Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Estimation of Mangosteen fruit size from Macro Camera Images using YOLOv5 and Regression Analysis

**Vidhya N P**

npvidu@rediffmail.com

University of Kerala https://orcid.org/0000-0002-6135-7690

**Priya R**

University of Kerala https://orcid.org/0000-0001-7426-9780

**Chitra M T**

Kerala Agricultural University

**Shajeesh Jan P**

Kerala Agricultural University

**Research Article**

**Keywords:** Mangosteen, Macro Camera, YOLOv5, mean Average Precision (mAP), Multiple Regression Analysis, Relative Error

**Posted Date:** June 10th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-4455628/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

# Estimation of Mangosteen fruit size from Macro Camera Images using YOLOv5 and Regression Analysis

Vidhya N P[1], Priya R [2, *,] Chitra M T[3] and Shajeesh Jan P[4]

[1]1 Department of Computer Science, University of Kerala.
[2*] Department of Computer Science, Government College Kariavattom, University of Kerala.
[3] Department of Computer Science, Regional Agricultural Research Station, Wayanad, Kerala Agricultural University.
[4] Department of Agricultural Meteorology, Regional Agricultural Research Station, Wayanad, Kerala Agricultural University.

Corresponding author's E-mail: priyanil2007@gmail.com

**Abstract.** Accurate determination of fruit size is crucial in agricultural research for quality assessment and yield prediction. This study presents a novel ensembled approach for fruit detection using YOLOv5 and fruit size estimation using regression analysis, from macro-camera images of mangosteen fruits. YOLO, a state-of-the-art deep learning algorithm, is employed to automatically detect and localize mangosteen fruits in images captured by the macro camera. YOLO's high accuracy and real-time processing capabilities streamline the fruit detection process, reducing manual efforts and increasing the efficiency of data collection. Monthly field trails were conducted from May to August using macro camera for measuring the diameter of the fruit, which was observed to range from 1cm to 6.1cm. The deep learning technology, YOLOv5, for fruit detection gave 99.5% training accuracy in the 20th epoch. The results demonstrate that the proposed ensembled approach significantly improves the accuracy and efficiency of mangosteen fruit size determination compared to traditional methods. The acquired fruit size data is subjected to regression analysis, which establishes mathematical relationships between actual fruit size, which is a measurable parameter and the image size. The size of the mangosteen fruit is estimated from the bounding box parameters of the fruit using the multiple regression model. This study offers a comprehensive and reliable solution for fruit size estimation in agricultural research with relative error ranging from -0.004 to 0.0596, facilitating improved yield, quality control, and overall fruit production management.

.

**Keywords:** Mangosteen, Macro Camera, YOLOv5, mean Average Precision (mAP), Multiple Regression Analysis, Relative Error

## 1 Introduction

The mangosteen (Garcinia mangostana Linn. - GML) is a tropical fruit known for its exquisite taste and potential health benefits. Its rich flavour and nutritional content have led to increasing demand in the global market. It is a tropical tree from India, Myanmar, Malaysia, Philippines, Sri Lanka, and Thailand. This tree can reach 6–25 m and it has leathery, glabrous leaves and is slow to grow. The mangosteen-fruit is dark purple or reddish, with white, soft and juicy edible pulp with a slightly acid and sweet flavour and a pleasant aroma. Mangosteen is known as "the queen of fruits" because it is one of the best tasting tropical fruits [1]. Mangosteen fruit is about the size of a tangerine, averaging five to seven centimeters in diameter. According to Hidayat et al. the geometrical mean diameter of the mangosteen fruit is 54.79 ± 9.15 mm [2].

Many deep learning algorithms have been significantly employed in precision agriculture during the past few decades. The impact of applying machine learning and deep learning algorithms for plant leaf identification and pest/disease detection is highlighted in the research paper [3]. The dimensions of fruits or vegetables, as well as the quantification of their numbers, can be investigated through the application of computer vision and image analysis [4, 5]. The machine learning and deep learning techniques were utilized initially to detect the fruits and subsequently to calculate their sizes. The deep learning algorithm YOLOv5 enables accurate measurements of on-tree fruit sizes. This information helps the farmers for taking decisions about fertilization and crop management. Precise size estimation and counting of fruits empowers the farmers to optimize harvesting schedules efficiently, prioritizing areas with fruits of the desired size and minimizing the risk of harvesting underripe or overripe fruits [6, 7].

Designed to address the challenges of mangosteen fruit sizing, this study employs advanced image analysis techniques and the YOLOv5 deep learning object detection algorithm, the goal is to offer a robust and efficient solution, enhancing accuracy and scalability for improved agricultural practices, with YOLOv5 specifically designed to excel in detecting small objects [8]. Taking the advantage of YOLOv5's specific capability, a pretrained model is suggested for training the mangosteen image dataset. The main objective of this work is to identify on-tree mangosteen fruits and determine the image size, and thereby evaluate the fruit size by applying regression analysis. In precision agriculture, the widespread implementation of YOLOv5 for on-tree fruit size estimation has a significant impact, enhancing precision, efficiency, and quality.

YOLOv5 is an evolution of the YOLO architecture that excels in real-time object detection tasks. The original YOLO paper introduced the concept of real-time object detection by framing the task as a regression problem (Redmon et al., 2016). YOLOv5 was introduced by Ultralytics and its official repository can be found on GitHub. Its architecture consists of a backbone network (CSP-Darknet53), feature pyramid network, and detection head. YOLOv5's accurate detection capabilities contribute to precise size measurements. The algorithm ensures consistent measurements, minimizing variations due to human error. It can also help growers with better approximations on harvest materials and labour requirements by quickly providing an estimate of the distribution of fruit size for their crop [5]. Additionally, it assists farmers in strategic decision-making regarding crop management and market supply.

