 and (h) 4.

This sub-section aids in determining the acquisition and processing steps of UAV dataset, reference data set generation for visualization and evaluation process and synthetic data generation for data augmentation process.

3) FEATURE TRANSFORMATION

Features are the initiating parameter evaluation in a computer vision problem and feature engineering is a domain of data preparation which uses domain knowledge of data to curate and extract features that help in item-item ''representations'', which increase the overall efficiency of DNN algorithms [48]. As discussed earlier, features are marked properties or attributes which are unique to an image. Deciding the type, importance and description of features play an integral role in target detection, by encoding useful information into numerical series which can be in turn used to differentiate pixel-based clusters from one another. Features can be regions (centroid co-ordinate, size, shape, orientation), lines (equations of lines) or points (coordinate scale).

Image features are of two types: local or global, where local features aim to detect key points in the images and describe the regions around these key points, hence it is safe to say that image feature points provide valuable and rich information on image content [49]. Local image features are generally used in object recognition and identification. While, global features describe an image as a whole, and can be interpreted as a particular property, and henceforth, being generally used for image retrieval, object detection and classification. Extracted features are much low dimensional than the original image, hence reduction in dimensionality reduces the overheads of processing the images [50].

TABLE 3. Color models, with their channels and conversion w.r.t RGB color model used in the study.

Color Model	Channel	RGB Conversion
RGB	R	r = R/(R+G+B)
	G	g = G/(R+G+B)
	B	b = B/(R+G+B)
HSV	V	V = M; where, M = max {R, G, B};
	S	(M-m)/M; where M = max {R, G, B} and, m = min {R, G, B}
	H	H = {p(G-B) if M = R; 120 + p(B-R) if M = G; and 240 + p(R-G) if M = B}, where p = 60m/M
Grayscale weighted average (luminosity method)	x	x = 0.299r + 0.587g + 0.114b

Color and texture are prominent image descriptors, which are used as features in our study. A color model is a digital representation of color-based information in an imagery and on the basis of applications, different color spaces can be utilized to one's benefit [51]. Hence, color becomes an important clue or identifying tool for localization and identification of targets. These color models provide various numerical and visual representations, aiding difference in the classes by enabling analysis and study of the relationships and properties of color. RGB, as a color model is depicted as a cube by mapping red, green and blue dimensions onto the x, y and z axis in 3D space, whereas HSV is represented as a cylindrical color model that remaps the RGB into dimensions such as hue, saturation and value; H, S and V, which are interdependent values [52]. RGB to HSV color model conversion aids in sharpening, smoothening and enhancing the edges of the image characteristics and has a higher utilization rate of color information and is an attractive color model. It is suitable for scene classification, target analysis and target segmentation. It also decouples intensity component from color carrying information components – giving us a way to display colors closer to our interpretation. Gray level, on the other hand has less noise and is adequate for locating low-level feature extraction such as edges and corners. The best method for RGB to grayscale color model conversion

is the luminosity method [37], [53]. In our study, we have experimented with RGB, HSV and grayscale color spaces (**Table 3**) to extract maximum feature combinations to train the deep network. UAV raster in the RGB color space shows variations in BV in the green channel between weed and crop to some extent, while HSV channels exhibit some varied level of detected weed in certain BV ranges, making their individual bands great intensity-based features. Moving over to grayscale channels of the raster, the variation in characterization of gray tones reveal texture-based features, which is extracted using Gray-level Co-occurrence Matrix (GLCM), which in turn exhibits an array of promising features with a lot of potential to understand the weed-crop pixel distribution [54]. Gray level co-occurrence matrix calculates how often a pixel with gray-level value "i" occurs either horizontally, vertically or diagonally to adjacent pixels with the value "j", where "i" and "j" are the gray level values (tone) of the image. In terms of feature extraction, there are three groups of texture measures: 1. contrast group (contrast, dissimilarity, homogeneity), 2. measures related to orderliness (angular second moment, entropy), and 3. descriptive statistics group of the GLCM measures (mean, variance, correlation) [36].

For a given input image, a set of features f_k, $k \in \{1,2\ldots n\}$ are extracted from the backbone of the network, which works from high-resolution low semantics to low-resolution high semantics. For two classes, like our problem, two features (say X1, and X2) will have different orientations and value sets as depicted in **Figure 8** [55]. An exhaustive feature set is curated and extracted from our study area, as tabulated below in **Table 4**.

