International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 15, No.1 (Feb-24)

http://dx.doi.org/10.12785/ijcds/150149

Predicting Apple Yield Based on Occurrence of Phenological Stage in Conjunction With Soil and Weather Parameters

Rakesh Mohan Datt[1] **and Vinay Kukreja**[1]

[1]*Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, 140401 India*

Received 12 May. 2023, Revised 6 Jan. 2024, Accepted 21 Jan. 2024, Published 1 Feb. 2024

Abstract: Accurate and reliable yield forecasting is required for efficient planning and management of an important crop like apple. Efforts have been made to predict apple yield, mostly through the use of statistical tools with limited indicator parameters. The proposed neural network (NN) based system predicts yield of apple crops in an orchard based on identification, characterization, time of arrival and duration (ICTD) of phenological phases interactively with soil and weather parameters. The task of automatic yield estimation in orchards is challenging. Despite the significant amount of work that has been put into developing automated methods for estimating yields, the majority of methods currently in use are based on fruit counting, which is only useful one to four weeks before harvest. Whereas, in the proposed system, we will be predicting yield, during each phenological phase, among five classes, taking into account time of phenological stage occurrence (i.e. early occurrence, normal occurrence, or delay occurrence), soil parameter, and parameter related to weather conditions. This model will help the growers to timely take decision to execute contingency plans in case of average or negligible yield. The F-measure of the proposed system is 0.94 and with 95% accuracy. It is compared with other popular machine learning (ML) algorithms like Logistic regression, Support vector machines (SVM) and K-nearest neighbors (KNN).

Keywords: Artificial Neural Network (ANN), Deep Learning (DL), Yield estimation.

1. Introduction

From an economic standpoint, apples are a significant agricultural crop. There is a greater need for increased food production because of the growing population across the globe [1]. This is made more difficult by the fact that the quantity of land available for agriculture has shrunk as a result of urbanization [2]. Predicting apple yields has become an important research area in the context of climate change [3]. The weather conditions and consequent arrival as well as duration of phenological phases in apple, which used to be somewhat consistent a few decades back, are not the same these days. In most of the apple growing areas this has affected the yield adversely. It has been projected that there is a chance of a loss in crop production in India of between 10 and 40 percent owing to global warming by the years 2080 and 2100, as stated in the IPCC report [4]. To avoid frost damage, winter chill is essential for apple plant that falls dormant in the winter. Insufficient chilling affects flowering, fruit colouring, texture, and flavour [5]. If this ‘would be’ effect of these weather variability, in interaction with other parameters, on phenological phases is predicted well in advance, then the yield losses can be mitigated to some extent.

From the initial visible biological event to the last, phenological phases are the phases of crop growth that take place during a crop season. These phenological phenomena are divided into four phases in the case of apple crops: dormancy from December to March, flowering and fruit set from April to May, growth and development from June to September, and pre-dormancy from October to November [6]. Numerous studies have been undertaken on particular phenological phases, although the majority of these studies rely on manual observations [7]. The precise ICTD of a phenological stage, particularly in high-value fruits such as apple, is crucial for crop management, reducing adverse climatic impacts, and maximising yields. Accurate ICTD of phenological phases has long been a barrier for scientists, but this can be overcome with the use of cutting-edge techniques such as DL/ML. In horticulture, however, it is a novel technology [8]. DL has achieved considerable strides in a variety of fields [9]. The ICTD of phenological phases in interaction with climatic and soil parameters affect the quantitative production of apples. Generally, yield estimation of apples is assessed by mathematical models or fruit counting techniques that are based on real-time observations and require advanced mathematical know-how and equipment [10].

The main objective of the proposed system is to predict the apple yield in an orchard by taking into consideration factors like the occurrence of phenological phases, soil

E-mail address: rakeshdatt@gmail.com, onlyvinaykukreja@gmail.com

parameters, and weather parameters. The approach employed in the present study is different from the majority of conventional methods that focused on fruit identification and counting techniques using hi-tech equipment like UAVs mounted with LiDAR sensors or sophisticated cameras. Moreover, variable lighting conditions, image anomalies, background clutter, occlusions with other vegetative organs, and green fruit colour are some of the main factors that cause imprecision in fruit identification. In cases of high-density plantations, image-based detection techniques may not work as they will restrict the movement of the UAVs and autonomous orchard vehicle due to the smaller spacing among plants. Such technology is expensive and not affordable for most orchardists. Another limitation of such techniques is that the estimation can be made only a few weeks prior to harvest. Whereas, the technique that is being proposed will anticipate the apple yield well in advance, thereby giving plenty of time to design and implement a contingency plan to minimise the losses, if any. Besides, the technology that will be developed based on the outcome of the proposed study will require relatively low input costs, thus making it affordable.

The occurrence of phenological phases has a significant relationship with yield, and this information can be used to predict fruit yield too. A number of models have been proposed to predict apple yield, but models based on the occurrence of phenological phases in conjunction with soil and weather parameters for yield estimation have never been attempted for apple crops. Yield estimation well in advance, prior to harvesting, can help planners and policymakers design suitable post-harvest crop planning like storing, marketing, fixing the minimum support price, etc.

2. Related Work

The yield of an apple crop can be predicted on the basis of soil parameters, metrological (climate) parameters, morphological features and fruit counting using image analysis. In this section, we have reviewed the different methods used to predict crop yield. These methods include statistical models, ML algorithms, and DL models. Statistical models are commonly used to predict crop yield by analysing historical data and identifying patterns between various. ML algorithms, on the other hand, utilise advanced techniques such as regression and decision trees to make estimation based on large datasets. DL models, which involve deep neural networks (DNN), have shown promising results in accurately predicting crop yield by learning complex relationships between various factors.

A. Yield Estimation Techniques Based on Soil Parameters

Soils are made up of various types of minerals, organic matter, moisture content, and air. Lo Bianco (2019) [11] predicted apple output under deficit irrigation using water-related variables (soil physical parameters). According to the regression model built, total stomatal growth conductance and leaf water deficit influenced yield estimation significantly in the Gala apple variety. According to his regression model, stomatal conductance accounted for 79% of the variation in anticipated yield. Peng et al., (2017) [12] also provided a simulation model for estimating yield in Chinese apple orchards based on soil moisture, water use, and fertiliser.

In the past, researchers [13], [14] established a few models that used soil data, often in conjunction with weather parameters, to predict yields. Some researchers have employed soil properties in software tools specifically developed for yield estimation [6]. Attempts to estimate soil-based yields began four decades ago [15] and are currently ongoing [15]. Furthermore, the majority of these models have been built and used for arable crops, with little information known about such models being applied for apple crops.

B. Yield Estimation Techniques Based on Metrological (Climate) Parameters

Yazdanpanah et al., (2010) [16] used meteorological parameters to create a neural network model to predict several phenological phases of apple crop in the Golmakan region. The quantitative index was 77% using the intended technique. Sen et al., (2015) [17] using secondary meteorological data investigated the influence of climate on apple productivity and biodiversity in the Kullu valley area. Regression research found that the minimum temperature in January, February, and November, as well as the rainfall in December and the highest temperature in March and October, were major predictors of apple yield. Changes in global circulation patterns, according to Li et al., (2018) [18], alter meteorological parameters, which in turn affect apple yield [19].