## 2 Related Works

Accurate size measurement of mangosteen fruits is crucial for various purposes, including yield, quality control, grading, packaging, and distribution. Currently, the processing centres calculate the fruit size and the quantity only after the fruits are harvested, weighed, and sorted using commercial grading machines. These manual measurement methods are time-consuming and subject to human errors. Application of deep learning techniques has been a breakthrough in this field of research. Qureshi et al. used K-nearest neighbour pixel classification and support vector machine classification for estimating the number of fruits in the images of mango trees [9]. On tree mango fruit size was assessed by Wang et al. using RGB-D Images, which is useful for estimation of fruit maturation and size distribution with RMSE 4.9mm [6]. The distance from camera to fruit obtained using RGB-D camera technology is a pre-requisite to find the fruit size in the work. Behera et al. used the Randomizes Hough Transform (RHT) technique to detect the mango fruit with 98% accuracy and the size was measured with 3.07% average error [10]. The colour of the fruit was also detected in the work. A colour CCD camera and a time-of-flight (TOF) light-based 3D camera was used by Gongal et al. to capture images of apple tree canopies for apple detection and size measurement [11]. The authors introduced a pixel size method to estimate the size of apple surface and the accuracy was increased by 84.8%. Using deep learning techniques, the yield and size of citrus fruits were calculated by Apolo-Apolo et al. using a UAV with standard Error 7.22% [7]. The apple fruits in an orchard were detected using Faster R-CNN, segmented using Mask R-CNN and size was estimated using Stereo vision camera images in the work of Mirbod et al. [5]. The diameter of the apple fruit was estimated with mean and standard deviation errors as 1.08 and 0.73 mm respectively.

The machine learning and deep learning techniques, including KNN, SVM, Alexnet, VGG19, ResNet50, DenseNet201, and MobileNetV2, were employed to identify and classify the healthy and two diseases – Sigatoka and Leafspot, of banana leaf [12, 13]. Incorporating deep learning algorithms like YOLOv5 can streamline and enhance the fruit detection process, enabling efficient and reliable size calculations. The application of deep learning techniques in the field of computer vision has revolutionized various industries, from autonomous vehicles to medical image analysis. According to Achyut Morbekar et al., YOLO effectively increases the speed of processing images to 45 frames per second in real-time, thereby improving the accuracy in the field of disease identification of plant leaf images [14]. The healthy and unhealthy plant leaves were classified with accuracy 82.38% with 2-3 seconds latencies using YOLO [15]. The objects which appear on the road can be classified into respective categories with the help of bounding boxes in the work of Sarda et al. with mAP 74.6% [16]. In the comparative study of Transformer, EfficientDet and YOLOv5 for crop circle detection in remote sensing imager, the YOLOv5 could detect more objects than others, which is addressed in the work of Mekhalfi et al. [17]. Bacterial spot disease in the ball pepper plant leaf was identified by Mathew and Mahesh et al. using YOLOv5 with mAP 90.7% [18]. Egi et al.

proposed a method which utilizes the drone footage of greenhouse to count the number of red and green tomato and the number of flowers using YOLOv5 with an accuracy of 0.618 at mAP 0.5 [19]. Ahmad et al. developed a model to identify and classify crop damaging insect pests to increase the production using YOLOv5 algorithm and proposed an automatic system in the form of a smartphone with accuracy 98.3% [20]. Tassel in RGB UAV Imagery was detected using improved YOLOv5 with mAP value of 44.7%, much better than other well-known approaches [21]. Dai et al. proposed a model to detect and grade the sprouted potatoes with accuracy 90.14% and mAP 88.1% using improved YOLOv5 model training [22]. Under the complex natural conditions, the plant diseases were recognized using an improved plant disease recognition model based on the original YOLOv5 network model. This improved YOLOv5 outperformed original YOLOv5 by 5.4% [23]. YOLOv5 network model algorithm gave faster and more accuracy than the previous versions of YOLO in the work of Karakaya et al. with 92.96% precision in the recognition and classification of five different diseases in tomato leaves [4]. Huang et al. proposed an improved YOLOv5 to detect tiny objects with 95% accuracy in mAP@0.5 [24]. The smaller objects were detected and classified by Benjumea et al. in the work for autonomous vehicle's performance, which showed 6.9% in mAP@0.5 [25]. The mango fruit were detected and count using computer vision and YOLOv5 deep learning technique and achieved 99.5% mAP@0.5 and mean absolute error of 1.5 [26].

By harnessing YOLOv5's object detection prowess, coupled with regression analysis to convert pixel coordinates into physical measurements holds immense promise in revolutionizing the assessment of mangosteen fruit dimensions. Multiple regression is often favoured due to its relatively lower complexity compared to other statistical models, making it a commonly used tool for addressing assumptions for the model predictions [27]. Between 2004 and 2013, Li et al. examined national food production and related indicators which were treated as independent variables, using multiple linear regression to assess multicollinearity [28]. The prediction model accurately estimated the grain output of the year 2014, demonstrating close alignment with actual values.

## 3 Materials and Methods

In this work, a novel ensembled approach is proposed to find the size of the mangosteen fruit from macro camera images. The proposed ensembled method incorporates YOLOv5 for on tree fruit identification and multiple regression analysis for fruit size calculation. The overview of the system is depicted in Fig. 1.