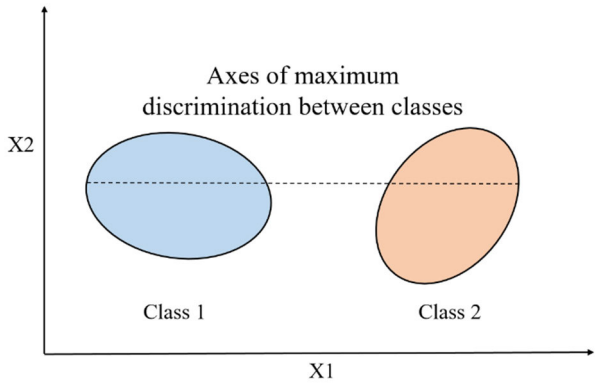

FIGURE 8. Feature set relation with classes.

Table 4 has been distributed into three major categories of feature extractors, namely, a.) Color intensity-based features, and b.) Texture: statistical (spatial distribution of gray values) features and c.) Texture: co-occurrence matrices (expression of local spatial structure) features. Intensity/color features such as R, G, B band values of raster, along with dominant color, viz, green band values play an integral role in determining the weed-non-weed color threshold by observing the variation in BV values. Similar pixel clustering for respective classes are also aided by intensity features such as mean and std deviation values of weed, non-weed classes along with color model values and interpretation in HSV and grayscale

TABLE 4. Feature list (Spectral, Statistical and Textural) for early weed detection.

Feature type	Name of the Feature	Remark
Color: intensity-based features	Red, green and blue bands of UAV raster(s)	Analysis of BV distribution in simple RGB color space: R, G and B channels for clusters and pattern recognition w.r.t weed detection [54] [53]
	Dominant color	Green band is the dominant color in weed patches which shows variation in BV and can be used as an intensity feature
	Mean values of red, green and blue bands	Avg. color band values used for similar pixel clustering [53] [57]
	Std Deviation of red, green and blue bands	Measure of statistical dispersion in different bands and identification of corelated information for similar pixel clustering [36]
	Hue, saturation and value channels of HSV color space of UAV raster(s)	HSV color space, where hue is used as a feature as it gives real time visual point of view and representation [53]
	Mean Grey Value	Description of a weed patch by its mean grey value
Texture: Statistical (spatial distribution of gray values)	Mean of histogram	1st moment of gray image describes central value [40], [58]
	Variance of histogram	2nd moment of gray image, describes measure of dispersion [40], [58]
	Skewness of histogram	3rd moment of gray image, describes asymmetry in pixel distribution
	Kurtosis of histogram	4th moment of gray image, describes peakedness
Texture: Co-occurrence matrices	GLCM mean	Soothing operation [40] [55]
	GLCM correlation	Measures the joint probability occurrence of

TABLE 4. *(Continued.)* **Feature list (Spectral, Statistical and Textural) for early weed detection.**

(expression of local spatial structure)		the specified pixel pairs [40] [55]
	Energy	Provides the sum of squared elements in the GLCM [40] [55]
	Entropy	Amount of texture in an image, in terms of disorder of the kernel values [40] [55] [41], degree of randomness
	Correlation	Correlation b/w GLCM matrices
	Contrast	Intensity contrast b/w GLCM matrices, measures the local variations in the image
	Local Homogeneity	Measures the closeness of the distribution of elements in the GLCM
	Dissimilarity Texture browsing descriptor	Degree of dissimilarity Characterization of the texture's regularity, directionality and depth. [39]
	Edge Histogram Descriptor	Records spatial distribution of the edges.

color models. On the other hand, texture features such as histogram spatial distribution of tonal gray values on the lines of mean, variance, skewness and kurtosis explicitly define class-wise central value, asymmetry and data points distributions in both the classes. Co-occurrence texture assessment gives insights into joint pixel occurrence probability, degree of randomness, local tonal variations in the images, characterization of the texture's regularity and the spatial distribution of class edges. These exhaustive features help aid feed the data with improved feature set to help segregate the classes, which in turn aids in good classification and localization of our target, i.e., weed.

This sub-section aids in determining the role of feature engineering in feeding the DNNs to make it more effective. The role of color and texture-based features have been explored and proposed.

B. APPLICATION OF DNN

1) CAMOUFLAGE TARGET DETECTION USING DEEP LEARNING MULTI-LAYER CLASSIFIER (TRAINING AND TESTING)