Kouzegaran et al., (2020) [20] investigated the impact of climate extremes on saffron yield. The effects of climatic extreme indicators were investigated using regression analysis, and a saffron yield model was proposed by selecting the best indices. Using the indices R^2 = 0.68, RMSE = 0.6, and NRMSE = 14, the model's accuracy was monitored and evaluated. As a result, the magnitude of change in these meteorological characteristics or global circulation patterns could be used to forecast or simulate apple yield. Based on these observations, Li et al., (2020) [3] used three types of representative concentration pathways related to meteorological parameters in his simulation studies (of 28 apple producing counties) for apple yield estimation: RCPrf (reference/contemporary scenario), RCP 4.5 and RCP 8.5. The estimation results revealed two crucial findings: (1) that climatic factors had a substantially higher effect than meteorological disaster factors, and (2) that spring factors had a far greater influence than other seasonal components. In the future scenario RCP 4.5, 09 counties showed a minor decline; however, 02 counties showed a substantial decrease, 15 counties either maintained the same level or showed a slight increase, and the findings of 02 counties indicated a significant increase. The unit yield differential ranged between 1.44 and -1.85 t/ha. In the future scenario

RCP 8.5, 10 counties showed a minor decline, 02 counties showed a substantial decrease, 12 counties either maintained the same level or showed a slight increase, and 04 counties showed a significant increase. The unit yield differential ranged from 2.43 to -2.78 t/ha. The yield uncertainty grew over time for both future scenarios.

Hahn et al., (2023) [21] evaluated the effects of nitrogen fertilization, orchards, and cultivars on predicting yields of 'Royal Gala' and 'Fuji Suprema' apples in a subtropical climate. In a study conducted by Kuradusenge et al. (2023) [22], crop harvest estimations were made utilizing historical weather data and yield information through various ML techniques. The researchers observed that the Random Forest (RF) model exhibited superior accuracy in forecasting crop yield in comparison to other ML methods. A few researchers have earlier built weather-based yield estimation models, such as Info Crop, but all of these were limited to arable crops only [12], and no such specific model for apple yield estimation has been produced so far.

C. Yield Estimation Techniques Using Fruit Detection and Counting

Aggelopoulou et al., (2011) [23] developed a yield estimation method that estimated yield by analysing photos of a full bloom tree in an orchard. For 53 apple trees, the suggested approach produced a yield estimation error of 18%. The measured and estimated yields were statistically examined. Similarly, Zhou et al., (2012) [24] developed an image processing foundation method to identify, count, and predict the production of an apple orchard in Bonn, Germany. The colour characteristics of *cv.*'Gala' apple fruits were used by the authors. By picking RGB colour pixels against a white background, an algorithm was utilised to recognise apple fruits. 50 apple fruits were taken twice in natural daylight, once during ripening and once after June drop. For fruit detection during the ripening process, RGB and HSI colour schemes were utilised. R^2 values of 0.80 and 0.85 were reported for fruit counting and manually counting. For the number of apples anticipated and actual yield, R^2 was between 0.58 and 0.71.

Črtomir et al., (2012) [25] presented yield forecasting using a hybrid model comprised of ANN and an imaging processing technique. The model was implemented using the commercial programme Alyda Neurointelligence 2.07. The proposed ANN model surpassed the analysis, which was primarily performed by other image processing techniques. As a performance metric, R^2 and the standard deviation (SEE) were utilised. R^2 was 0.83 in the instance of "Golden Delicious," and SEE was 2.83 (kg/image). R^2 was 0.78 and SEE was 2.55 (kg/image) for the 'Braeburn' variety.

Wang et al (2013) [26] sought to estimate apple crop yield based on fruit count in an orchard in Washington State. The authors produced a bespoke dataset that included a high-resolution colour image of 1072 X 712 pixels captured by a dual camera installed on an autonomous orchard truck. Using the MATLAB programme 2010a, the authors created an image processing method for detecting red and green apples. The yield estimation error for a red and green apple block respectively, was -3.2% and 1.2%. Fruit load estimation is a key approach for planning and informing harvest resourcing and management, as well as marketing.

Gongal et al., (2016) [27] suggested an apple crop-load estimation technique based on fruit counting utilising an over-the-row machine vision system. The authors built a custom dataset comprising 424 photos. An image processing technology was used to detect and count apples. The system was built with the MATLAB R2012a programme. The proposed system's apple identification accuracy was 79.8%, while its crop load estimation accuracy was 82%.

Bargoti and Underwood (2016) [28] created a Multi-Layered Perceptron-based picture classifier (MLP). In an image, the suggested classifier model identified fruit and non-fruit portions. The unique training dataset with metadata was created by randomly picking 1100 sub-images at a Melbourne apple farm. Pylearn2, an open-source DL package, was used to develop the technique. The detection algorithm's f1 score was 0.721 without metadata and 0.743 with metadata. The proposed classification approach attained a R^2 value of 0.69 and a yield estimation accuracy of 81.6% without the need of metadata. With metadata, however, the R^2 value climbed to 0.78 and the yield estimation accuracy increased to 86.8%. Furthermore, Sa et al., (2016) [29] used VGG16 to create a model for fruit detection based on Faster R-CNN. Both rock melon and sweet pepper fruit were detected by the model. The model was trained and tested using 122 images in total. The Caffe DL framework was used to implement the proposed model. The model's F-measure was 0.838.

Chen et al., (2017) [30] created an apple-and-orange counting system using DL. In order to train the model, very few images of oranges and apples were used. The Caffe framework was used to implement the system. The orange crop had a mean IU of 0.813 and the apple crop had a mean IU of 0.838. Rahnemoonfar and Sheppard (2017) [31] developed a fruit counting technique to predict the number of tomatoes. The researchers built a dataset of 24K synthetic photos. Using the TensorFlow API, the authors constructed and implemented a modified Inception-ResNet CNN model. For over 100 photos, the system's average accuracy was 91.03%. Faster R-CNN based models were created by Bargoti and Underwood (2017a) [32] to identify apples, mangoes, and almonds. Models were developed using a unique dataset. In the cases of apples and mangoes, the F-score was more than 0.9. Bargoti and Underwood (2017b) [33] developed image segmentation-based apple yield estimation method based on multi-scale MLP and CNN. A custom data set of two apple kinds, Kanzi apple and Pink Lady, was generated, with almost 8,000 high resolution photos of 1232 X 1616 pixels apiece. From each high-resolution image, 32 sub-images of 308 X

202 pixels were extracted. Pixel-level labels were manually annotated for 1100 sub-images in both the fruit and non-fruit categories. The CNN model was fed image patches with a size of 48 X 48. Pylearn2 DL library was used to build the system. Pixel-wise segmentation techniques were used to identify and count fruits. The F-measure of the purposed approach was 0.861.

Dias et al., (2018) [34] created an apple blossom detection deep convolutional network (flowering stage). A previously built CNN model was adjusted to be especially sensitive to flowers. Using superpixel segmentation, this CNN model was then used to extract features. Extracted features were passed into a classifier, which determined whether or not each image contained flowers. The authors generated a bespoke collection of high-resolution photographs taken from various perspectives and distances. Using picture augmentation techniques, the samples were quadrupled to boost the training data. The dataset was labelled using the MATLAB GUI. The authors reported F-measure of more than 0.90. In neither of the aforementioned investigations was yield estimation performed. Cheng et al., (2017) [35], on the other hand, suggested a NN for early yield estimation based on apple tree canopy and fruit attributes using image analysis. During the growth season, a bespoke dataset of 150 samples was developed. Pixels were divided into multiple classes using image segmentation techniques, such as fruit, foliage, and background. The images were all scaled to 512 X 683 pixels. Matlab 2011b was used to implement the proposed model. The R^2 was 0.81, and RMSE was 2.34 kg/tree.

