2022 International Conference on Futuristic Technologies (INCOFT)
Karnataka, India. Nov 25-27, 2022

2022 International Conference on Futuristic Technologies (INCOFT) | 978-1-6654-5046-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/INCOFT55651.2022.10094447

Evaluation and Segregation of Fruit Quality using Machine and Deep Learning Techniques

Mohit Kedar Mali
Dept. of ESE, National Institute of Electronics and Information Technology.
Aurangabad, Maharashtra, India.
noni4577@gmail.com

Shubham Rajendra Devake
Dept. of ESE, National Institute of Electronics and Information Technology.
Aurangabad, Maharashtra, India.
shubhamdevake1999@gmail.com

Satyam Manoj Kharpude
Dept. of ESE, National Institute of Electronics and Information Technology.
Aurangabad, Maharashtra, India.
satyamkharpude09@gmail.com

Yogesh Kumar
Scientist B
Dept. of ESE, National Institute of Electronics and Information Technology
Aurangabad, Maharashtra, India.
yogeshkumar@nielit.gov.in

Prashant Pal
Scientist-B
Dept. of ESE, National Institute of Electronics and Information Technology
Aurangabad, Maharashtra, India
prashantpal@nielit.gov.in

Shashank Kumar Singh
Scientist-B
Dept. of ESE, National Institute of Electronics and Information Technology
Aurangabad, Maharashtra, India.
shashank@nielit.gov.in

Saurabh Bansod
Scientis-C
Dept.of ESE, National Institute of Electronics and Information Technology
Aurangabad, Maharashtra, India
saurabhbansod@nielit.gov.in

***Abstract*: Identifying fruit fruits is an essential part of fruit plantation smart management. This paper presents a mechanism based on the available deep learning model to determine the fruit fast and reliably in a complicated orchard environment. We employed the YOLOv3 method to detect the deep characteristics of fruit fruits using a stereo camera and an indoor fruit dataset, resulting in efficient identification of varied fruit sizes. In this paper, we proposed segmentation of fruit using the UNET framework using various deep learning backbones such as densenet, efficientnet, mobilenet, vggnet etc. This research has been done on field fruit images, so it contains some noise in the background of the image; this problem impacts accurate detection and classification. The YOLOv3 model has been used for fruit object detection with boundary regions. It returns the normalized image in square form with the detection of edges of the object the UNET framework has used for classification. The various backbones obtain different accuracy, but UNET-VGG19 brings Dice Coefficient of 90.35%, which is better than other methods. As a result, various experimental analyses were done on real-time images and demonstrated the dice score, precision, recall, etc.**

Keywords: Yolov3 model, VGG16, U-Net, Fruit Detection, Segmentation, image processing, deep learning, fruit detection

I. INTRODUCTION

With the paucity of agricultural work and the speedy development of artificial intelligence (AI) , robots have received much interest in the agricultural area. Agricultural robots perform time-consuming farm tasks, allowing farmers to concentrate more on farm management. Amongst the most famous agricultural robots is the harvesting robot. Harvesting robots have grown greatly in speed and efficiency in recent times, and people are becoming more enthusiastic in using agricultural robots to extract fruits and veggies. Several investigations on fruit identification have been conducted recently. The best way to achieve automated harvesting is a vision technology, and correct detection is the foundation of follow-up operations including picking in fruit or vegetable harvesting. Due to the resemblance or obstruction of fruits, branches, and also other environmental difficulties and the unpredictability of the orchard surroundings, developing an accurate and robust fruit identification method is a major task.

Crop development data must be collected smoothly and quickly in order for intelligent agriculture to work. Fruit identification and accurate location are required of robot harvesting systems in the orchard; fruit identification can also instantly check the number of fruits in the area. The orchard management can use computerized fruit scanning to monitor fruit drop, predict yield, and prepare for the market accordingly. Fruit detection method related to machine vision can already identify fruit development data, provide numerous indications for illness and insect outbreaks, forecast yield, place the harvesting, and do other activities. Robotics is increasingly being used in orchards, especially for yield estimation, yield modeling, as well as automated picking. Simultaneously, machine vision serves as the lenses of a smart robot, allowing it to detect its surroundings, develop its knowledge, and so enhance its work precision and effectiveness.

Fruit is perhaps the most common fruit on the planet and an essential source of sustenance. Because fruits have an inconsistent color and texture, reliable detection is becoming the major challenge of a fruit harvesting robot in the wild. Environmental factors have a big impact on fruit setting percentage; the quantity of fruit fruits fluctuates depending on climate and environment. Physically collecting the color, structure, shape, as well as other properties of the fruits was one of the first techniques to systematic fruit detection. This article describes a deep learning-based method for detecting fruits in a field. We obtained fruit photos at the farm using a conventional RGB color camera, and fruit fruits were recognized under various illumination and occlusion circumstances.