A deeply connected, feature-extensive computational deep-learning modelling with iterative and multiple processing layers, trained on the concept of representation learning, extracted from the UAV rasters to effectively detect small and multiple weed patches and finally deliver a binary classified map is the main theme of this paper [9]. The network design, as shown in **Figure 2**, is sequential forward in nature and starts with fully connected layers and nodes which are the building blocks of the network. The scheme takes in consideration increasing level of representations, from raw UAV images to polygon-based ROIs and complex feature representation maps. The modules of the network are simple but non-linear in nature, making it easy to learn complex mapping at the pixel-level. The network starts with the input layer, which inputs UAV rasters and their representation as an array of pixel values, moving to intensity-texture based statistical analysis, and polygon-based ROI labelling, which is the initiation of the network. As the feature inputs are numeric in nature, Keras [58], plays a significant role in data interpretation and introduces a thorough layering approach. The input influences each successive layer as it moves ahead. Feature extraction and abstract representations on the basis of intensity (color) and texture descriptors help interpret spatial and spectral distribution within the image pixels and is the optimal foundation for weed detection. At the hidden layer level, these abstractions take place, layer-by-layer, which are composed of exhaustive feature sets that are self-learning in nature and help eventually train the network. Feature sets are further, standardized, due to data-transformations sensitivity. Activation functions aid in establishing non-linearity in the network. They determine the rate of information transfer from one node to another. To increase the accuracy and overall speed of the training process and the stability of the network, RelU is preferred in our problem statement [59], formulated as:

$$f\,(x) = \max[0, x) \tag{10}$$

where, $f(x)$ is the activation function, x is the neuron input, both f(x) and derivative are monotonic, returning positive value $(0, \infty)$.

The design is a feed-forward stack of multilayers where inputs from the previous layers keep on passing to the other layer, to finally reach the output layer, where data transformations take place, resulting in class values (0 and 1). In the first run, random weights are assigned, which are then auto-appended further, also known as backpropagation [60], for which we need: a) objective function (loss function) and b) optimizer. For the purpose of binary mask classified raster as output, multi-categorical cross entropy is used. For this network, binary cross entropy [61] is preferred, formulates as:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} y_i . \log\left(p\left(y_i\right)\right) + \left(1 - y_i\right) . \log\left(1 - p\left(y_i\right)\right) \tag{11}$$

In terms of optimizer, Adaptive Moment Estimation Algorithm, commonly known as Adam is selected for our binary image classification problem owing to its adaptive

learning rate, great compatibility with targets and considering its exemplary results with large datasets, according to literature [62]. Its robust nature and expedited efficient learning rate make it an ideal optimizer for our problem statement. Overall, the proposed DNN classifier explores the features and is trained to recognize pixels belonging to their respective classes. With deep layers, a hierarchy of increased learning, abstraction and complexity is established, also known as feature hierarchy, which helps in detecting multiple weed patches of varying shapes and sizes, successfully. Post feature extraction, the following algorithm aids in camouflage target detection using deep learning multi-layer network.

Algorithm:
WeedFound ← **false**
while true do
$Label_i$ ← assign "1"
while WeedFound == **false do**
detect_weed(f)
if weed is detected **then**
WeedFound ← **true**
class ← weed
end if
end while
bg(f)
if bg is false **then**
WeedFound ← **false**
class ← non-weed
$Label_i$ ← assign "0"
end if
end while

The algorithm depicts the presence of "weed" in an area of interest if the $Label_i$ has weed present, along with "WeedFound" function which assigns label "class" to the detected area, leading to a generation of a binary masked output with "weed" and "non-weed" patches. The classified binary raster, when obtained, is further transformed as weed, non-weed and extra detection via image subtraction (using decision tree) for more further visual understanding and assessment, as shown in **Figure 9.**

VI. IMPLEMENTATION, RESULTS, AND DISCUSSION

In the case of background-merging camouflage target, such as weed merged with sugarcane crop, if observed, at the pixel level, two subsequent pixels of nearly similar spectral reflectance might belong to completely different object/feature/class. Accurately and minutely assigning pixels to their correct class label is the ultimate objective of this paper. Post-performance metrics and accuracy assessment for the classifier is the final and validating step of the proposed methodology. In terms of machine learning, the ground truth determines the accuracy of the training sets classified. In other words, classifier accuracy is achieved by comparing ground reference map/ground truth acquired from UAV rasters to the binary masked output obtained post-classification. But before jumping into the accuracy metrics, it is noteworthy to also realize and keep in mind the assumption that classification results may be mildly erroneous due to variating labelling clusters, incorrect labelling and correlation between bands or imperfect classification algorithm.

For overall improved accuracy, our algorithm/technique needs to be:

1. Accurate and fault tolerant in nature
2. Must have capability to learn spatial association in terms of target-background item-item co-occurrence resulting in distinguished weed and crop (non-weed) target localization and detection
3. Accurate target detection with binary classified maps as output
4. Reduced false alarm rate

An error matrix (confusion matrix), post-classification aids in computing the effectiveness of the DNN-based classifier w.r.t the ground truth and the reference data generated. Simply put, it is a comparative analysis in the form of a mathematical matrix which compares the ground truth values, i.e., actual values with respect to the classifier's predicted values. Based on the confusion matrix, several post-classification metrics can be computed, as shown in **Table 5**.