In their research, Roy et al., (2019) [36] suggested a yield estimation technique for apple orchards. Based on the fruit counting approach, the programme calculates the productivity of apple crops. The authors developed their own dataset of video clips. For apple recognition and fruit counting, a semi-supervised and GMM-based clustering technique was used. The suggested fruit detection algorithm had an F-measure of 0.95-0.97, the fruit counting method had an accuracy of 89-98%, and the yield estimation algorithm had an overall accuracy of 91.98-94.81%.

Tian et al., (2019) [37] created a YOLO-based algorithm that detects apples in an orchard throughout distinct growth phases in real time. The custom dataset of 480 original photos was expanded to 4800, which was then utilised to train the model. The Darknet framework was used to implement the proposed concept. The model's F-measure was 0.817. Yu et al., (2019) [38] suggested a DL model based on Fast RCNN and Single Shot Detection (SSD) to estimate fruit crop productivity. The proposed approach identified, counted, and estimated the yield of various types of fruit crops. By browsing Google Images with a Python crawler, a dataset of several sorts of fruit crops was built. Photos were then enhanced used for training of the model. Faster-RCNN has an accuracy of 89%, whereas SSD accuracy was 82%.

Gutiérrez et al., (2019) [39] created a unique on-tree yield forecast technique based on hyperspectral images (HSI) obtained from an autonomous ground vehicle (UGV) with 3D LIDAR sensor. The authors used a very simple CNN model. The model generated two output classes (mango and non-mango). During testing, R^2 versus manual count was 0.75 and 0.83 against RGB mango count. Gené-Mola et al., (2019) [36] use RGB-D cameras to create a faster R-CNN model for Fuji apple identification. The authors constructed a unique KFuji RGB-DS database with 967 images. The VGG-16 model is used in the initial convolutional layers for fruit detection. The F-measure of the suggested model is 0.898, and the AP is 94.8%. Meaningful information about yield estimation can also be derived without human involvement from multidimensional raw data acquired by advanced devices and sensors [40].

Apolo-Apolo et al., (2020) [41] created an advance hybrid DL algorithm using Faster R-CNN and LSTM. The suggested approach detects, counts, and estimates citrus fruit output and size. The authors built a bespoke dataset of 300 high-resolution photos captured with a UAV. Images were resized and enhanced further by rotating them at various angles. The training set includes 900 photos. To identify the citrus fruits in each image, a manual labelling method was used with the LabelImg software. TensorFlow API and Keras were used to create the proposed models. The LSTM model was trained to predict total production and yield per citrus tree. The proposed model has an F-measure of more than 89%. The standard error between manual counting and the model's autonomous fruit recognition was 6.59% on average.

Kang and Chen (2020) [42] developed a fruit detection technique in apple orchards. As training data, a custom dataset of 800 photos was employed, while 400 images were used for validation. The resolution of the training image was 320 X 320. To improve the training dataset, various data augmentation techniques such as brightness, saturation, contrast, and image rotation were used. Auto label creation was accomplished using the clustering-RCNN technique. A final f1- score of 0.826 was attained by the model.

Gené-Mola et al., (2020) [43] sought to develop a system for detection of fruit, estimation of yield, and canopy geometry characterisation utilising 3D LiDAR sensors. A method for identifying fruits based on reflectance and SVM was developed. MATLAB R2018a was used to implement the proposed approach. An RMSE of below 6% was achieved by the suggested model in predicting the yield. Using LiDAR and multispectral images data from UAVs, researchers create a channel for the automatic extraction of spectral and morphological aspects of apple trees. The combination of two frequently used algorithms i.e. SVN and K-NN led to the development of an ensemble ML yield estimation model. With an R^2 of 0.813, the ensemble learning model performs better than all base learners [44]. Oikonomidisa et al., (2022) [45] developed a hybrid CNN-

DNN model trained on soybean dataset that predicted the yield with R^2 of 0.87. Ge et. al., (2022) [46] created an apple yield estimation algorithm based on multi-feature fusion and SVM. The F-measure of the proposed method for yield estimation was 94.93%. The Hough transform and HSV conversion are used in the three-part recommended technique for forecasting by Saddik et al., (2023) [47]. It combines fruit detection, image acquisition, and counting operation. On the test dataset, the system was accurate and performed between the desired range of 95.04

A lot of studies about some selected phenological phase have been conducted for yield estimation but most of these studies are based fruit detection and counting techniques. Fruit identification and counting methodologies necessitate the use of specialised image data collection equipment such as sophisticated cameras, 3D LiDAR sensors and UAVs. Variable lighting circumstances, image anomalies, background clutter, occlusions with other vegetative organs, and green fruit colour are all major contributors to fruit identification imprecision. Furthermore, image-based identification and counting techniques can only be used one to four weeks before harvesting during the fruit growth phase, but the suggested technique will predict apple production well in advance, giving plenty of time to establish contingency plans to minimise any losses. Furthermore, the technology that will be created based on the results of the proposed study would have relatively low input costs, making it inexpensive. To fill these gaps and maximise returns from per unit resource input, some method that can predict yield based on ICTD of phenological phases in conjunction with soil and weather parameters is required. Such a study has never been conducted, resulting in a void in research continuity, particularly with regard to apple yield forecasts.

3. Materials and Method

A. Data Collection and Pre-processing

In this study, we utilized two distinct datasets. The first dataset included images of diverse phenological phases of apple crops, specifically curated for training a pre-trained CNN model, facilitating the recognition of various phenological phases. Additionally, a set of 75 manually devised rules was employed in the training process of the apple yield estimation model. These rules were designed, keeping in view all possible scenarios, with the help of domain experts after carefully analysing the historical raw data for the past 20 years. The aim was to ensure that the rules would effectively address any potential issues or challenges that may arise. The collaboration with domain experts and the thorough analysis of historical data allowed for a comprehensive understanding of the various scenarios that could occur.

Whereas, a dataset of high-quality photographs of eight phenological phases of apple crop was collected from an orchard in Badgam (34°1'12"N, 74°46'48"E), Srinagar, JK(UT), India. Around 1290 coloured photos were collected, and data augmentation techniques were used to increase the dataset size to around 8761 images. A domain specialist's expertise was used for manual classification and labeling of images, ensuring precise categorization and reliability for further analysis and research. The custom dataset was carefully chosen to achieve a diverse representation of apple crop phenological phases.