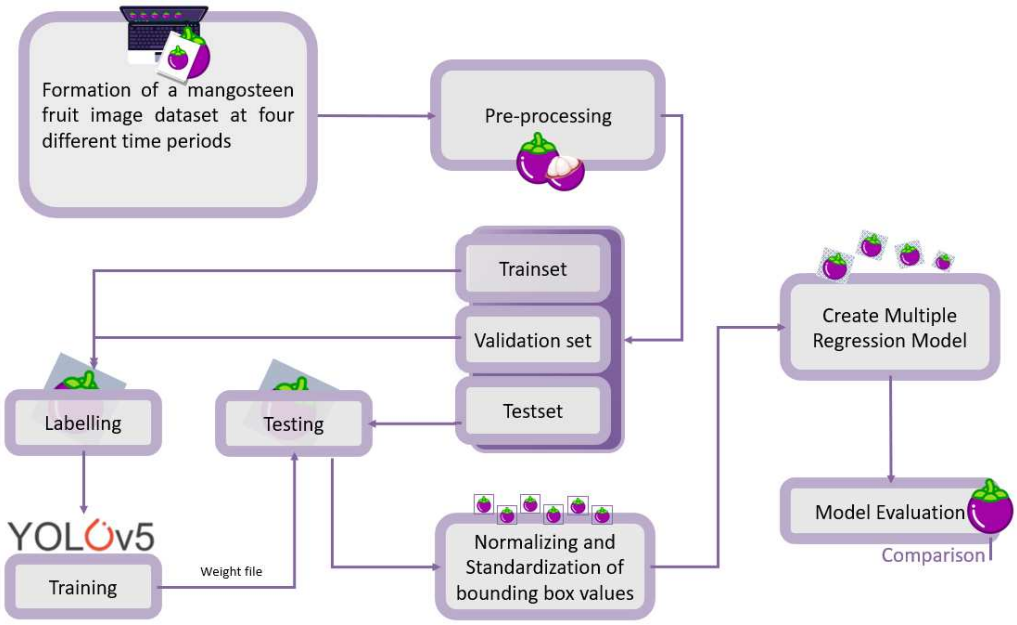



**Fig. 1.** Overview of the Mangosteen fruit size calculation.

### 3.1 Image collection using macro camera

The mangosteen images are collected from the research station of Kerala Agricultural University, Wayanad using the macro camera Nikon Z50 BKW/16-50MM F/3.5-6.3 VR KIT. The Z50 is an APS-C mirrorless camera announced by Nikon on October 10, 2019. The 16-50mm lens kit pairs the small but mighty Z 50 with the smallest Nikon DX lens ever, the NIKKOR Z 16-50mm f/3.5-6.3. This lens offers a versatile zoom range of 16-50mm, with Vibration Reduction (VR) image stabilization built-into the lens; and quiet operation, making it ideal for video as well as still images. The camera's image resolution is 5568 x 3712 pixels, and the images are saved in jpg format. The lens has a focal length of 50mm and the image sensor of the lens is 23.5 x 15.7 mm (APS-c) CMOS. Since the size of the image depends on the distance between the camera and object, six different distances are used to capture pictures of a single fruit, to normalize the value. The fruits are captured uniformly from early growth stage through the harvesting stage, ensuring an equal distribution across the dataset. It also includes images captured under various environmental conditions, with variations in lighting and the presence of shadows. The images are taken over different months, ranging from May to August.

### 3.2 Building the image dataset and annotation

The image set utilized for this work comprises pictures of Mangosteen fruits taken from the research station of Kerala Agricultural University, Wayanad, Kerala, India. The camera-to-fruit distance is adjusted between 20cm and 70cm, employing a 50mm focal length. The precise diameter of the fruits is determined using vernier callipers (Fig. 2), ranging from 1cm to 6.1cm, and these measurements are the ground truth in the study. The average size a mature mangosteen fruit is 5.2 cm ranging from 3.2 cm to 6.2 cm.



**Fig. 2.** Measuring fruit size using vernier callipers.

The dataset is pre-processed by renaming the images, normalizing pixel values, and handling class labels to ensure consistency and compatibility with the model. Each fruit was named with a combination of its distance and diameter. This collection of images was used to train the algorithm for fruit detection. Each individual fruit in the image was annotated by drawing bounding box using the 'labelImg' annotation tool. The fruit labelling process is depicted in Fig. 3.




**Fig. 3.** System overview of fruit labelling.

The text file generated as a result of the annotation process comprises of five values. These values include the image's class, followed by the coordinates of its center (x and y), and subsequently, the width (w) and height (h) of

the bounding box. The values x, y, w and h, referred to as normalised values, are converted to pixel values using the equations (1) to (4). A sample set of the normalized values and their corresponding pixel values of the bounding boxes are depicted in Table 1.

$$x\ in\ pixel\ =\ Norm.x\ *\ 5568 \tag{1}$$

$$y\ in\ pixel\ =\ Norm.y\ *\ 3712 \tag{2}$$

$$w\ in\ pixel\ =\ Norm.w\ *\ 5568 \tag{3}$$

$$h\ in\ pixel\ =\ Norm.h\ *\ 3712 \tag{4}$$

**Table 1.** Sample of the normalized values and their corresponding pixel values.

| Norm. x | Norm. y | Norm. w | Norm. h | x in pixel | y in pixel | w in pixel | h in pixel |
|---|---|---|---|---|---|---|---|
| 0.576149 | 0.518858 | 0.364583 | 0.565927 | 3207.9976 | 1926.0009 | 2029.99814 | 2100.72102 |
| 0.613685 | 0.583244 | 0.268267 | 0.329893 | 3416.9981 | 2165.00173 | 1493.71066 | 1224.56282 |
| 0.599946 | 0.360587 | 0.285062 | 0.363955 | 3340.4993 | 1338.49894 | 1587.22522 | 1351.00096 |
| 0.536099 | 0.425916 | 0.329537 | 0.492759 | 2984.9992 | 1581.00019 | 1834.86202 | 1829.12141 |
| 0.58046 | 0.378772 | 0.3916092 | 0.581918 | 3232.0013 | 1406.00166 | 2180.48003 | 2160.07962 |
| 0.57148 | 0.359375 | 0.163247 | 0.241853 | 3182.0006 | 1334 | 908.959296 | 897.758336 |
| 0.597971 | 0.39507 | 0.179533 | 0.272058 | 3329.5025 | 1466.49984 | 999.639744 | 1009.8793 |
| 0.216146 | 0.35695 | 0.200858 | 0.303198 | 1203.5009 | 1324.9984 | 1118.37734 | 1125.47098 |
| 0.527927 | 0.334995 | 0.26398 | 0.39695 | 2939.4975 | 1243.50144 | 1469.84064 | 1473.4784 |

This work utilized four quarter of datasets collected during various stages of fruit growth. Each quarter of dataset comprises of fruits with varying diameters, captured from six different distances 20cm, 30cm, 40cm, 50cm, 60cm and 70cm. Since the relationship between object's (fruit's) size and its distance from camera is an inverse linear relationship, capturing images at different distances helps improve the accuracy of fruit size estimation. A total of 855 images are included in the datasets. The first quarter contains 228 images, the second quarter contains 212 images, the third quarter contains 215 images and the fourth quarter contains 200 images. The dataset is then divided into 80% for training, 10% for validation, and 10% testing. The training dataset is employed to train the YOLOv5 algorithm using a pretrained weight file.