TABLE 5. Post-classification accuracy assessment.

Performance Metric	Formula	Value
Recall (Sensitivity)	$\frac{TP}{TP+FN}$	**0.806**
Precision	$\frac{TP}{TP+FP}$	0.899
Specificity	$\frac{TN}{TN+FP}$	**0.947**
Accuracy	$\frac{TP+TN}{TP+FP+FN+TN}$	**90.50 %**
Misclassification rate/Error rate	$\frac{FP+FN}{TP+FP+FN+TN}$	0.105 %
F-measure	$\frac{2 \times Recall \times Precision}{Recall + Precision}$	0.85
where, TP: true positive, TN: true negative, FN: false negative and FP: false positive		

Different post-classification performance metrics have different significance, such as recall or sensitivity indicates the ability of the classifier to predict and resolve the weed patches correctly. Precision on the other hand explains, that when the classifier predicts a raster, how often is the classifier correct, in determining the weed patches. Accuracy is an overall performance metric as to how often the classification is completely correct while, misclassification or error rate is how

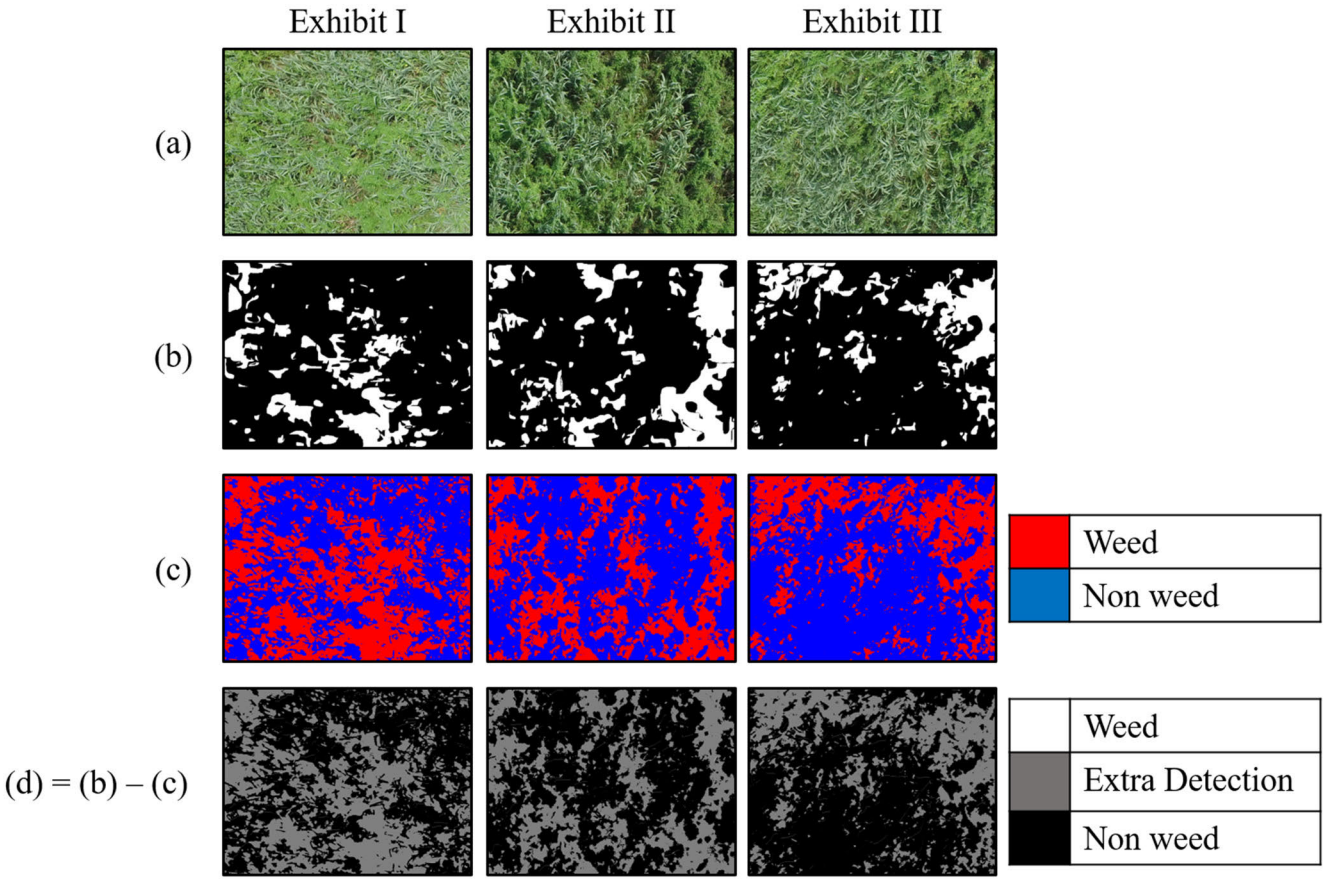

FIGURE 9. Exhibit I, II and III are (a) UAV rasters with variating contrast, saturation intensity and luminescence factors, with their (b) respective weed masks, (c) classified output and (d) spectral subtraction, i.e. classified binary mask - weed mask to highlight weed, extra weed and non-weed patches, along with legends for reference.