B. Proposed System Architecture

The proposed system consists of two sub-models i.e., a multilayer NN model having 5 layers for Yield Estimation (YE) and a custom CNN model for phenological phase recognition (PPR) [48] as shown in fig. 1. Using a CNN model, the PPR model was trained to distinguish eight distinct phenological phases of an apple crop. The PPR shows F-measure of 0.98. The YE model was trained using the handcrafted 75 rules as shown in Table III. The model was trained using leave-out-one cross validation (LOOCV) technique. In which, we leave out one entry from the dataset and use the remaining entries to train the model. The process is repeated for all the entries in the dataset. The hyperparameters like learning rate was set to 0.01, batch size was set to 5 and number of epochs was set to 25. Depending on the objectives and specific context, a model may be trained utilising the LOOCV technique if the dataset is small and well matched to the task at hand. After the training, the model was able to predict the yield with high accuracy among five distinct classes' i.e. high yield, moderately above average yield, average yield, moderately below average yield, negligible yield in an apple orchard. The structure of the YE model is shown in table II.

In order to make estimations, the YE model takes soil parameters like pH, Texture, Nitrogen, phosphorus, potassium, calcium, organic carbon, and cation exchange

TABLE I. Dataset of diverse phenological phases [48]

Stage	Phenological Phase	Total Images
1	Sprouting / bud development	1893
2	Leaf development	1777
3	Shoot development	481
4	Inflorescence emergence	1849
5	Flowering	1505
6	Development of fruit	430
7	Maturity of fruit	283
8	Senescence, beginning of dormancy	543

TABLE II. Structure of YE Model

Layer(type)	Output Shape	Param No.
dense (Dense)	(None, 13)	182
dense_1 (Dense)	(None, 13)	182
dense_2 (Dense)	(None, 26)	364
dense_3 (Dense)	(None, 13)	351
dense_4 (Dense)	(None, 6)	84

TABLE III. Training Dataset for YE Model.

Rule ID	Phenological Stage Arrival	Weather conditions	Soil conditions	Expected Yield
1	Delay	Very Poor	Very Poor	Negligible Yield
2	Delay	Very Poor	Below Average	Negligible Yield
3	Delay	Very Poor	Average	Negligible Yield
4	Delay	Very Poor	Above Average	Negligible Yield
5	Delay	Very Poor	Excellent	Moderately Below Average Yield
6	Delay	Below Average	Very Poor	Negligible Yield
7	Delay	Below Average	Below Average	Negligible Yield
8	Delay	Below Average	Average	Negligible Yield
9	Delay	Below Average	Above Average	Moderately Below Average Yield
10	Delay	Below Average	Excellent	Moderately Below Average Yield
11	Delay	Average	Very Poor	Negligible Yield
12	Delay	Average	Below Average	Negligible Yield
13	Delay	Average	Average	Moderately Below Average Yield
14	Delay	Average	Above Average	Moderately Below Average Yield
15	Delay	Average	Excellent	Moderately Below Average Yield
16	Delay	Above Average	Very Poor	Negligible Yield
17	Delay	Above Average	Below Average	Negligible Yield
18	Delay	Above Average	Average	Moderately Below Average Yield
19	Delay	Above Average	Above Average	Moderately Below Average Yield
20	Delay	Above Average	Excellent	Average Yield
21	Delay	Excellent	Very Poor	Negligible Yield
22	Delay	Excellent	Below Average	Negligible Yield
23	Delay	Excellent	Average	Moderately Below Average Yield
24	Delay	Excellent	Above Average	Average Yield
25	Delay	Excellent	Excellent	Average Yield
26	Advance	Very Poor	Very Poor	Negligible Yield
27	Advance	Very Poor	Below Average	Negligible Yield
28	Advance	Very Poor	Average	Negligible Yield
29	Advance	Very Poor	Above Average	Negligible Yield
30	Advance	Very Poor	Excellent	Moderately Below Average Yield
31	Advance	Below Average	Very Poor	Negligible Yield
32	Advance	Below Average	Below Average	Negligible Yield
33	Advance	Below Average	Average	Moderately Below Average Yield
34	Advance	Below Average	Above Average	Moderately Below Average Yield
35	Advance	Below Average	Excellent	Moderately Below Average Yield
36	Advance	Average	Very Poor	Negligible Yield
37	Advance	Average	Below Average	Moderately Below Average Yield
38	Advance	Average	Average	Moderately Below Average Yield
39	Advance	Average	Above Average	Average Yield
40	Advance	Average	Excellent	Average Yield
41	Advance	Above Average	Very Poor	Negligible Yield
42	Advance	Above Average	Below Average	Moderately Below Average Yield
43	Advance	Above Average	Average	Average Yield
44	Advance	Above Average	Above Average	Moderately Above Average Yield
45	Advance	Above Average	Excellent	Moderately Above Average Yield
46	Advance	Excellent	Very Poor	Negligible Yield
47	Advance	Excellent	Below Average	Moderately Below Average Yield
48	Advance	Excellent	Average	Average Yield
49	Advance	Excellent	Above Average	Moderately Above Average Yield
50	Advance	Excellent	Excellent	Moderately Above Average Yield

Rule ID	Phenological Stage Arrival	Weather conditions	Soil conditions	Expected Yield
51	Normal	Very Poor	Very Poor	Negligible Yield
52	Normal	Very Poor	Below Average	Negligible Yield
53	Normal	Very Poor	Average	Negligible Yield
54	Normal	Very Poor	Above Average	Negligible Yield
55	Normal	Very Poor	Excellent	Negligible Yield
56	Normal	Below Average	Very Poor	Negligible Yield
57	Normal	Below Average	Below Average	Negligible Yield
58	Normal	Below Average	Average	Moderately Below Average Yield
59	Normal	Below Average	Above Average	Moderately Below Average Yield
60	Normal	Below Average	Excellent	Moderately Below Average Yield
61	Normal	Average	Very Poor	Negligible Yield
62	Normal	Average	Below Average	Moderately Below Average Yield
63	Normal	Average	Average	Moderately Below Average Yield
64	Normal	Average	Above Average	Average Yield
65	Normal	Average	Excellent	Average Yield
66	Normal	Above Average	Very Poor	Negligible Yield
67	Normal	Above Average	Below Average	Moderately Below Average Yield
68	Normal	Above Average	Average	Average Yield
69	Normal	Above Average	Above Average	Moderately Above Average Yield
70	Normal	Above Average	Excellent	Moderately Above Average Yield
71	Normal	Excellent	Very Poor	Negligible Yield
72	Normal	Excellent	Below Average	Average Yield
73	Normal	Excellent	Average	Moderately Above Average Yield
74	Normal	Excellent	Above Average	High Yield
75	Normal	Excellent	Excellent	High Yield

capacity, Boron, Zinc, Iron and Microbial biomass as input which are than compared with the optimal values as prescribed by Sharma and Kumawat, (2019) [6] to generate Soil Quality Index (SQI) as shown in Table IV. Similarly,

TABLE IV. Soil Quality Index (SQI) and Weather Quality Index (WQI)

S.No.	Description	Soil Quality Index (SQI)	Weather Quality Index (WQI)
01	Very Poor	1	1
02	Below Average	2	2
03	Average	3	3
04	Above Average	4	4
05	Excellent	5	5

TABLE V. Phenological Stage Arrival Index (PSAI)

S.No.	Description	Phenological Stage Arrival Index (PSAI)
01	Delay	1
02	Advance	2
03	Normal	3

weather parameters like temperature, humidity, wind, rainfall during dormant and cropping season are compared with the optimal values [6] to generate Weather Quality Index (WQI) as shown in Table IV. phenological stage is detected with the help of PPR model [48] which takes an image of growth stage as an input. Arrival of the Phenological stage is determined by the comparing the detected stage with the Phenological phase calendar, in Julian day, of apple crop [6] to generate Phenological Stage Arrival Index (PSAI) as show in Table V. The present model is specifically designed for Golden Delicious variety of apple. The output of the model helps the grower to implement the contingency plans at the right time. Depending upon the output, the growers can plan the use of fertilizers, agrochemicals irrigation etc. If the output of the model indicates that the yield is “Average” it means that the grower needs to either increase the use of fertilizers / agrochemicals or increase the plant watering to maximize the yield. In case the predicted yield is “High Yield” then the growers can avoid the use of fertilizers / agrochemicals and save their valuable recourses. Apple growers can improve operational decisions and maximise yields by using this model.