### 3.3 Deep learning technique for Fruit detection using YOLOv5

YOLOv5 was trained on the annotated trainset using transfer learning. The model learns to detect mangosteen fruits and predict their bounding boxes. One aspect of precision agriculture is estimation of crop yield at different stages of crop growth, to gain logistics decisions around harvesting, packing, transport and marketing. Crop yield estimation for fruit trees is currently based on manual counting of fruit and estimation of fruit size on a subset of trees. The manual process is tedious, time consuming and error prone, due to the high variability of crop yield in such crops. Although image analysis and machine vision have the potential to improve the speed and accuracy of fruit crop yield prediction, automated counting of fruits on-trees and calculation of fruit size remains challenging, given the unstructured, cluttered environment and the lack of distinguishing features on the fruit surface, relative to, for example, a human face.

In this work, the mangosteen fruits on tree are detected using the deep learning technique, YOLOv5, which provides accurate results for fruit identification. The pretrained model of YOLOv5 is used to perform inference on new images of mangosteen fruits. YOLOv5 includes techniques like auto-orientation, contrast adjustments, and

cropping which performs well in detecting the fruits across all these conditions, consistently producing accurate bounding boxes. YOLOv5's built-in techniques such as advanced data augmentation, robust feature extraction and enhanced architecture helps to improve model performance under challenging conditions like noise and blur (Redmon et al., 2016). The model is capable of effectively estimating the size of small fruits. However, estimating the size of occluded fruits is more challenging and the accuracy may vary depending on the extent of occlusion.

During the training time the fruits are detected in all images with 99.5% in mAP@50 and 92% in mAP@50-95. Since the mangosteen fruits are circular in shape, variations in orientation and rotation do not significantly affect the model's performance, however, the dataset contains different orientation images. A sample set of labelled fruits during training are depicted in Fig. 4.

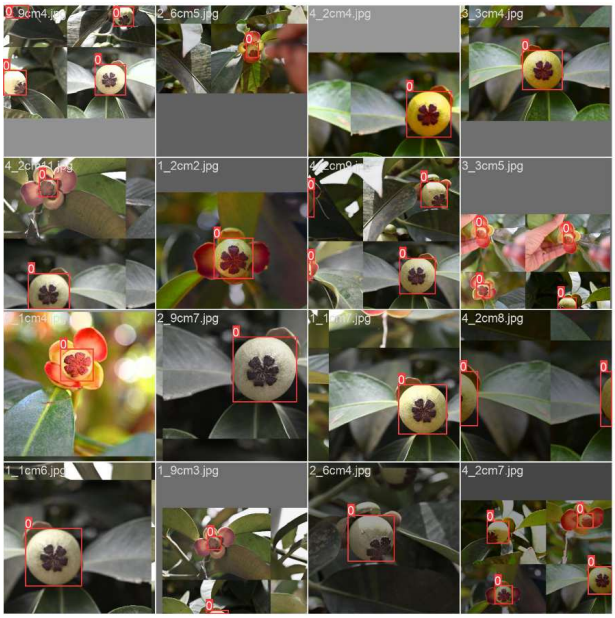



**Fig. 4.** Labelled images during training using YOLOv5.

Training and testing are performed on the data collected during the first, second, third, and fourth quarters, to determine the bounding box values for each identified fruit. Detected mangosteen fruit instances are outlined with bounding boxes and labelled with class identifiers. The pixel coordinates of the bounding box vertices are used to calculate the size of the detected mangosteen fruits. Conversion factors are applied to map pixel distances to real-world measurements. This approach reduces manual effort and accelerates the measurement process.

### 3.4 Computer Vision system of Fruit size evaluation

The computer vision, which enables systems to derive information from images, can be applied in various domains, such as agriculture, food processing, and quality control. The goal is to replace manual measurements with automated processes, saving time and reducing human error. The images captured, on which the system processes, should be of good quality and should provide clear views of the fruits for accurate fruit identification. Preprocessing techniques are applied to enhance the quality of the images. In order to accurately measure the size of individual fruits, the system needs to identify and isolate each fruit in the image. YOLOv5 transfer learning technique was used to separate the fruits from the background and from each other with bounding box. The system could be trained on a dataset of annotated fruit images to learn patterns and relationships between image features and actual fruit sizes. After labelling the fruits, the system extracts pertinent features for size evaluation. These features could encompass measurements related to the fruit's dimensions, including width, height, and diameter. Depending on the fruit's shape, various mathematical formulas may be employed to precisely calculate its size. In the case of a mangosteen, which exhibits a spherical shape, the approach utilized entails averaging the width and height derived from the bounding box to determine the fruit's diameter. Initially the diameter presented in pixel units, can be converted into millimeters (mm) (Equations (5) and (6)) and Table 2 shows the sample values.

$$width = \frac{23.5 * width\ in\ pixel}{5568} \quad (5)$$

$$height = \frac{15.7 * height\ in\ pixel}{3712} \quad (6)$$

where, 23.5 x 15.7 is the image sensor of the lens in mm and 5568 x 3712 is the image resolution in pixel.