TABLE 6. Comparative analysis with literature.

Model	Classifier Accuracy (in percent)
Support Vector Machine (SVM) [64]	66.23
Random Forest classifier (RF) [65]	79.23
3-Layer Multi-layer perceptron (MLP) [66]	85.5
2-D ConvNet [67]	89.5
Tuned DNN to detect small and multiple weed patches	90.5

often can the classifier be wrong or misclassify. Our classifier exhibits 90.5% accuracy in detecting multiple weed patches of varying shapes and sizes and demonstrates a sensitivity of 0.806. The final output of the classifier is shown in **Figure 9** where multiple small patches of weed of varying shapes and sizes are efficiently detected even after presence of appearance variance and green-on-green background complexity.

Table 6 exhibits the comparative analysis of our fine-tuned feature-based DNNs performance with algorithms from popular literature, when implemented on the same problem set.

VII. CONCLUSION AND FUTURE SCOPE

It can be concluded that our DNN-based representation learning style classifier exhibits immense capabilities w.r.t traditional supervised learning models (**Table 6**) and maps out multiple and small patches of weed of varying shapes and sizes effectively. The ability to learn intelligently from color and texture-based features is the advantage of this classifier, which resolved weed even at minimal pixel size (2.13 cm/px), from a complex natural background. The classifier addresses a potential feature set that defines and localizes the real time classes to a respectable accuracy of 90.5%. The classifier successfully resolves camouflaged target (weed) under different illumination conditions. It is safe to say that an attempt to discover a pattern in our

dataset, and to detect camouflage target(s) in concealed background matching template has been explored in this work. Future work with regard to this problem statement can focus on acquiring and processing UAV datasets with the same problem at different heights, varying GSDs and even on different target-background imageries such as various other crops, along with improving this classifier to accommodate more natural yet complex camouflage target based scientific datasets.

ACKNOWLEDGMENT

The authors are thankful to Drone Research Centre, IIT Roorkee, for their generous support and resources for this research work.

REFERENCES

[1] H. M. Schaefer and N. Stobbe, "Disruptive coloration provides camouflage independent of background matching," *Proc. Roy. Soc. B, Biol. Sci.*, vol. 273, no. 1600, pp. 2427–2432, Jul. 2006, doi: 10.1098/rspb.2006.3615.

[2] B. G. Hogan, I. C. Cuthill, and N. E. Scott-Samuel, "Dazzle camouflage, target tracking, and the confusion effect," *Behav. Ecology*, vol. 27, no. 5, pp. 1547–1551, Jan. 2016, doi: 10.1093/beheco/arw081.

[3] S. Merilaita, "Visual background complexity facilitates the evolution of camouflage," *Evolution*, vol. 57, no. 6, pp. 1248–1254, Jun. 2003, doi: 10.1111/j.0014-3820.2003.tb00333.x.

[4] J. Skelhorn and C. Rowe, "Cognition and the evolution of camouflage," *Proc. Roy. Soc. B, Biol. Sci.*, vol. 283, no. 1825, Feb. 2016, Art. no. 20152890, doi: 10.1098/rspb.2015.2890.

[5] C. Michalis, N. E. Scott-Samuel, D. P. Gibson, and I. C. Cuthill, "Optimal background matching camouflage," *Proc. R. Soc. B, Biol. Sci.*, vol. 284, no. 1858, Jul. 2017, Art. no. 20170709, doi: 10.1098/rspb.2017.0709.

[6] V. Singh and D. Singh, "Advanced image processing approach for color-texture analysis of UAV imagery for weed detection in sugarcane crop," in *Proc. IEEE Int. India Geosci. Remote Sens. Symp.*, India, Dec. 2021, pp. 421–424.

[7] F. Xue, C. Yong, S. Xu, H. Dong, Y. Luo, and W. Jia, "Camouflage performance analysis and evaluation framework based on features fusion," *Multimedia Tools Appl.*, vol. 75, no. 7, pp. 4065–4082, Apr. 2016, doi: 10.1007/s11042-015-2946-1.

[8] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016, doi: 10.1109/TGRS.2016.2584107.