4. Results and Discussions

The multilayer neural network trained with 75 handcrafted rules shows high F-measure of 0.94 and training accuracy of 95%. Table VI shows the classification report and fig. 2 shows the training loss / accuracy curve. Fig. 3 depicts the model’s confusion matrix. The suggested YE

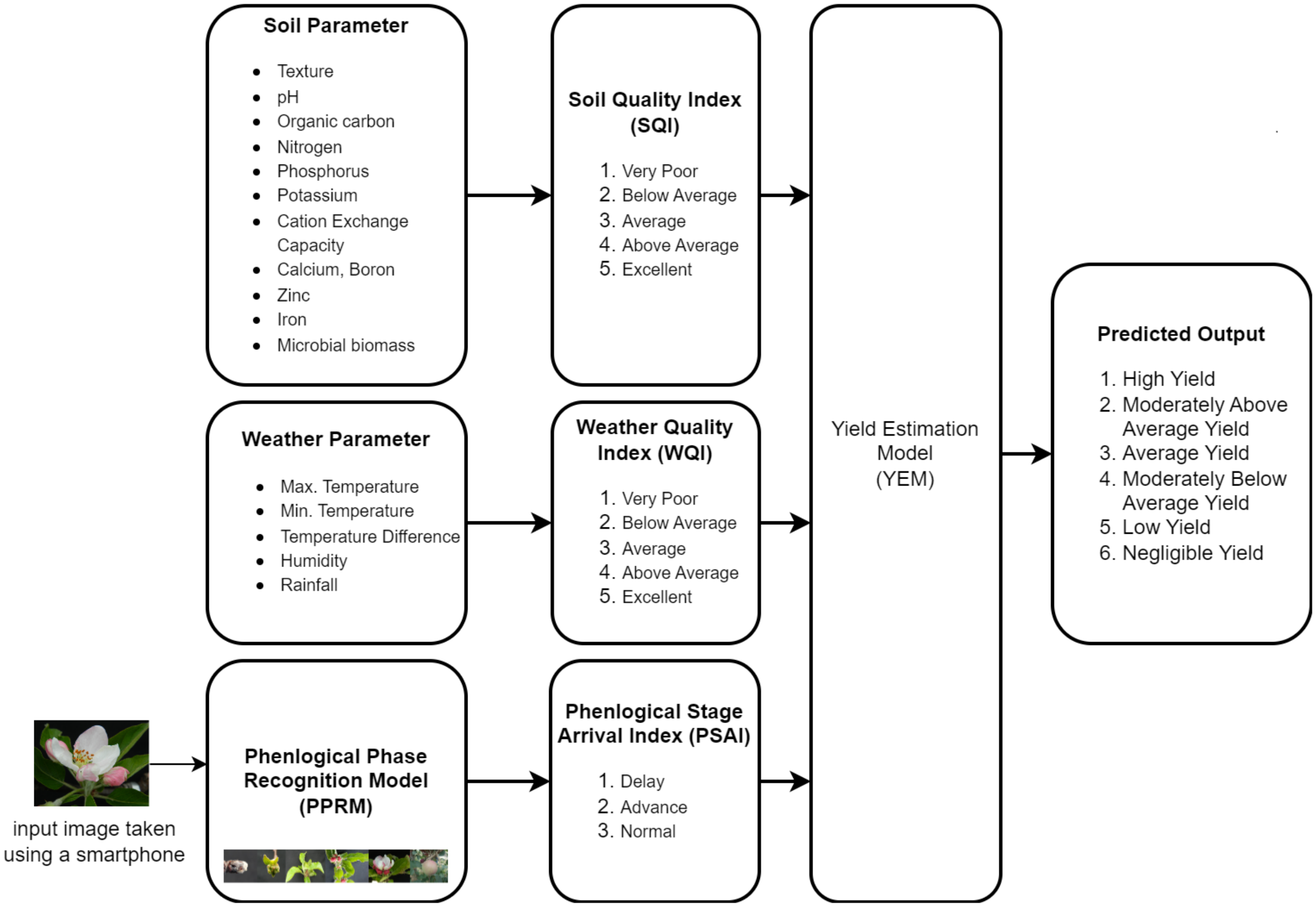


Figure 1. Proposed Model

model was evaluated with a number of other well-known ML techniques. Every one of the models was trained on the same dataset of manually generated rules using LOOCV technique. The confusion matrix generated using logistic regression can be seen in Fig. 4, while the classification report can be seen in Table VII. It was determined that the F-measure of the logistic regression was 0.60, and the accuracy of the logistic regression model was 0.78. The confusion matrix generated by the K-nearest neighbors (K-NN) model is presented in Fig. 5, respectively. The accuracy of the KNN was 0.60, while its F-measure was 0.43 as shown in Table VII. Since the Support Vector Machines (SVM) model uses the OneVsRest classification strategy

TABLE VI. Proposed Model Classification Report

S.No.	Precision	Recall	F-measure	Total Samples
1	0.94	0.94	0.94	32
2	0.92	0.96	0.94	23
3	0.91	0.91	0.91	11
4	1.0	0.86	0.92	7
5	1.00	1.00	1.00	2
Acc.	-	-	0.95	75
Avg.	0.95	0.93	0.94	75

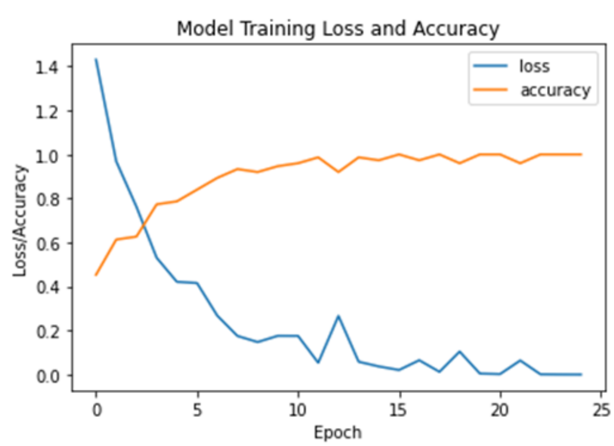


Figure 2. Training loss / accuracy curve.