The primary research content includes-

- The outcomes in this article were compared with the fruit detection based on U-Net algorithm with MobileNetv2 backbone, to authenticate the applicability as well as high effectiveness of the suggested technique in fruit detection.
- Under diverse environmental conditions, the proposed detection method YOLOv3 allows for fast and precise detection of fruit fruits.

The remainder of this article is organized in the following manner. The second section goes through the work that has been done in the past. The final section explains how the fruit detection method works and how it is implemented in the plantation. The test data and comparative analysis are introduced in parts four and five. The overview and action strategy for the future are included in the sixth section.

II. LITERATURE SURVEY

Zhing Qu et al. [1] develop a concrete pavement fracture detection technique using Convolutional Neural Network (CNN) in 2020. The improved VGG16 framework is developed to remove the concrete crack features in the crack identification section. Two crack databases were created for evaluation process; with CCD1500 serving as the crack categorization dataset and CDD861 serves as the crack recognition database. On the CFD corporus, Cracktree200 corpora, and DeepCrack database, the suggested approach has the maximum F1 value in comparison to VGG16, U-Net, as well as Percolation techniques. The created approach maintains VGG16's significant PFS value while improving the system's crack detecting effectiveness. It has a comparatively decent accuracy and recall value for all types of cracks in various situations, allowing for speedy and precise crack picture identification. Hua Bai et al. [2] suggested a deep learning-based chromosomal extraction approach in 2020. The raw micrographs were segmented using U-Net to eliminate background noise like nuclei as well as other disruptions. Then, using YOLOv3, every chromosome was detected and extracted. U-Net was also utilized to accurately retrieve the individual chromosomes. The test findings reveal that this technology can efficiently eliminate interferences from elsewhere in the chromosomes while precisely extracting overlapping and sticky chromosomes. The extraction of chromosomes from unprocessed G-band chromosomal pictures is 99.3 percent accurate. This technique is critical for the advancement of computerized karyotype advanced analytics.

Waqar Riaz et al. [3] used the AOLP corpora for license plate identification in 2020. The suggested technique used YOLOv3 for identification and CRNN for categorization, following multitask learning approach for text string identification. For the purposes of assessment, 40 percent of the photos were assigned to the training dataset, 20 percent to the validation dataset, and 40 percent to the testing dataset. For the test set examination, a lesser threshold of 0.125 was adopted, resulting in a 99.82 recall rate. The suggested methodology has 86 percent recognition accuracy, with 88 percent of 3-letter plates and 99 percent of 4-letter plates being recognized. Finally, utilizing temporal redundancy, the ultimate identification accuracy is greatly enhanced, reaching 96%. The suggested scheme boosts detection accuracy between 93.58 to 96.1 percent, exceeding Sighthound & OpenALPR by 9percent and 4.9 percent, correspondingly. Pranjal Ranjan et al. [4] suggest a hybrid design for categorization of remote sensing - based pictures collected from U-Net in 2020, enhanced by Multi-Resolution Analytical processing (MRA). While the same data on various scales is accessible for extracting the features and training, the channel's contextual perspective improves. The impacts of several wavelets for deconstruction on recognition rate are also investigated in this study. When compared to the ordinary U-Net system, the MRA-aided system can divide regions with greater precision and average overlap over union, with Daub basis of MRA providing the highest efficiency. Gliomas account for 23.41 percent of every basic brain tumors in the United States, with over 2000 occurrences detected annually. Grade II gliomas, also known as Low-Grade Gliomas (LGG), are sluggish brain tumors classified by the World Health Organization (WHO). LGG is an illness that kills teenagers (among 35 to 44 years of age). LGG can progress to High-Grade Gliomas (HGG) or WHO classifications III and IV in the majority of individuals, resulting in mortality. Surgical removal, radiation, and chemotherapeutic are the most common treatments for LGG sufferers. While performing surgical excision, fluid-attenuated inverted recovery scanning is required to detect the tumor site. In 2021, Dwilaksana Abdullah Rasyid et al. [5] proposed a dynamic system for tumor segmentation that combines U-Net as well as VGG16 design innovations with transfer learning. The deep learning method obtains a good outcome with the Dice Similarity Coefficient of 99 percent and the Area Under Curve (AUC) of 98 percent using pretreatment FLAIR visual information from 110 individuals with LGG out from Cancer Genome Atlas.