[9] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 53, no. 9, p. 7553, May 2015, doi: 10.1109/MC.2020.3004171.

[10] K. N. Qureshi, A. Alhudhaif, A. A. Shah, S. Majeed, and G. Jeon, "Trust and priority-based drone assisted routing and mobility and service-oriented solution for the Internet of Vehicles networks," *J. Inf. Secur. Appl.*, vol. 59, Jun. 2021, Art. no. 102864, doi: 10.1016/j.jisa.2021.102864.

[11] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," 2012, *arXiv:1206.5538*.

[12] A. Rehman, A. Paul, A. Ahmad, and G. Jeon, "A novel class based searching algorithm in small world Internet of Drone network," *Comput. Commun.*, vol. 157, pp. 329–335, May 2020, doi: 10.1016/j.comcom.2020.03.040.

[13] I. Ahmed, M. Ahmad, and G. Jeon, "A real-time efficient object segmentation system based on U-Net using aerial drone images," *J. Real-Time Image Process.*, vol. 18, no. 5, pp. 1745–1758, Oct. 2021, doi: 10.1007/s11554-021-01166-z.

[14] E. Akleman, "Deep learning," *Computer*, vol. 53, no. 9, pp. 1–17, Sep. 2020.

[15] *Unmanned Aerial Vehicles (UAVs): Practical Aspects, Applications, Open Challenges, Security Issues, and Future Trends|SpringerLink*. Accessed: Oct. 14, 2023. [Online]. Available: https://link.springer.com/article/10.1007/s11370-022-00452-4

[16] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[17] Y. Shendryk, N. A. Rossiter-Rachor, S. A. Setterfield, and S. R. Levick, "Leveraging high-resolution satellite imagery and gradient boosting for invasive weed mapping," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4443–4450, 2020, doi: 10.1109/JSTARS.2020.3013663.

[18] C. A. Pulido-Rojas, M. A. Molina-Villa, and L. E. Solaque-Guzmán, "Machine vision system for weed detection using image filtering in vegetables crops," *Revista Facultad De Ingeniería Universidad De Antioquia*, vol. 2016, no. 80, pp. 1–15, Sep. 2016, doi: 10.17533/udea.redin.n80a13.

[19] I. Luna and A. Lobo, "Mapping crop planting quality in sugarcane from UAV imagery: A pilot study in Nicaragua," *Remote Sens.*, vol. 8, no. 6, p. 500, Jun. 2016, doi: 10.3390/rs8060500.

[20] J. Ryu, M. U. Rehman, I. F. Nizami, and K. T. Chong, "SegR-Net: A deep learning framework with multi-scale feature fusion for robust retinal vessel segmentation," *Comput. Biol. Med.*, vol. 163, Sep. 2023, Art. no. 107132, doi: 10.1016/j.compbiomed.2023.107132.

[21] N. Razfar, J. True, R. Bassiouny, V. Venkatesh, and R. Kashef, "Weed detection in soybean crops using custom lightweight deep learning models," *J. Agricult. Food Res.*, vol. 8, Jun. 2022, Art. no. 100308, doi: 10.1016/j.jafr.2022.100308.

[22] Z. Wu, Y. Chen, B. Zhao, X. Kang, and Y. Ding, "Review of weed detection methods based on computer vision," *Sensors*, vol. 21, no. 11, p. 3647, May 2021, doi: 10.3390/s21113647.

[23] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2015, *arXiv:1506.02640*.

[24] F. Dang, D. Chen, Y. Lu, and Z. Li, "YOLOWeeds: A novel benchmark of YOLO object detectors for multi-class weed detection in cotton production systems," *Comput. Electron. Agricult.*, vol. 205, Feb. 2023, Art. no. 107655, doi: 10.1016/j.compag.2023.107655.

[25] *Drone Mapping Software | Drone Mapping App | UAV Mapping | Surveying Software | DroneDeploy*. Accessed: May 5, 2023. [Online]. Available: https://www.dronedeploy.com/

[26] *Professional Photogrammetry and Drone Mapping Software*. Accessed: May 5, 2023. [Online]. Available: https://www.pix4d.com/

[27] (2023). *Unmanned Aerial Vehicle*. Accessed: May 5, 2023. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Unmanned_aerial_vehicle&oldid=1153124328

[28] *Making Orthomosaics With Drones: Everything You Need to Know*. Accessed: May 7, 2023. [Online]. Available: https://enterprise-insights.dji.com/blog/orthomosaics

[29] (2023). *Pixel*. Accessed: Jul. 7, 2023. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Pixel&oldid=1156571776

[30] *Ground Sampling Distance (GSD) in Photogrammetry*. Accessed: Jul. 7, 2023. [Online]. Available: https://support.pix4d.com/hc/en-us/articles/202559809-Ground-sampling-distance-GSD-in-photogrammetry