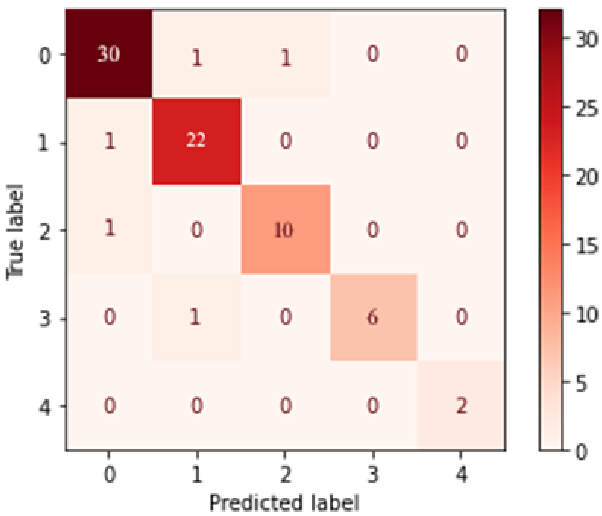


Figure 3. Confusion Matrix of the proposed model.

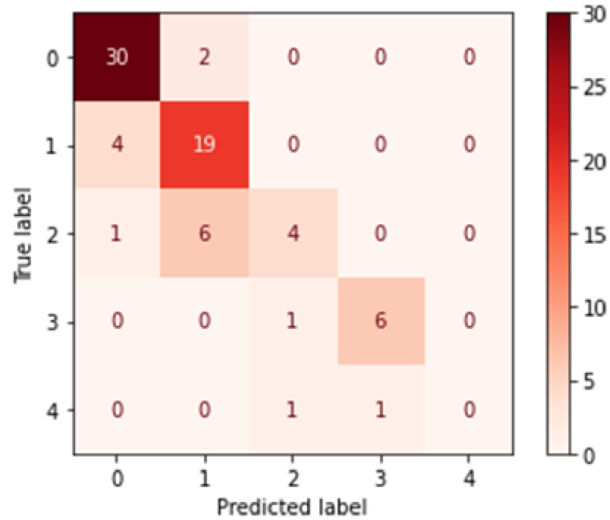


Figure 4. Logistic regression confusion matrix.

because it is a binary classifier, Fig. 6 presents the confusion matrix. The accuracy of the SVM was 0.81, while its F-measure was 0.80 as shown in Table VII. Fig. 7 shows the performance comparison all the four models.

With a macro average F-measure of 0.94, the suggested method is accurate and outperforms existing ML methods such as logistic regression, SVM, and K-NN when trained on the same dataset of manually generated rules using LOOCV. Out of all the models being studied, the accuracy of the K-NN model comes in at 60%. Comparatively, the SVM and logistic regression had accuracy of 81% and 79%, respectively. Despite the tremendous effort invested into creating automated systems for yield estimation, the majority of techniques in use today are based on fruit/ flower detection and counting or statistical algorithms with constrained indicator values. The proposed system predicts yield of apple crops in an orchard based on ICTD of phenological phases interactively with soil and weather parameters. We can anticipate the yield in the current system for each phenological stage, which will assist farmers in creating contingency plans to reduce any potential losses. The intended model was created exclusively for a certain apple variety, Golden Delicious, which is cultivated in the north-western region of Himalaya. For future research, on similar terms dataset and model can be also being designed for other well-known apple varieties.

5. Conclusions

In orchards, predicting yield automatically is a challenging task. The proposed NN based YE model estimates apple crop yield in orchards based on phenological stage ICTD, along with soil and weather parameters. The process of automating yield estimation has been extensively researched, but the majority of methods in use today are based on fruit counting, which can only be used few weeks prior to harvest. Contrarily, in the proposed system, we will

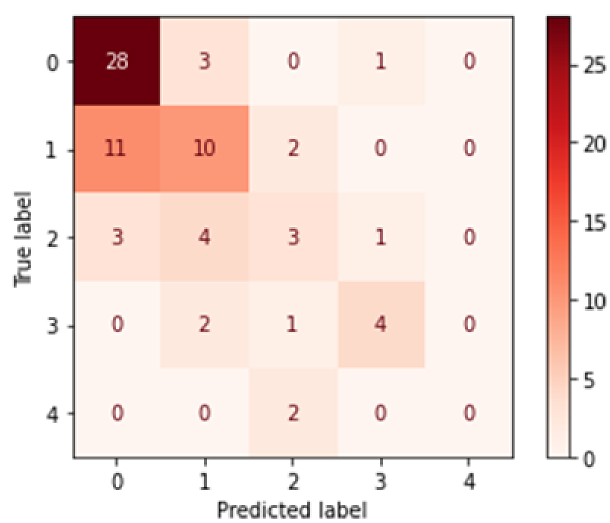


Figure 5. K-NN confusion matrix.

predict yield in five classes (High Yield, Moderately Above Average Yield, Average Yield, Moderately Below Average Yield, and Negligible Yield) by considering the timing of phenological stage occurrence (i.e. early, normal, or delayed occurrence), soil parameter, and weather parameter. The model's output enables the farmer to apply backup strategies at the appropriate moment. The usage of fertilizer, irrigation, and other measures may be planned by producers based on the predicted yield. If the model's output shows that the yield is "Average," the farmer must apply more fertilizer, agrochemicals, or water the plants more often in order to boost the production. If the expected output is "High Yield," growers can forego the application of agrochemicals and fertilizers and conserve vital resources. The proposed system performs better than other ML algorithms,

TABLE VII. Classification Report of ML models

	Logistic regression			K-NN			SVM		
	precision	recall	F-measure	precision	recall	F-measure	precision	recall	F-measure
1	0.86	0.94	0.9	0.67	0.88	0.76	0.86	0.97	0.91
2	0.7	0.83	0.76	0.53	0.43	0.48	0.7	0.83	0.76
3	0.67	0.36	0.47	0.38	0.27	0.32	0.75	0.27	0.4
4	0.86	0.86	0.86	0.67	0.57	0.62	1	0.86	0.92
5	0	0	0	0.00	0.00	0.00	1	1	1
Acc.	-	-	0.79	-	-	0.60			0.81
Avg.	0.62	0.6	0.6	0.45	0.43	0.43	0.86	0.78	0.81

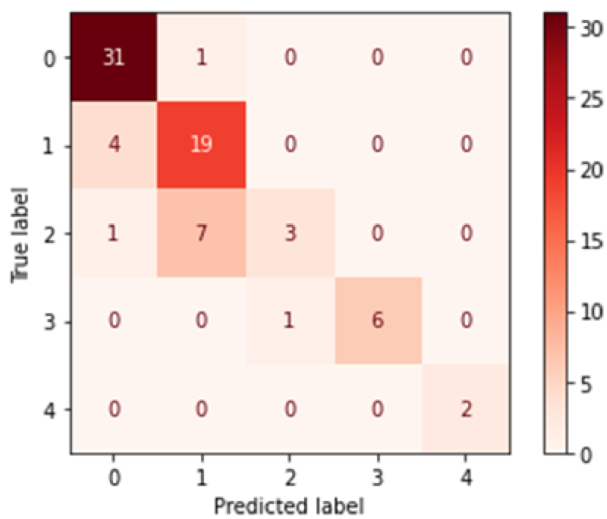


Figure 6. SVM confusion matrix.