In 2019, Wei Cui et al. [6] used a sanitation van to record an actual photograph of litter on the street. For trash identification, the YOLOv3-darknet method was established, which would be dependent on adaptive clustering anchor boxes. Items that are commonly not identified are used in multi-category dialog training. The geographical coordinate data of the rubbish is calculated using additional supplementary range devices, allowing the trash to be cleared with a focus. Tests demonstrate that the developed method is effective at detecting rubbish. On the GTX1080ti, a single image takes 60ms to identify. A novel neural network for autonomous Head and Neck Cancer (HNC) classification using Magnetic Resonance Imaging (MRI) was published by Baixiang Zhao et al. [7] in 2019. The suggested neural network is built on the U-net, which incorporates characteristics from multiple configurations to enable end-to-end clinical picture finding and fragmentation. The expanded convolution is incorporated into U-net in this paper, in order to obtain a bigger dynamic range and retrieve multi-scale characteristics. Dice loss is also used in this system to smooth out the categories. On actual MRI data, the developed methodology is designed and evaluated. The novel network surpassed the initial U-Net by 5percent (Dice score) upon head or neck tumor fragmentation, according to cross-validation data. Chia-Wei Chuang et al. [8] investigated the effectiveness of YOLOv3 small dependent deep-learning inference networks for ocular biometric verification in 2020. After labeling the image data of eye with partial iris as well as sclera regions, the suggested YOLOv3 tiny based classification algorithm succesfully infers the individual's appearance. The implemented YOLOv3 tiny based inference mode provides a mean Average Precision (mAP) of up to 99.92 percent just by using single anchoring unit, according to the UBIRIS dataset. The suggested low complexity approach outperforms earlier ocular biometric research employing iris or sclera data in terms of effectiveness and eliminates the necessity for iris and sclera division.

In 2020, Biao Zheng et al. [9] System assisted rectal medical diagnosis is of tremendous importance for the earlier diagnosis of rectal cancer. According to the indistinct border among lesions as well as good colonic mucosa, segmentation and classification is tough to accomplish in information processing. To solve these challenges, a rectal tumor classification technique based on an altered U-net is developed, which improves detection precision and effectiveness. To begin, the central dimensions of the rectal section must be calculated in order to obtain the area of interest. The YOLOv3 method is then used to identify the tumor site in the CT picture. Furthermore, the remaining link and focus mechanisms are employed to enhance the precision of the classic U-net model by reducing the probability of misjudging normal rectal tissues as tumors, and the revised U-net system has been used to partition the colon tumor area. Tests show that this approach's Dice coefficient could achieve 83.45 percent, which in itself is roughly 7 percent better than the classic Unet approach, demonstrating the validity and usefulness of the suggested method. Deepanshu Waiker et al. [10] used the U-Net and Seg-Net models for lung X-Rays in 2020 and found that U-Net outperformed Seg-Net on the very same corpora. When various loss factors were used on the very same system, it was discovered that the customized loss function, which would be a convolution of BCE, Dice, as well as Inverse Dice, produced some of the best outcome. On either a testing data, the U-Net was evaluated with custom losses and received a mean IoU of 0.95. On SegNet, a median IoU of 0.9322 was reached by doing various work. For almost the same information, U-Net significantly increased verification and testing outcomes than SegNet. Dice or BCE loss did not extend the system as well as Custom loss did. As a consequence, the testing results yielded improved outcomes. Among the models tested, the U-Net using custom loss operation performed much better. Although U-Net outperformed SegNet in terms of performance, along with its fully associated tiers, U-Net takes substantially longer to learn than SegNet. U-Net has a lot many variables than SegNet. For X-Ray fragmentation, using SegNet as well as U-Net is a tradeoff between processing speed and precision. SegNet can only be used if the program's compute capability is limited and the outcomes are satisfactory.

Gan Fang et al. [11] published a paper in 2019. White blood cells (WBC) play a critical role in clinical research and diagnostics. Artificial intelligence is now commonly used in clinical area, thanks to its rapid advancement. Deep learning is the primary research technique now in use. Regarding WBC's, many neural networks fail miserably. YOLOv3 is optimized to enhance the prediction performance in this research. Darknet retrieves characteristics in yolov3, whereas k-mean generates boxes. Yolov3 is a multi-scale forecasting algorithm that is superior for detecting small targets. Simultaneously, the data is normalized in this article by dividing it into chunks as per a specific ratio in enhancing the prediction accuracy. As seen in the outcomes, the modified data set outperforms in training than that of the primary corpora. Nan Xiang et al. [12] The issue of poor identification rates and increased failed detection accuracy in deep learning models for identifying traffic signals, and also the shortage of street light databases in China, are the focus of this research. The updated YOLOv3 technique is used to present actual traffic signal identification and tracking approach. To begin, the linear scale amplification approach is utilized to maximize the screen resolution of the previous box formed by K-means clustering, as well as the clustering outcome is proportionally computed to determine the appropriate anchor box dimension. The signal information is then enhanced using the enhanced Mosaic technique. Furthermore, an SPP dimension is connected after the network structure to minimize the recurrent extraction of features of the picture by the CNN, and a 4 step up-sampling tier is introduced to effectively combine high-level contextual features with deep location data. At the very same moment, the prediction model is reduced and the quantity of convolutional layers as in collar region is lowered when contrasted to the YOLOv3 method, Test findings demonstrate that the suggested technique obtains better precision mostly on the Lara database and the Chongqing trTaffic Light Dataset (CQTLD) . On CQTLD, the identification speed is up 11.8 percent, and the mAP is up 3.78 percent.