**Table 2.** Sample values for average diameter in mm.

| Norm. w | Norm. h | w in pixel | h in pixel | w in mm | h in mm | Diameter (mm) |
|---|---|---|---|---|---|---|
| 0.364583 | 0.565927 | 2029.9981 | 2100.72102 | 8.5677005 | 8.8850539 | 8.7263772 |
| 0.268267 | 0.329893 | 1493.7107 | 1224.56282 | 6.3042745 | 5.1793201 | 5.7417973 |
| 0.285062 | 0.363955 | 1587.2252 | 1351.00096 | 6.698957 | 5.7140935 | 6.20652525 |
| 0.329537 | 0.492759 | 1834.862 | 1829.12141 | 7.7441195 | 7.7363163 | 7.7402179 |
| 0.3916092 | 0.581918 | 2180.48 | 2160.07962 | 9.2028162 | 9.1361126 | 9.1694644 |
| 0.163247 | 0.241853 | 908.9593 | 897.758336 | 3.8363045 | 3.7970921 | 3.8166983 |
| 0.179533 | 0.272058 | 999.63974 | 1009.8793 | 4.2190255 | 4.2713106 | 4.24516805 |
| 0.200858 | 0.303198 | 1118.3773 | 1125.47098 | 4.720163 | 4.7602086 | 4.7401858 |
| 0.26398 | 0.39695 | 1469.8406 | 1473.4784 | 6.20353 | 6.232115 | 6.2178225 |

For accurate measurements, the system needs to be calibrated using known reference objects or scales. This ensures that the pixel measurements correspond to real-world measurements. The actual diameter of the trained fruit image was measured in centimeters using vernier callipers (actual_diameter). The system can also be equipped to assess the quality of the fruits based on their sizes. For example, in the agricultural industry, larger or smaller than average sizes might indicate different stages of ripeness or potential defects.

### 3.5 Regression analysis model for size estimation

Regression analysis is a statistical technique that allow to create a model that estimates the actual size of the fruit based on the size of the fruit image. Multiple linear regression analysis is used to create the model by identifying a mathematical equation that describes the relationship between the known independent variables (predictor variable) and the dependent variable (response variable). A linear relationship should exist between the response variable and predictor variables. Regression analysis offers a structured and quantitative approach to size estimation, allowing the incorporation of various factors that influence an object's size. It's important to note that the success of this approach relies heavily on the quality and relevance of the collected data, as well as the assumptions underlying the chosen regression model. The independent variables used in this study are the distance from the camera to the fruit and the diameter of the fruit in the image and the dependent variable is the actual diameter of the fruit..These two variables exhibit a linear relationship, as an increase in distance typically results in a decrease in the observed diameter. Since these are the only input variables considered in the study, Multiple Linear Regression is an appropriate choice for estimating the actual size of the fruit based on their combined influence. Moreover, the variation in distance significantly influences the accuracy of the dependent variable, i.e., the estimated fruit size. The multiple linear regression equation (Equation (7)), is used to estimate the fruit size.

Here,

$y$ is the actual diameter of the fruit (measured using vernier callipers)

$x_1$ is the distance from the camera to fruit

$x_2$ is the diameter of the fruit image calculated from bounding box values

$\hat{y}$ is the predicted diameter of the fruit

Then the multiple linear regression equation can be

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 \tag{7}$$

where $b_0$, $b_1$, $b_2$ are the regression coefficients of the model, which are calculated from the equations Equation (8) to Equation (10)

$$b_0 = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 \tag{8}$$

$$b_1 = \frac{\sum {x_2}^2 . \sum x_1 y - \sum x_1 x_2 . \sum x_2 y}{\sum {x_1}^2 . \sum {x_2}^2 - (\sum x_1 x_2)^2} \tag{9}$$

$$b_2 = \frac{\sum {x_1}^2 . \sum x_2 y - \sum x_1 x_2 . \sum x_1 y}{\sum {x_1}^2 . \sum {x_2}^2 - (\sum x_1 x_2)^2} \tag{10}$$

Where $\sum {x_1}^2$, $\sum {x_2}^2$, $\sum x_1 x_2$, $\sum x_1 y$, and $\sum x_2 y$ are regression sums and calculated from the equations Equation (11) to Equation (15).

$$\sum {x_1}^2 = \sum_n {x_1}^2 - \frac{(\sum_n x_1)^2}{n} \tag{11}$$

$$\sum {x_2}^2 = \sum_n {x_2}^2 - \frac{(\sum_n x_2)^2}{n} \tag{12}$$

$$\sum x_1 x_2 = \sum x_1 x_2 - \frac{\sum_n x_1 \sum_n x_2}{n} \tag{13}$$

$$\sum x_1 y = \sum x_1 y - \frac{\sum_n x_1 \sum_n y}{n} \tag{14}$$

$$\sum x_2 y = \sum x_2 y - \frac{\sum_n x_2 \sum_n y}{n} \tag{15}$$

The bounding box values of the trained dataset and the distance of the camera to fruit are mapped to the corresponding dependent variable, to derive an equation using Equation (7). The equation thus derived is proved with the test fruit datasets, which indicates that the bounding box values obtained using YOLOv5 is accurate.

## 4 Results and Discussions

The dataset is formed by merging four quadrant datasets, which is then split into three parts, the train set (80%), the validation set (10%), and the test set (10%). The transfer learning method, YOLOv5, is applied to train the trainset for object detection with learning rate of 0.001, image size of 640x640 and batch size of 16. Python programming language is utilized for implementing the transfer learning procedure. The model is trained for 100 epochs and achieved an accuracy of 99.5% in the 20$^{th}$ epoch. The accuracy of mAP@0.5 is 99.5% and that of mAP@0.95 is 92% in the training period. Fig. 5 and Fig. 6 depicts the graphs generated during the training period. As the number of epochs increases, the box loss during the training and validation periods decreases. Moreover, precision, recall, mAP@0.5, and mAP@0.95 metrics show improvement with increasing epochs. These observations suggest that better results can be achieved through prolonged training and iteration.