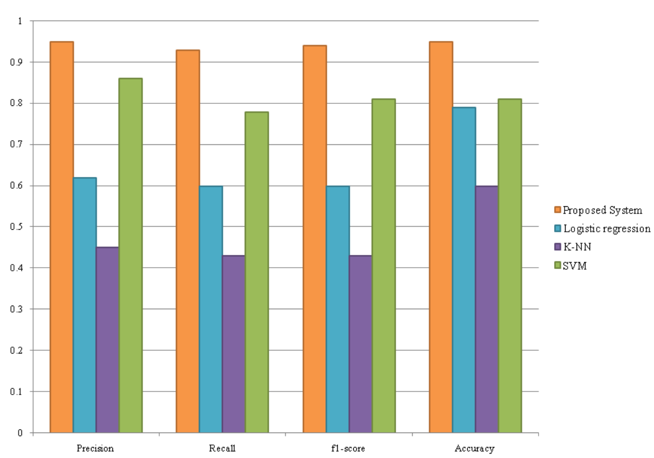


Figure 7. Performance comparison.

as evidenced by its macro average F-measure of 0.94 with accuracy of 0.95

In this study, we have developed a predictive model tailored for the cv. Golden Delicious, considering the intricate relationship between phenological phases and apple yield. However, it is important to acknowledge the limitations of our model, primarily its specificity to the Golden Delicious variety. The timing of phenological phases, a critical factor in apple yield estimation, varies significantly across different apple varieties. Therefore, for future research, there is a crucial need to enhance the versatility and applicability of our model. One promising avenue for future work involves the development of a more generalised predictive model that can accommodate the unique characteristics of various apple varieties. By incorporating a broader spectrum of apple types into our analysis, we can create a comprehensive framework that caters to the diverse needs of apple growers worldwide. This generalised model would enable accurate yield estimations for a wide array of apple cultivars, fostering a more inclusive and practical approach to orchard management.

Additionally, our current model primarily focuses on the relationship between phenological phases and yield. While phenology plays a pivotal role, it is essential to recognise the multifaceted nature of apple production. Several other factors significantly influence crop yield and quality. To enhance the predictive accuracy of our model, future research should explore the integration of additional variables such as detection of disease, pest, planting age and morphological features of apple trees. Understanding the intricate interplay between these factors and their impact on apple yield will provide a more holistic perspective. By collecting comprehensive data on these variables and leveraging advanced analytical techniques, we can refine our predictive model.

References

[1] K. Kour, D. Gupta, K. Gupta, S. Juneja, M. Kaur, A. H. Alharbi, and H. N. Lee, "Controlling agronomic variables of saffron crop using iot for sustainable agriculture," *Sustainability*, vol. 14, p. 5607, 2022.

[2] K. Kour, D. Gupta, K. Gupta, D. Anand, D. H. Elkamchouchi, C. M. Pérez-Oleaga, M. Ibrahim, and N. Goyal, "Monitoring ambient parameters in the iot precision agriculture scenario: An approach

to sensor selection and hydroponic saffron cultivation," *Sensors*, vol. 22, p. 8905, 2022.

[3] M. Li, J. Guo, J. He, C. Xu, J. Li, C. Mi, and S. Tao, "Possible impact of climate change on apple yield in northwest china," *Theoretical and Applied Climatology*, vol. 139, pp. 191–203, 2020.

[4] R. Pachauri and L. Meyer, "Climate change 2014: Synthesis report. contribution of working groups i, ii and iii to the fifth assessment report of the intergovernmental panel on climate change," 2014.

[5] R. Rai, S. Joshi, S. Roy, O. Singh, and A. Chandra, "Implications of changing climate on productivity of temperate fruit crops with special reference to apple," *Journal of Horticulture*, vol. 2, p. 1000135, 2015.

[6] A. Sharma and K. L. Kumawat, "Understanding flowering phenology and productivity of apple in response to climate change in prime apple growing areas of the country," ICAR-Central Institute of Temperate Horticulture, Srinagar, Tech. Rep., 2020.

[7] A. Sharma, A. Jain, P. Gupta, and V. Chowdary, "Machine learning applications for precision agriculture: A comprehensive review," *IEEE Access*, vol. 9, pp. 4843–4873, 2021.

[8] V. Meshram, K. Patil, V. Meshram, D. Hanchate, and S. Ramkteke, "Machine learning in agriculture domain: A state-of-art survey," *Artificial Intelligence in the Life Sciences*, vol. 1, p. 100010, 2021.

[9] L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis, "Machine learning in agriculture: A comprehensive updated review," *Sensors 2021, Vol. 21, Page 3758*, vol. 21, p. 3758, 2021.

[10] R. Akhter and S. A. Sofi, "Precision agriculture using iot data analytics and machine learning," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, pp. 5602–5618, 2022.

[11] R. L. Bianco, "Water-related variables for predicting yield of apple under deficit irrigation," *Horticulturae*, vol. 5, p. 8, 2019.

[12] X. Peng, Z. Guo, Y. Zhang, and J. Li, "Simulation of long-term yield and soil water consumption in apple orchards on the loess plateau, china, in response to fertilization," *Scientific Reports*, vol. 7, p. 5444, 2017.

[13] A. Sharma and S. Arora, "Soil quality indices and relative production efficiency for maize and wheat crops in agroclimates of northwest india," *Soil Science*, vol. 175, pp. 44–49, 2010.

[14] A. Sharma, N. Ahmed, S. Kumar, D. Kumar, O. C. Sharma, and S. A. Khan, "Production efficiency based land-use planning for almond - a new modus operandi," *Fruits*, vol. 72, pp. 247–257, 2017.

[15] D. D. L. Rosa, F. Cardona, and J. Almorza, "Crop yield predictions based on properties of soils in sevilla, spain," *Geoderma*, vol. 25, pp. 267–274, 1981.

[16] H. Yazdanpanah, D. Ohadi, and T. M. Soleimani, "Forecasting different phenological phases of apple using artificial neural network," *JOURNAL OF RESEARCH IN AGRICULTURAL SCIENCE*, vol. 6, pp. 97–106, 2010.

[17] V. Sen, R. S. Rana, R. C. Chauhan, and Aditya, "Impact of climate variability on apple production and diversity in kullu valley, himachal pradesh," *Indian Journal of Horticulture*, vol. 72, pp. 14–20, 2015.

[18] M. Li, J. Guo, C. Xu, Y. Lei, and J. Li, "Identifying climatic factors and circulation indices related to apple yield variation in main production areas of china," *Global Ecology and Conservation*, vol. 16, p. e00478, 2018.

[19] R. Bhattacharya, S. Bordoni, K. Suselj, and J. Teixeira, "Parameterization interactions in global aquaplanet simulations," *Journal of Advances in Modeling Earth Systems*, vol. 10, pp. 403–420, 2018.

[20] S. Kouzegaran, M. M. Baygi, I. Babaeian, and A. Khashei-Siuki, "Modeling of the saffron yield in central khorasan region based on meteorological extreme events," *Theoretical and Applied Climatology*, vol. 139, pp. 1207–1217, 2020.

[21] L. Hahn, C. Basso, J. M. Moura-Bueno, L. C. Argenta, M. Toselli, C. Carranca, M. Rech, I. S. Hahn, and G. Brunetto, "Yield prediction models for lsquo;royal galarsquo; and lsquo;fuji supremarsquo; apple varieties cultivated under a subtropical climate," *Agronomy*, vol. 13, no. 2, 2023.