Htet Aung et al. [13] Facial recognition is not just one of the best researched problems in the world of machine vision, but is also a crucial challenge in a spectrum of uses, including safety access controls, video monitoring, human-computer interfaces, as well as picture database administration. Face monitoring systems also use a variety of approaches, such as Viola-Jones, Recurrent Convolutional Neural Network (RCNN), SSD, and others. Several academics are still experimenting with different pictures, stances, skin hues, and actual identification to enhance facial recognition algorithms. This research proposes an enhancement for facial detection techniques by combining the You Only Look Once (YOLO) method with the VGG16 pre-trained CNN. The suggested technique identified the testing image dataset with an average accuracy of above 95 percent, according to the findings. In addition, the suggested technique significantly boosted the efficiency of facial algorithm to detect live stream. The Image Analysis

Package as well as the Deep Learning Toolkit in MATLAB was used in this investigation. Yi-Lin Tu et al. [14] Graphic based plant phenotyping has been a popular study issue in the 21st century. It can aid research into the complicated relationship among genes and behavioral features. Traditional herbal phenotyping, on the other hand, requires physical assessment of plant features and is time-consuming. Duration and personal labor can be greatly diminished in this procedure with the aid of advanced picture processing. An computerized approach for leaf identification was established in the pursuit of the sole goal of automated phenotyping. There are 2 kinds of image-based plant identification methods currently available. First is focused on fragmentation of occurrences, while another is focused on regression forecasting. Object recognition algorithms have been proposed as an alternative to current methods. The YOLOv3 technique, especially, has been tweaked to best suit the leaf counting requirement Cauliflower photos from the Academia Sinica – Biotechnology Center in Southern Taiwan (ASBCST) dataset and Arabidopsis pictures from the Computer Vision Problems in Plant Phenotyping (CVPPP) dataset were used to test the suggested technique. On these databases, the suggested technique produces impressive outcomes.

Masahiro Takahashi et al. [15] hope to build a gentle item detector in 2020 that takes a dimension and hue picture from a stereo camera as feed. It is feasible to produce in the depth direction through expanding the network design of YOLOv3 to three dimensions in the center. In particular, in three - dimensional space, Intersection over Union (IoU) is used to check the correctness of area to determine outcomes. Item detectors that employ distance data as feed are being intensively explored in the area of deep learning for its use in self-driving cars. The typical sensor, on the other hand, has a huge framework, which reduces its actual capability. Datasets are used to verify the efficiency of the sensor built as mentioned previously. The study demonstrated that the suggested approach can generate 3D bounding boxes and identify humans with partially covered bodies. The system's computational power was also 44.35 frames per second. Dongdong Zhang et al. [16] published a paper in 2019. Clinical image processing study involves a component called cell identification. Counting red and white cells in tiny photographs is among the most major tasks. The dispersion of WBC is thin and simple to detect in a picture, but the dispersion of Red Blood Cells (RBC) is thick and susceptible to overlap stickiness. In this study, a novel technique for identifying RBC as well as WBC in microscopic images is developed. To begin, in a microscopic picture, the YOLOv3 net is utilized to recognize distinct RBC, WBC, and aggregation RBC's. Subsequently, using an image intensity estimate technique, researchers quantify the collected RBC. This technique can deliver better precise count findings than the YOLOv3 and picture intensity estimate methods. Naomi Yagi et al. [17] suggested utilizing U-Net to split the belly region for tumor radiation assistance in 2019. Two approaches were tested for classification performance: the very first segmented the urethra, prostate, and colon separately, whereas the other segmented their aggregate area. As a consequence, the very first technique outperformed the others, with a mean Dice coefficient of 0.96.

Even though nuclei offer quantitative data regarding tumor malignancy, nuclei identification is a challenging stage in cytopathology for pleural detection of cancer. CNN are well-known for their superior quality in image classification. Busranur Kilic et al. [18] presented YOLOv3 as a convolutional item detector for detecting nuclei in pleural effusion (PE) cytopathology pictures in 2019. The tests were carried out on 80 PE cytopathology pictures with 11157 nuclei. The suggested model had an accuracy of 94.10 percent, a recall of 98.98 percent, and an Fmeasure of 96.48 percent. The YOLOv3's greatest significant addition is a 10x speed increase over certain established state-of-the-art approaches, which would be critical for actual Computer-Aided Diagnosis (CAD) uses in electronic cytopathology. Xueyan Gao et al. [19] developed a novel approach of semantic fragmentation in 2020, with the goal of improving the program's segmentation accuracy. A residual network is appended to the regular U-NET net, deepening the network's thickness while preventing the performance reduction related to increased height. Simultaneously, the Attention Mechanism (AM) is implemented, allowing the structure to flexibly concentrate on multi-scale characteristics, improving multi-scale contextual usage. The suggested technique's testing was conducted on the BRATR 2017 corpora, and it produced good fragmentation results. The collection of clinical images is challenging in 2019, according to Wang Yijie et al. [20], and the tiny volumes of information is a big issue for image classification. U-peculiarity net's allows it to lead to better outcomes even with few data. U-net is utilized to divide arteries in a fundus picture in ability to forecast certain eye disorders upfront in this article. The planned U-net has been transformed from a traditional to a 7-layer net, and different factors, like patch size, have been improved. The test findings reveal that the fundus arteries produced by this fragmentation are quite near to the markers and have superior accuracy than previous approaches. The technique is quite useful for finding a solution of limited clinical picture data classification. Alavikunhu Panthakkan et al. [21] u sing modern machine learning approaches and lung X-rays, latest study attempts to predict COVID-19 (+) with high accuracy. The potential VGG16 transfer learning approach for the efficient and timely detection of COVID-19 (+) is provided in this research. The technique divides the lung X-ray picture into two categories: COVID-19 (+) and Healthy. Performance indicators like correctness, specificity, recall, and f1 score are used to assess the success of the developed system. 2000 X-ray samples were used in the investigations. The suggested VGG16 approach has an amazing detection performance of 99.5 percent for the two-class categorization of the given sample size, which is higher than any other current approaches in the literature. Because the proposed method is both precise and exact, it can be utilized to assist and assist radiologists and health workers in identifying COVID19 (+) using lung X-rays.