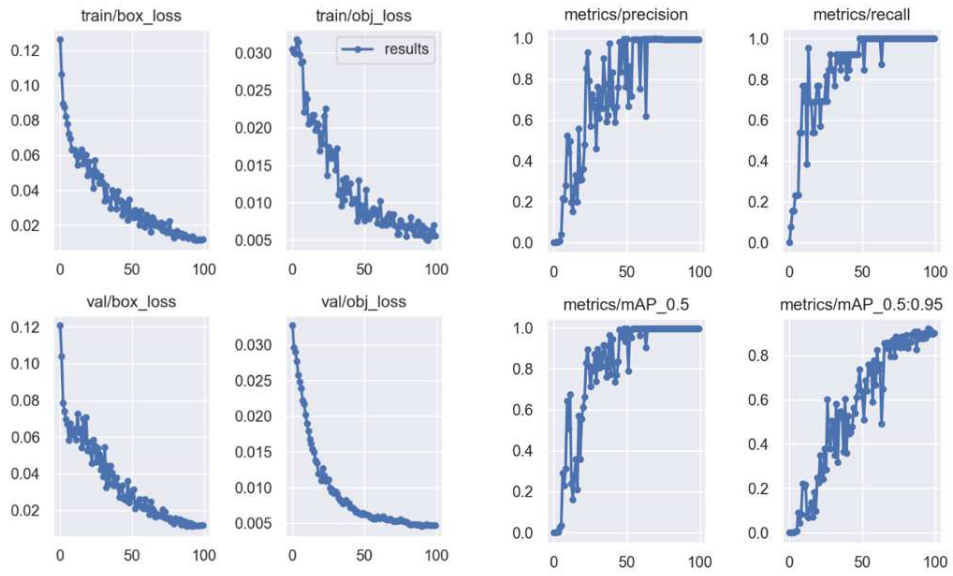



**Fig. 5.** Plot of Loss function convergence, precision, recall, and mAP after training the model.

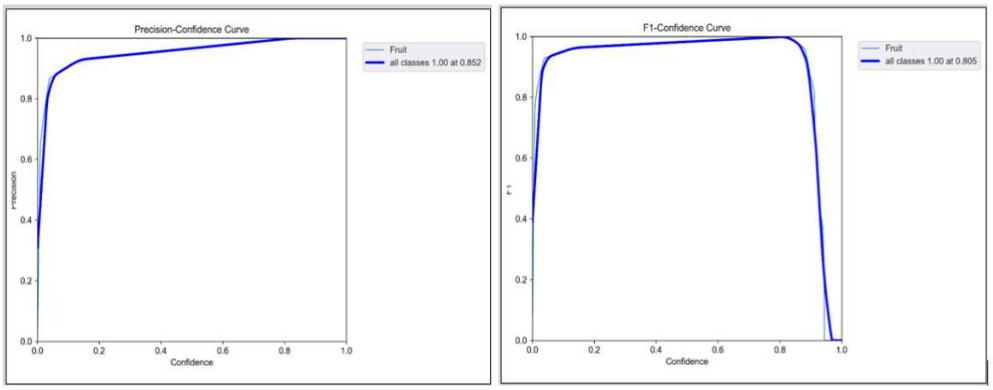



**Fig. 6.** Precision-Confidence and F-Confidence Curves.

The mAP@0.5 is the average precision of the fruit detection when IoU (Intersection over Union) is 0.5 and in mAP0.5:0.95, IoU value ranges from 0.5 to 0.95. The IoU can be calculated using Equation (16).

$$IoU = \frac{Area\ of\ overlap}{Area\ of\ union} \tag{16}$$

$mAP = \frac{\sum AP_k}{n}$, where AP is the average precision and n is the number of classes.

The Fig. 7 depicts the predicted bounding bod (red colour), and the ground truth bounding box (green colour) of the object tested. The IoU can be calculated using the Equation (16) and then calculate the confidence score of the object.

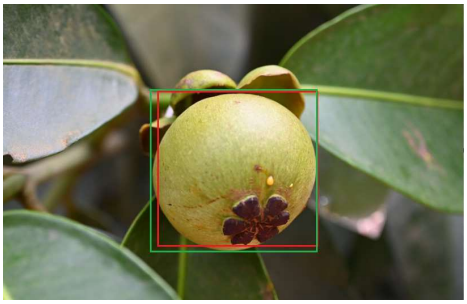


**Fig. 7.** Illustration for IoU estimation.

The training model was tested with remaining datasets and Fig. 8 shows the sample results obtained with fruit name and confidence score. The confidence score, expressed as a percentage, indicates the probability that the image will be correctly detected by the algorithm. The confidence score is computed through a logistic regression function, which relies on the IoU metric between the predicted bounding box and ground truth bounding box (Equation (17)).

$$Confidence\ Score\ =\ Probability\ Pr(Object)\ *\ IoU \tag{17}$$




**Fig. 8.** Testing images with bounding box and confidence score.

Since the mangosteen fruit is spherical in shape, the diameter of the fruit image was measured by calculating the average of the width (w) and height (h) values in the bounding box obtained. This average will provide an estimate of the fruit's image diameter (image_diameter). The diameter of the fruit was estimated from the image_diameter and distance from the camera to fruit. The multiple linear regression analysis is used to estimate the diameter of the fruit image (estimated_diameter). The distance from the camera to fruit and the estimated_diameter are considered as independent variables and actual_diameter as dependent variable of the multiple linear regression analysis. Table 3 gives the sample calculations of the multiple linear regression. The independent variables $x_1$ corresponds to the distance from camera to fruit and $x_2$ corresponds to image_diameter of image fruit, $\hat{y}$ is the estimated_diameter using regression analysis, and y is the actual_diameter measured using Vernier Callipers.

Multiple linear regression model thus predicts the size of the fruit from image, when distance from camera to fruit is known. The relative error is defined as the ratio of the estimated size of the fruit to the actual size. This error calculation determines the magnitude of the absolute error in terms of the actual size. The relative error indicates how accurate the estimation is relative to the size of the fruit being measured. Relative error can be calculated using Equation (18).

$$Relative\ Error\ =\ \frac{Estimated\ size - Actual\ size}{Actual\ Size} \tag{18}$$

Where, Estimated size is estimated_diameter, Actual size is actual_diameter.