[31] *What is Pixel Resolution?*. Accessed: Jul. 7, 2023. [Online]. Available: https://www.educative.io/answers/what-is-pixel-resolution

[32] *Image Statistics*. Accessed: Jul. 7, 2023. [Online]. Available: https://web.media.mit.edu/

[33] *4.5—Eigenvalues and Eigenvectors | STAT 505*. Accessed: Jul. 7, 2023. [Online]. Available: https://online.stat.psu.edu/stat505/lesson/4/4.5

[34] *Equations for 3D Haralick Texture Feature - FarsightWiki*. Accessed: Jul. 7, 2023. [Online]. Available: http://farsight-toolkit.ee.uh.edu/wiki/Equations_for_3D_Haralick_Texture_Feature

[35] B. S. V, "Grey level co-occurrence matrices: Generalisation and some new features," *Int. J. Comput. Sci., Eng. Inf. Technol.*, vol. 2, no. 2, pp. 151–157, Apr. 2012, doi: 10.5121/ijcseit.2012.2213.

[36] *Textural Features for Image Classification | IEEE Journals & Magazine | IEEE Xplore*. Accessed: May 12, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/4309314

[37] R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, no. 5, pp. 786–804, May 1979, doi: 10.1109/PROC.1979.11328.

[38] A. K. Maurya, N. Bhargava, and D. Singh, "Efficient selection of SAR features using ML based algorithms for accurate FVC estimation," *Adv. Space Res.*, vol. 70, no. 7, pp. 1795–1809, Oct. 2022, doi: 10.1016/j.asr.2022.06.039.

[39] D. J. I. Enterprise. *Ground Sample Distance | DJI Enterprise*. Accessed: May 4, 2023. [Online]. Available: https://enterprise-insights.dji.com/blog/ground-sample-distance
[40] *Spatial Resolution—An Overview | ScienceDirect Topics*. Accessed: May 6, 2023. [Online]. Available: https://www.sciencedirect.com/topics/engineering/spatial-resolution
[41] *Image Masking*. Accessed: Jul. 7, 2023. [Online]. Available: https://www.xinapse.com/Manual/masking.html
[42] *Region of Interest (ROI) Tool*. Accessed: May 12, 2023. [Online]. Available: https://www.l3harrisgeospatial.com/docs/regionofinteresttool.html
[43] J. R. Schott, S. D. Brown, R. V. Raqueño, H. N. Gross, and G. Robinson, "An advanced synthetic image generation model and its application to multi/hyperspectral algorithm development," *Can. J. Remote Sens.*, vol. 25, no. 2, pp. 99–111, Jun. 1999, doi: 10.1080/07038992.1999.10874709.
[44] S. Hinterstoisser, V. Lepetit, P. Wohlhart, and K. Konolige, "On pre-trained image features and synthetic images for deep learning," presented at the *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, 2023, pp. 1–18.
[45] A. Rozantsev, V. Lepetit, and P. Fua, "On rendering synthetic images for training an object detector," *Comput. Vis. Image Understand.*, vol. 137, pp. 24–37, Aug. 2015, doi: 10.1016/j.cviu.2014.12.006.
[46] *Saturation Stretch*. Accessed: May 12, 2023. [Online]. Available: https://www.l3harrisgeospatial.com/docs/saturationstretch.html
[47] A. R. Gillespie, A. B. Kahle, and R. E. Walker, "Color enhancement of highly correlated images. I. Decorrelation and HSI contrast stretches," *Remote Sens. Environ.*, vol. 20, no. 3, pp. 209–235, Dec. 1986.
[48] F. Nargesian, H. Samulowitz, U. Khurana, E. Khalil, and S. D. Turaga, "Learning feature engineering for classification," pp. 2529–2535, Aug. 2017, doi: 10.24963/ijcai.2017/352.
[49] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[50] *Introduction To Dimensionality Reduction*. Accessed: Jul. 7, 2023. [Online]. Available: https://www.geeksforgeeks.org/dimensionality-reduction/
[51] *Color Models and Color Spaces—Programming Design Systems*. Accessed: Jul. 7, 2023. [Online]. Available: https://programmingdesignsystems.com/color/color-models-and-color-spaces/index.html
[52] S. Sural, G. Qian, and S. Pramanik, "Segmentation and histogram generation using the HSV color space for image retrieval," in *Proc. Int. Conf. Image Process.*, vol. 2, Sep. 2002, pp. 1–12.
[53] R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image analysis using mathematical morphology," *IEEE Trans. Pattern Anal. Mach. Intell.*, vols. PAMI-9, no. 4, pp. 532–550, Jul. 1987, doi: 10.1109/TPAMI.1987.4767941.
[54] B. Sebastian V, A. Unnikrishnan, and K. Balakrishnan, "Gray level co-occurrence matrices: Generalisation and some new features," 2012, *arXiv:1205.4831*.
[55] Y. Wang and W. Zhang, "Generalized nearest neighbor decoding," 2020, *arXiv:2010.06791*.
[56] *Standard RGB Color Spaces*. Accessed: Jun. 15, 2023. [Online]. Available: https://infoscience.epfl.ch/record/34089
[57] G. N. Srinivasan, "Statistical texture analysis," vol. 36, 2008.
[58] *Keras: Deep Learning for Humans*. Accessed: Jul. 1, 2023. [Online]. Available: https://keras.io/
[59] A. Fred Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*.
[60] R. Hecht-Nielsen, "Theory of the backpropagation neural network**based on 'nonindent,'" in *Neural Networks for Perception*. New York, NY, USA: Academic, 1992, pp. 65–93.
[61] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2020, doi: 10.1109/ACCESS.2019.2962617.
[62] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[63] *Support Vector Machines | SpringerLink*. Accessed: Jul. 1, 2023. [Online]. Available: https://link.springer.com/referenceworkentry/10.1007/978-0-387-30162-4_415
[64] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/a:1010933404324.
[65] *Multilayer Perceptron—An Overview | Science Direct Topics*. Accessed: Jul. 1, 2023. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/multilayer-perceptron
[66] L. O. Chua and T. Roska, "The CNN paradigm," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 40, no. 3, pp. 147–156, Mar. 1993, doi: 10.1109/81.222795.