Xiaoxiong Zheng et al. [22] published a paper in 2020. With the advancing technology of deep learning in latest years, the area of remote sensing image processing has begun to employ few deep learning methods to accomplish smart and fast image processing, with outcomes that have increased to a certain level when opposed to prior techniques. In 2015, the U-Net CNN for clinical picture categorization was developed. The U-Net was applied to remote sensing image fragmentation to achieve pixel level semantic fragmentation of remote sensing data end-to-end, depending on prior studies. The total performance of train corpora is 93.83 percent, whereas the precision of testing dataset is 82.27 percent, the kappa coefficient is 0.7721, as well as the Mean intersection Over Union (MiOU) is 0.6405. These results were obtained using U-Net training as well as learning on GF-2 remotely sensed picture. The test's findings indicate that it had a high level of segmentation performance and generalization capacity. Wei Zhang et al. [23] published a paper in 2019. In the area of earth observation, building retrieval from remotely sensed photos is often a difficult task. Furthermore, fully convolutional networks-based techniques have reached state-of-the-art efficiency in semantic classification and enabled dense pixel-wise categorization possible, with the U-Net design is among the most prominent models. For generating extraction, 3 U-Nets comprised of regular convolution blocks, remnant units, and conception units, dubbed Normal-U-Net, Residual-U-Net, as well as Inception-U-Net, are proposed. The Inception U-Net is perhaps the most successful design overall, according on test findings on the open Massachusetts building corpora. Furthermore, the three individual U-Nets are integrated into a tougher method known as EU-Net, have shown outstanding results in building extraction.

The potential of methodologies for automated brain tumor identification has improved significantly, according to Neil Micallef et al. [24], with the introduction of deep learning techniques for picture identification. This study proposes a framework for this job that is based on U-Net++ and improves training efficiency while simultaneously improving the reliability. On the total tumor, tumor base, and improved tumor core categories, the technique received Dice Scores of 0.90, 0.85, and 0.68, respectively. This study relied on a 68-scan outlier set from the BraTS 2019 training sample. In addition, the suggested system uses 50% of the variables of a famous U-Net adaption that are using residual blocks, leading to a rapid training. Including all 3 categories in our configuration, the method performs 8.44 percent higher than the latter in terms of Dice scores. Jierong Cheng et al. [25] introduced a deep Convolutional Neural Net U-Net-based automated crack detecting technique in 2018. Unlike many other machine learning depending crack identification approaches, U-encoder-decoder Net's structure allows an image to be analyzed without patching. Rather than gathering from neighboring chunks, the categorization outcome is produced from the net. Furthermore, a new cost function depending upon the distance transform is provided to apply pixel-level value based on the segmentation's least proximity. The evaluation is performed on 2 corpora of road fracture photos in tests. The pixel-level fragmentation preciseness is greater than 92 percent, greatly outperforming other state-of-the-art approaches.

III. PROPOSED SYSTEM ARCHITECTURE

Stacking numerous standard convolutional layers leads the network to suffer from gradient vanishing issue, which hinders the optimization of network weights, yet the classical U-net has achieved great accuracy in segmentation. Four residual blocks replace the encoder's initial basic convolutional unit to make deep network training easier. Due to down sampling in the encoder, attention gate (AE) and squeeze-and-excitation (SE) blocks have been added into the decoder in order to improve the richer low level and high-level information and offer a larger weight coefficient for more important channels. It is shown in Figure 1 how the proposed Unet network, as well as the acquisition equipment for fluorescence pictures, data processing, and training and testing of the proposed networks, are used to detect oil leaks from a transformer, as explained below.