**Table 3.** Sample of calculation of $\hat{y}$.

| actual_diameter (mm) | Distance (mm) | image_diameter (mm) | estimated_diameter (mm) |
|---|---|---|---|
| y | **x1** | **x2** | **y'** |
| 15 | 200 | 3.8166983 | 14.47837306 |
| 48 | 200 | 8.7263772 | 48.87743678 |
| 26 | 300 | 4.24516805 | 25.86812087 |
| 59 | 300 | 9.1694644 | 60.36959997 |
| 41 | 300 | 6.5169422 | 41.78502813 |
| 49 | 300 | 7.79506845 | 50.74006312 |
| 49 | 300 | 7.91540145 | 51.58316154 |
| 37 | 400 | 4.7401858 | 37.72412909 |
| 59 | 400 | 7.7402179 | 58.74348646 |
| 27 | 400 | 3.4392824 | 28.60950879 |
| 27 | 400 | 3.41183995 | 28.41723663 |
| 27 | 400 | 3.4096576 | 28.40194626 |
| 31 | 400 | 3.89330165 | 31.79053904 |
| 51 | 500 | 5.7417973 | 53.12952435 |
| 58 | 500 | 6.20652525 | 56.3855838 |
| 55 | 500 | 6.2178225 | 56.4647366 |

The relative error calculated and sample values of $\hat{y}$ are depicted in Table 4. It is found that the relative error ranges from -0.004 to 0.0596. Since the error range in size estimation is negligible, this model is suitable for real-world application, particularly for use by farmers in precision agriculture.

**Table 4.** Sample of calculation of relative error.

| y | y' | Relative Error |
|---|---|---|
| 15 | 14.47837 | -0.03477513 |
| 48 | 48.87744 | 0.018279933 |
| 26 | 25.86812 | -0.005072274 |
| 59 | 60.3696 | 0.023213559 |
| 41 | 41.78503 | 0.019147028 |
| 49 | 50.74006 | 0.035511492 |
| 49 | 51.58316 | 0.052717583 |
| 37 | 37.72413 | 0.019571056 |

| | | |
|---|---|---|
| 59 | 58.74349 | -0.004347687 |
| 27 | 28.60951 | 0.059611437 |
| 27 | 28.41724 | 0.052490245 |
| 27 | 28.40195 | 0.051923935 |
| 31 | 31.79054 | 0.025501259 |
| 51 | 53.12952 | 0.04175538 |
| 58 | 56.38558 | -0.027834762 |
| 55 | 56.46474 | 0.026631575 |

The proposed method is then compared with the other state-of-arts methods and for various fruits of different sizes are presented in Table 5. The proposed method demonstrates minimal error when compared to other methods. Dataset creation employs a macro camera, which is highly cost effective compared to 3D cameras and UAVs. However, the process of dataset creation can be time consuming. But the trained model can be used to detect the size of any image of the object.

**Table 5.** Comparison of proposed method with other state-or-the-arts methods.

| Sl. No. | Authors | Fruit | Fruit Size Estimation Method | Accuracy |
|---|---|---|---|---|
| 1 | W. S. Qureshi et al., 2016 [9] | Mango | K-nearest neighbour pixel classification and contour segmentation<br>Super-pixel over-segmentation and classification using SVM | F1 measure 0.68 and above |
| 2 | Z. Wang et al., 2017 [6] | Mango | RGB-D depth information, HOG feature, thin lens formula | RMSE 4.9 (length) and 4.3 (width) |
| 3 | S. K. Behara et al., 2018 [10] | Mango | Randomizes Hough Transform (RHT) | Average Error 3.07% |
| 4 | A. Gongal et al., 2018 [11] | Apple | 3D Coordinates (3DC) and Pixel Size (PS) | Average 30.9% (3DC), 15.2% (PS) |
| 5 | O. E. Apolo-Apolo et al., 2020 [7] | Citrus | Faster R-CNN Deep Learning Model using UAV | Standard Error 6.59% |
| 6 | O. Mirbod et al., 2023 [5] | Apple | Faster R-CNN and Mask R-CNN | Average Error 4.8% |
| 7 | Proposed Method | Mangosteen | Macro Camera Images, YOLOv5, and Regression Analysis | Relative Error ranges from -0.004 to 0.0596 (Average: 2.78%) |

## 5 Conclusion and Future Work

In conclusion, employing the YOLOv5 deep learning algorithm for mangosteen fruit detection and regression model for fruit size calculation represents a significant advancement in the field of fruit processing and distribution. This model is also versatile and can be used for calculating the size of any spherical-shaped fruit. By automating the size measurement process, this approach enhances accuracy and efficiency, contributing to improved quality control and customer satisfaction in the global mangosteen market. The computer vision system can be integrated into larger automated processes, such as sorting and grading systems in food processing facilities. This integration can

streamline the entire production or quality control process. Accurate size assessment aids in maintaining product quality and meeting market demands. The system can be scaled for large-scale fruit processing and distribution. YOLO makes detection faster and it outperforms other object detection methods. Since the model maintains its accuracy regardless of the camera used to capture the fruit images, it can be directly used by farmers as a mobile app, making it highly beneficial for them.

Though the data collection using macro camera is a tedious task, it is cost effective when compared to 3D cameras. As a future work, the results can be improved by generating a denser 3D point cloud to obtain a contour representation of the fruit. Using a sequence of images for each tree at different viewpoints will reduce the manual labour and also increase the dataset accuracy. Higher resolution lenses which can better resolve individual pixels will help to improve fruit boundary definition. Accurate camera calibration in combination with lower lens distortion should help to remove inaccuracies in diameter estimation for fruits located closer to the edges of an image. The thermal imaging technology could provide complementary information that enhances detection accuracy and system robustness across diverse environmental conditions. The total number of fruits on a tree can also be determined using the pretrained YOLOv5 model. Additionally, the system could be further trained to assess fruit maturity, enabling more precise harvesting decisions and contributing to better post-harvest management. This model can be extended to estimate the size of elliptical or any other shaped fruits in an orchard as well. The error calculation can be analysed to examine errors at different distances and also for different sizes of fruits. This analysis could lead to targeted improvements in the model's performance. Additionally, future studies should focus on quantifying the economic impact of deploying such a system in agricultural settings. Metrics such as labor savings, reduction in manual errors, and increase in harvest quality and yield can provide valuable insights into the profitability and scalability of this approach.

**Compliance with Ethical Standards**

**Disclosure of potential conflicts of interest**

The authors declare that he/she has no conflict of interest.