VYOMIKA SINGH (Graduate Student Member, IEEE) received the integrated dual B.Tech. and M.Tech. degrees in computer science and engineering and in software engineering from the School of Information and Communication Technology, Gautam Buddha University, Gautam Budh Nagar, Uttar Pradesh, India, in 2017. She is currently pursuing the Ph.D. degree with the Department of Electronics and Communication Engineering, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India. Her research interests include digital image processing and analysis, image enhancement and fusion, computer vision, pattern recognition, object detection, machine leaning, and drone-based applications.

DHARMENDRA SINGH (Senior Member, IEEE) received the Ph.D. degree in electronics engineering from the Indian Institute of Technology (BHU), Varanasi, India. He has more than 30 years of experience in teaching and research. He has been a Visiting Scientist Postdoctoral Fellow with the Department of Information Engineering, Niigata University, Niigata, Japan; the German Aerospace Center, Cologne, Germany; the Institute for National Research in Informatics and Automobile, France; the Institute of Remote Sensing Applications, Beijing, China; Karlsruhe University, Karlsruhe, Germany; and the Polytechnic University of Catalonia, Barcelona, Spain. He has also dedicated his time and expertise in several other prominent laboratories in countries, such as the USA, Switzerland, South Korea, Canada, and many more. He is currently a Professor with the Department of Electronics and Communication Engineering, Indian Institute of Technology (IIT) Roorkee, Roorkee, India, where he is also a Coordinator with the Drone Research Center. He has published more than 340 papers in various national/international journals and conferences. Recently, he was ranked among the top 2% scientists in the world in the field of electronics and telecommunication, by an independent study done by Stanford University, in 2020. His research interests include digital image processing and analysis, image enhancement and fusion, artificial intelligence, machine learning, the IoT, drone-based application, anti-drone, satellite data-based application for ICT, microwave and optical remote sensing, along with an expertise in electromagnetic modeling, EM wave absorbing materials, microwave/millimetre/THz imaging, and radar polarimetry and interferometry.

HARISH KUMAR received the Bachelor of Engineering degree in computer science and engineering from Visvesvaraya Technological University, India, and the Master of Technology degree in computer cognition technology from Mysore University, India, and the Ph.D. degree in computer science and engineering from the Indian Institute of Technology Roorkee. He did his postdoctoral research with the INRIA Bordeaux Sud-Ouest, France. He was the Chief Engineer with SEL, Samsung India Electronics Private Ltd., India. He is currently the University Professor with the Department of Computer Science, College of Computer Science, King Khalid University, Abha. He did his internship with the Samsung Global Internship Program 2008, Samsung Electronics Company Ltd., Suwon, South Korea (he was one among two students from India, selected for the year 2008). He has several research papers in well-reputed journals. His research interests include augmented reality, virtual reality, security and privacy, HCI, steganography, machine learning, pattern recognition, natural language processing, computer vision, insider threats, fuzzy logic, deep learning, and soft computing. He has received the Young Scientist Award 2010 from the Second National Young Scientist Symposium 2010, Uttarakhand, India.