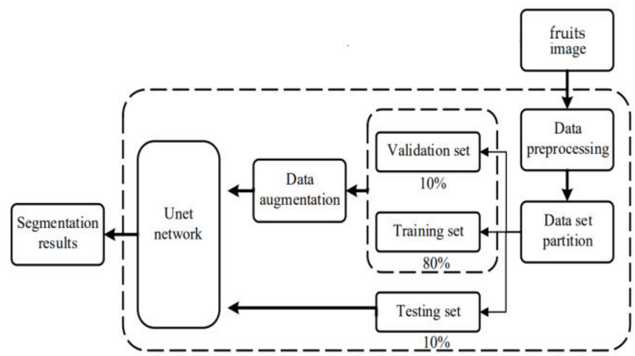

Fig. 1. Proposed system architecture for fruit segmentation

A. Image acquisition

This module collects plant image datasets from various sources, such as on filed fruit image datasets. The dataset may be imbalanced sometime, and it contains noisy images. In the second section, we pre-process and normalized the entire dataset to achieve the best results for training and testing.

B. Data Pre-processing

Pre-processing reduces distortion, making post-processing simpler. Pre-processing includes colour space transformation, cropping, smoothness, and enhancement. This module's use varies with image quality. Color space converting is accompanied by filtering and augmentation. The YOLOv3 module has been used for fruit object detection and to remove the noise from the image. YOLOv3 returns the output image with edges and boundaries of fruit existing in the entire image. The major purpose of the utilization of that module the eliminate the background effect of the image as well as accurate detection of fruit in the input image.

C. Feature Extraction

The Color, texture, and form features are often used to understand images. Moments and histograms are widely used to determine colour. Texture may have properties including contrast, homogeneity, variance, and entropy. Similarly, features such as roundness, area, eccentricity, and concavity are discovered for form. Heterogeneous datasets need a variety of characteristics; however, texture has been recognized as the best feature for plant disease diagnosis. For feature extraction, a variety of approaches are utilized.

D. Segmentation

Along with the items of interest, segmentation separates the image into sections with strong association. The number of histogram peaks, for example, is one feature of a correctly segmented picture that aids in the simple identification of healthy or contaminated samples. Plant disease detection systems have been demonstrated to function effectively using edge, threshold, location, and colour-based segmentation approaches. As a result of the considerable colour disparities between the infected leaf region and its native colour, spot colour-based segmentation emerges. In segmentation, determining a threshold value is critical.

E. Algorithm Design

To implement this work, we design a new modified deep learning-based convolutional neural network classifier called CNN. This algorithm is divided into two phases such as training and resting. The training module generates the rules for the entire module, while the testing phase validates disease detection and classification tests.

F. Execution of Training

Input: Train_DB[] as training dataset, set of activation function AF[].

Output: Trained module in .PKL file for entire splited dataset

Step 1: Initialize the both algorithms Train_DB[], AF[], epoch_size

Step 2 : Extracted_Features ← ExtractFeatures(Train_DB[])

Step 3 : Selecetd_Features[] ← optimized(Extracted_Features)

Step 4 : Train.pkl ← Build_Classifier(Selecetd_Features[])

Step 5: Return Train.pkl

G. Execution of Testing

Input: Test_DB [] as testing instance set or individual patient record, Training Background Knowledge Train.pkl, User defines threshold Th

Output: Output_Map <Predicted_class_label, Similarity_weight> optimized instance recommend by classifier.

Step 1: Read all testing records by using below equation

$$test\_Feature(m) = \sum_{m=1}^{n} (.\ feature\_Set[A[i] \ldots\ldots A[n] \leftarrow \text{Test\_DB})$$

Step 2 : Extract selected attribute features from entire test record $testFeature(m)$ using below equation.

$$\text{Extracted\_Feature\_Set\_x}[t\ldots\ldots n] = \sum_{x=1}^{n}(t) \leftarrow test\_Feature\ (m)$$

Extracted_Feature_Set_x[t] contains the feature vector of respective domain

Step 3: Extract all training instance from trained modules using below function

$$train\_Feature(m) = \sum_{m=1}^{n} (.\ feature\_Set[A[i] \ldots\ldots A[n] \leftarrow \text{Train.pkl})$$

Step 4 : extract each feature as a hot vector or input neuron from $testFeature(m)$ using below equation.

$$\text{Extracted\_Feature\_Set\_y}[t\ldots\ldots n] = \sum_{x=1}^{n}(t) \leftarrow test\_Feature\ (m)$$

Extracted_Feature_Set_x[t] contains feature vector for entire class labels.

Step 5 : Now evaluate each testing instance with all train features

$$calc\_weight = calcSim\ (\text{Feature\_Set\_x} \,||\, \sum_{i=1}^{n} \text{Feature\_Set\_y[y]})$$

Step 6 : *Return calc_wSeight*

IV. Results and Discussions

A. Dataset Description

The dataset validation has done on filed fruit images, the 1500 on-field fruit plant images and 1 hr video captured at Warkute village, Baramati with the time of 09/01/2022, 4 pm to 6 pm. and 10/01/2022, 9 am to 12 pm.