[22] M. Kuradusenge, E. Hitimana, D. Hanyurwimfura, P. Rukundo, K. Mtonga, A. Mukasine, C. Uwitonze, J. Ngabonziza, and A. Uwamahoro, "Crop yield prediction using machine learning models: Case of irish potato and maize," *Agriculture*, vol. 13, no. 1, 2023.

[23] A. D. Aggelopoulou, D. Bochtis, S. Fountas, K. C. Swain, T. A. Gemtos, and G. D. Nanos, "Yield prediction in apple orchards based on image processing," *Precision Agriculture*, vol. 12, pp. 448–456, 2011.

[24] R. Zhou, L. Damerow, Y. Sun, and M. M. Blanke, "Using colour features of cv. 'gala' apple fruits in an orchard in image processing to predict yield," *Precision Agriculture*, vol. 13, pp. 568–580, 2012.

[25] R. Črtomir, C. Urška, T. Stanislav, S. Denis, P. Karmen, M. Pavlovič, and V. Marjan, "Application of neural networks and image visualization for early forecast of apple yield," *Erwerbs-Obstbau*, vol. 54, pp. 69–76, 6 2012.

[26] Q. Wang, S. Nuske, M. Bergerman, and S. Singh, "Automated crop yield estimation for apple orchards," *Experimental Robotics. Springer Tracts in Advanced Robotics*, vol. 88, pp. 745–758, 2013.

[27] A. Gongal, A. Silwal, S. Amatya, M. Karkee, Q. Zhang, and K. Lewis, "Apple crop-load estimation with over-the-row machine vision system," *Computers and Electronics in Agriculture*, vol. 120, pp. 26–35, 2016.

[28] S. Bargoti and J. Underwood, "Image classification with orchard metadata," 2016, pp. 5164–5170.

[29] I. Sa, Z. Ge, F. Dayoub, B. Upcroft, T. Perez, and C. McCool, "Deepfruits: A fruit detection system using deep neural networks," *Sensors*, vol. 16, p. 1222, 2016.

[30] S. W. Chen, S. S. Shivakumar, S. Dcunha, J. Das, E. Okon, C. Qu, C. J. Taylor, and V. Kumar, "Counting apples and oranges with deep learning: A data-driven approach," *IEEE Robotics and Automation Letters*, vol. 2, pp. 781–788, 2017.

[31] M. Rahnemoonfar and C. Sheppard, "Deep count: Fruit counting based on deep simulated learning," *Sensors*, vol. 17, p. 905, 2017.

[32] S. Bargoti and J. Underwood, "Deep fruit detection in orchards," 2017, pp. 3626–3633.

[33] S. Bargoti and J. P. Underwood, "Image segmentation for fruit detection and yield estimation in apple orchards," *Journal of Field Robotics*, vol. 34, pp. 1039–1060, 2017.

[34] P. A. Dias, A. Tabb, and H. Medeiros, "Apple flower detection using deep convolutional networks," *Computers in Industry*, vol. 99, pp. 17–28, 2018.

[35] H. Cheng, L. Damerow, Y. Sun, and M. Blanke, "Early yield prediction using image analysis of apple fruit and tree canopy features with neural networks," *Journal of Imaging*, vol. 3, p. 6, 2017.

[36] P. Roy, A. Kislay, P. A. Plonski, J. Luby, and V. Isler, "Vision-based preharvest yield mapping for apple orchards," *Computers and Electronics in Agriculture*, vol. 164, p. 104897, 2019.

[37] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved yolo-v3 model," *Computers and Electronics in Agriculture*, vol. 157, pp. 417–426, 2019.

[38] H. Yu, S. Song, S. Ma, and R. O. Sinnott, "Estimating fruit crop yield through deep learning," 2019, pp. 145–148.

[39] S. Gutiérrez, A. Wendel, and J. Underwood, "Ground based hyper-spectral imaging for extensive mango yield estimation," *Computers and Electronics in Agriculture*, vol. 157, pp. 126–135, 2019.

[40] A. Koirala, K. B. Walsh, Z. Wang, and C. McCarthy, "Deep learning – method overview and review of use for fruit detection and yield estimation," *Computers and Electronics in Agriculture*, vol. 162, pp. 219–234, 2019.

[41] O. Apolo-Apolo, J. Martínez-Guanter, G. Egea, P. Raja, and M. Pérez-Ruiz, "Deep learning techniques for estimation of the yield and size of citrus fruits using a uav," *European Journal of Agronomy*, vol. 115, p. 126030, 2020.

[42] H. Kang and C. Chen, "Fast implementation of real-time fruit detection in apple orchards using deep learning," *Computers and Electronics in Agriculture*, vol. 168, p. 105108, 2020.

[43] J. Gené-Mola, V. Vilaplana, J. R. Rosell-Polo, J.-R. Morros, J. Ruiz-Hidalgo, and E. Gregorio, "Multi-modal deep learning for fuji apple detection using rgb-d cameras and their radiometric capabilities," *Computers and Electronics in Agriculture*, vol. 162, pp. 689–698, 2019.

[44] R. Chen, C. Zhang, B. Xu, Y. Zhu, F. Zhao, S. Han, G. Yang, and H. Yang, "Predicting individual apple tree yield using uav multi-source remote sensing data and ensemble learning," *Computers and Electronics in Agriculture*, vol. 201, p. 107275, 2022.

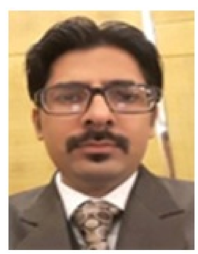
[45] A. Oikonomidis, C. Catal, and A. Kassahun, "Hybrid deep learning-based models for crop yield prediction," *https://doi.org/10.1080/08839514.2022.2031823*, vol. 36, 2022.

[46] L. Ge, K. Zou, H. Zhou, X. Yu, Y. Tan, C. Zhang, and W. Li, "Three dimensional apple tree organs classification and yield estimation algorithm based on multi-features fusion and support vector machine," *Information Processing in Agriculture*, vol. 9, pp. 431–442, 2022.

[47] A. Saddik, R. Latif, A. Z. Abualkishik, A. El Ouardi, and M. Elhoseny, "Sustainable yield prediction in agricultural areas based on fruit counting approach," *Sustainability*, vol. 15, no. 3, 2023.

[48] R. Datt and V. Kukreja, "Phenological stage recognition model for apple crops using transfer learning," 2022.

Rakesh Mohan datt an Assistant Professor in the dept. of Computer Application, Govt. College for Women, Gandhi Nagar, Jammu, JK(UT), India. Completed his Post Graduation (MCA) from the University of Jammu. Currently, he is pursuing PhD from the Dept. of Computer Science Engineering, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab. His area of scientific interest is artificial intelligence, machine learning, and deep learning.



Vinay Kukreja a Professor of Computer Science Engineering at Chitkara University in Punjab, India, earned his Ph.D. from the same university. He has a remarkable teaching experience of 17 years and has guided several Ph.D. and ME scholars. In addition to his teaching and guidance, he has an impressive track record with 87 articles indexed in Scopus and over 50 patents filed. Moreover, he has authored 3 books and edited 1 book. He received the first prize in SIH Hackathon in 2018 under the auspices of the Ministry of Housing Urban Affairs, India, which is a notable achievement. His research interests are primarily focused on machine learning, deep learning, agile software development, image processing, data analysis, and structural equation modeling