The systems were deployed on the widow's platform with Python 3.7 and the RESNET-100 deep learning framework for an extended experimental examination. The hardware configuration has built a Inference (NVIDIA tesla v100 GPU) (including pre and post processing). In this experiment, we use a real-time plant disease dataset to illustrate the classification accuracy of CNN (Sigmoid). Figure 2 displays the results of similar trials utilizing various cross validation techniques. According to the findings, 10-fold cross validation has the highest average classification accuracy of 92.10%. With CNN and sigmoid function, the 5-fold cross validation likewise obtains 89.60%. Figure 2 shows the results of a 10-fold data cross validation. During module testing, both functions obtain a comparable level of accuracy.

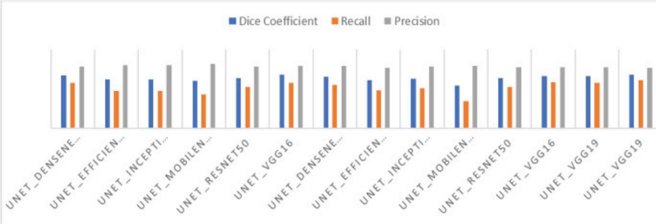

Fig. 2. system validation with various cross validation using proposed framework

B. *Comparative analysis of existing Deep Learning algorithms*

In another investigation, the probability of disease detection using supervised deep learning classification. System describes four evaluations between this research results and some existing systems results has calculated on the similar as well as multiple datasets.

TABLE I. COMPARATIVE ANALYSIS OF PROPOSE MODEL WITH OTHER BACKBONE

Backbone	input	Dice Coefficient	Recall	Precision
unet_densenet121	RGB	90%	87.14%	93.55%
unet_efficientnetb0	RGB	88.45%	84.03%	93.96%
unet_inception_resnetv2	RGB	88.46%	84.06%	94.07%
unet_mobilenetv2	RGB	87.97%	82.87%	94.50%
unet_resnet50	RGB	89.11%	85.60%	93.47%
unet_vgg16	RGB	90.26%	87.35%	93.80%
unet_vgg19	**RGB**	**90.35%**	**88.23%**	**93.04%**
unet_densenet121	HSV	89.70%	86.55%	93.70%
unet_efficientnetb0	HSV	88.40%	84.53%	93.07%
unet_inception_resnetv2	HSV	88.92%	85.30%	93.36%
unet_mobilenetv2	HSV	86.26%	80.21%	93.65%
unet_resnet50	HSV	88.98%	85.70%	93.20%
unet_vgg16	HSV	89.87%	87.43%	93.09%
unet_vgg19	HSV	89.90%	87.26%	93.29%

The system proposed CNN based hybrid model for segmentation of fruit using VGGNET-16 with YOLOV3 model. The precision achieved here is 93.04 % which is better than other Deep Learning models. Figure 5 compares the proposed algorithms' classification accuracy to that of different known machine learning techniques. For data organization or classification, the most recent predicted sample employs a training set and a test set. The training package is made up of input function modules and their associated class labels. This learning set is used to create a classification model that organizes the input data into appropriate template files or labels. The model is then validated using a test set derived from the class labels in entire test dataset.

V. CONCLUSION

The precise detection of fruits is critical for the intelligent administration of fruit detection and classification. In this paper, we propose a identification approach for fruit detection in the natural surroundings depending on the proposed YOLOv3 model for preprocessing and UNET with various backbones. We also looked at how well the standard Machine Learning (ML) technique, Neural Network (NN) approaches, U-Net, performed in fruit detection. The following conclusions were drawn based on empirical findings:

- In the plantation, we discovered a viable deep learning system for fruit detection. The proposed approach can extract deeper fruit characteristics while reducing background and uneven fruit influence.
- The YOLOv3 reduces the background noise of image by using accurate detection of boundary regions of input image.
- The different deep learning frameworks have been utilized for classification after preprocessing those images using YOLOv3, and the UNET-vgg19 brings a higher dice score, precision, and recall.
- For fruit identification in the orchard, the Deep Learning (DL) model outperforms than the Machine Learning (ML) approach. To detect the fruit fruit, the YOLOv3 was utilized, which removed the backdrop of the discovered fruit. The YOLOv3 output is sent into the U-Net with VGG19 backbone for segmentation. The suggested model reduces false segmentation and achieves 93% accuracy.

REFERENCES

[1] Zhong Qu, Jing Mei, Ling Liu and Dong-Yang Zhou, "Crack Detection of Concrete Pavement With Cross-Entropy Loss Function and Improved VGG16 Network Model", 2020, IEEE

[2] Hua Bai, Tianhang Zhang, Changhao Lui, Wei chen, Fangyun Xu and Zhi-Bo, "Chromosome Extraction Based on U-Net and YOLOv3", 2020, IEEE

[3] Waqar Riaz, Abdullah Azeem, Gao Chenqiang, Zhou Yuxi, Saifullah, Waqas Khalid, "YOLO Based Recognition Method for Automatic License Plate Recognition", 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), IEEE

[4] Pranjal Ranjan, Sarvesh Patil, Rizwan Ahmed Ansari, "U-net based MRA framework for segmentation of remotely sensed images", 2020 International Conference on Artificial Intelligence and Signal Processing (AISP), IEEE

[5] Dwilaksana Abdullah Rasyid, Guan Hua Huang and Nur Iriawan, "Segmentation of Low-Grade Gliomas using U-Net VGG16 with Transfer Learning", 2021, IEEE

[6] Wei Cui, Wei Zhang, Juli Green, Xu Zhang and Xiang Yao, "YOLOv3-darknet with Adaptive Clustering Anchor Box for Garbage Detection in Intelligent Sanitation", 2020, IEEE

[7] Baixiang Zhao, John Soraghan, Gaetano Di Caterina and Derek Grose, "Segmentation of Head and Neck Tumours Using Modified

U-net", 2019, 27th European Signal Processing Conference (EUSIPCO), IEEE

[8] Chia-Wei Chuang, Chih-Peng Fan, and Robert Chen-Hao Chang, "Design of Low-Complexity YOLOv3-Based Deep-Learning Networks with Joint Iris and Sclera Messages for Biometric Recognition Application", 2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)

[9] Biao Zheng, Chenxiao Cai and Lei Ma, "CT images segmentation method of rectal tumor based on modified U-net", 2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV), IEEE

[10] Deepanshu Waiker, Pratik Das Baghel, K Ramgopal Varma and Satya Prakash Sahu, "Effective Semantic Segmentation of Lung X-Ray Images using U-Net Architecture", Proceedings of the Fourth International Conference on Computing Methodologies and Communication (ICCMC 2020), IEEE

[11] Gan Fang, Yang Suhua and Jiang shaofeng, "Detection of white blood cells using YOLOV3 network",2019 14th IEEE International Conference on Electronic Measurement & Instruments

[12] Nan Xiang, Zhao Cao, Yuedong Wang and Qianqian Jia, "A Real-Time Vehicle Traffic Light Detection Algorithm Based on Modified YOLOv3", 2021 IEEE 4th International Conference on Electronics Technology

[13] Htet Aung, Alexander V. Bobkov and Nyan Lin Tun, "Face Detection in Real Time Live Video Using Yolo Algorithm Based on Vgg16 Convolutional Neural Network", 2021 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)

[14] Yi-Lin Tu, Wei-Yang Lin and Yao-Cheng Lin, "Automatic Leaf Counting Using Improved YOLOv3", 2020 International Symposium on Computer, Consumer and Control (IS3C), IEEE

[15] Masahiro Takahashi, Yonghoon Ji, Kazunori Umeda and Alessandro Moro, "Expandable YOLO: 3D Object Detection from RGB-D Images", 2020, IEEE

[16] Dongdong Zhang, Pengfei Zhang and Lisheng Wang, "Cell Counting Algorithm Based on YOLOv3 and Image Density Estimation", 2019 IEEE 4th International Conference on Signal and Image Processing

[17] Naomi Yagi, Manabu Nii and Syoji Kobashi, "Abdominal Organ Area Segmentation using U-Net for Cancer Radiotherapy Support", 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)

[18] Busranur Kilic, Elif Baykal, Murat Ekinci, Hulya Dogan, Mustafa Emre Ercin and Safak Ersoz, "Automated Nuclei Detection on Pleural Effusion Cytopathology Images using YOLOv3", 2019, IEEE

[19] Xueyan Gao and Lijin Fang, "Improved U-NET Semantic Segmentation Network", 2020, Proceedings of the 39th Chinese Control Conference, IEEE

[20] Wang Yijie, LI Shixuan, Cao Guogang , Cao Cong, Li Mengxue and Zhu Xinyu, "Improved U-net fundus image segmentation method", ICIIBMS 2019, Track 1: Image Processing, Computer Science and Information Engineering, Shanghai, China, IEEE

[21] Alavikunhu Panthakkan, Anzar S. M, Saeed Al Mansoori and Hussain Al Ahmad, "Accurate Prediction of COVID-19 (+) using AI Deep VGG16 Model", 2020, 3rd International Conference on signal Processing and Information Security (ICSPIS)

[22] Xiaoxiong Zheng and Tao Chen, "Segmentation of High spatial Resolution Remote Sensing Image Based on U-Net convolutional Networks", 2020, IEEE

[23] Wei Zhang, Ping Tang, Lijun Zhao and Qingqing Huang, "A Comparative Study of U-Nets with Various Convolution Components for Building Extraction", 2019, IEEE

[24] Neil Micallef, Dylan Seychell and Claude Julien Bajada, "A Nested U-Net Approach for Brain Tumour Segmentation", 2020, IEEE

[25] Jierong Cheng, Wei Xiong, Wenyu Chen, Ying Gu and Yusha Li, "Pixel-level Crack Detection using U-Net", Proceedings of TENCON 2018 - 2018 IEEE Region 10 Conference, IEEE