**Research involving human participants and/or animals**

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

**Informed Consent**

This article does not contain any studies with human participants or animals performed by any of the authors.

**References**


[1] Pedraza-Chaverri, J., Cárdenas-Rodríguez, N., Orozco-Ibarra, M., Perez-Rojas, J. M.: Medicinal properties of mangosteen (Garcinia mangostana). Food and chemical toxicology. 46(10), 3227-3239 (2008). https://doi.org/10.1016/j.fct.2008.07.024.

[2] Hidayat, D. D., Andriansyah, R. C. E., Indriati, A., Surahman, D. N., Luthfuiyanti, R.: Studies on geometrical, physical, mechanical and colour properties of mangosteen fruits. In IOP Conference Series: Earth and Environmental Science. Vol. 462, No. 1, p. 012045. IOP Publishing (2020). DOI 10.1088/1755-1315/462/1/012045.

[3] Vidhya, N. P., Priya, R.: Plant Leaf Disease Detection using Computer Vision-A Review. NeuroQuantology 20.15, 2261 (2022).

[4] Karakaya, M., Celebi, M. F., Gök, A. E., Ersoy, S.: Discovery of Agricultural Diseases by Deep Learning and Object Detection. Environmental Engineering and Management Journal. Vol. 21, no. 1, pp. 163-173 (2022). DOI:10.30638/eemj.2022.016.

[5] Mirbod, O., Choi, D., Heinemann, P. H., Marini, R. P., He, L.: On-tree apple fruit size estimation using stereo vision with deep learning-based occlusion handling. Biosystems Engineering. 226, 27-42 (2023). https://doi.org/10.1016/j.biosystemseng.2022.12.008.

[6] Wang, Z., Walsh, K. B., Verma, B.: On-tree mango fruit size estimation using RGB-D images. Sensors. 17(12), 2738 (2017). https://doi.org/10.3390/s17122738.

[7] Apolo-Apolo, O. E., Martínez-Guanter, J., Egea, G., Raja, P., Perez-Ruiz, M.: Deep learning techniques for estimation of the yield and size of citrus fruits using a UAV. European Journal of Agronomy. 115, 126030 (2020). https://doi.org/10.1016/j.eja.2020.126030.

[8] Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition. (pp. 779-788) (2016).

[9] Qureshi, W. S., Payne, A., Walsh, K. B., Linker, R., Chen, O., Dailey, M. N.: Machine vision for counting fruit on mango tree canopies. Precision Agriculture. 2016; 18, 224-244 (2016). DOI https://doi.org/10.1007/s11119-016-9458-5.

[10] Behera, S. K., Sangita, S., Sethy, P. K., Rath, A. K.: Image processing based detection & size estimation of fruit on mango tree canopies. International Journal of Applied Engineering Research. 13(4), 6-13 (2018).

[11] Gongal, A., Karkee, M., Amatya, S.: Apple fruit size estimation using a 3D machine vision system. Information Processing in Agriculture. 5(4), 498-503 (2018). https://doi.org/10.1016/j.inpa.2018.06.002.

[12] Vidhya, N. P., Priya, R.: Detection and Classification of Banana Leaf diseases using Machine Learning and Deep Learning Algorithms. IEEE 19th India Council International Conference (INDICON). IEEE (2022).

[13] Vidhya, N. P., Priya, R.: Image-Based Plant Leaf Disease Detection and Classification Using Deep Learning Models, 6th International Conference on "Emerging Technologies in Computer engineering: Industrial IOT and Cyber Physical systems" ICETCE 2023 (2023).

[14] Morbekar, A., Parihar, A., Jadhav, R.: Crop Disease Detection Using YOLO. International Conference for Emerging Technology (INCET), Belgaum, India. pp. 1-5 (2020). doi: 10.1109/INCET49848.2020.9153986. DOI: 10.1109/INCET49848.2020.9153986.

[15] Ponnusamy, V., Coumaran, A., Shunmugam, A. S., Rajaram, K., Senthivelavan, S.: Smart Glass: Real-Time Leaf Disease Detection using YOLO Transfer Learning. International Conference on Communication and Signal Processing (ICCSP), Chennai, India. pp. 1150-1154 (2020). doi: 10.1109/ICCSP48568.2020.9182146.

[16] Sarda, A., Dixit, S., Bhan, A.: Object Detection for Autonomous Driving using YOLO [You Only Look Once] algorithm. In Proceedings of the 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India. 4–6; pp. 1370–1374 (2021). DOI: 10.1109/ICICV50876.2021.9388577.

[17] Mekhalfi, M. L., Nicolo, C., Bazi, Y., Al Rahhai, M. M., Alsharif, N. A., Al Maghayreh, E.: Contrasting YOLOv5, Transformer, and EfficientDet Detectors for Crop Circle Detection in Desert. IEEE Geosci. Remote. Sens. Lett. 19, 288–302 (2022). DOI: 10.1109/LGRS.2021.3085139.

[18] Mathew, M. P., Mahesh, T. Y.: Leaf-based disease detection in bell pepper plant using YOLO v5. Signal Image Video Process. 16, 841–847(2022). DOI https://doi.org/10.1007/s11760-021-02024-y.

[19] Egi, Y., Hajyzadeh, M., Eyceyurt, E.: Drone-computer communication based tomato generative organ counting model using YOLO V5 and deep-sort. Agriculture (2022). https://doi.org/10.3390/agriculture12091290.

[20] Ahmad, I., Yang, Y., Yue, Y., Ye, C., Hassan, M., Cheng, X., Wu, Y., Zhang, Y.: Deep Learning Based Detector YOLOv5 for Identifying Insect Pests. Appl. Sci. 2, 10167 (2022). https://doi.org/10.3390/app121910167.

[21] Liu, W., Quijano, K., Crawford, M. M.: YOLOv5-Tassel: Detecting Tassels in RGB UAV Imagery With Improved YOLOv5 Based on Transfer Learning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. vol. 15, pp. 8085-8094 (2022). DOI: 10.1109/JSTARS.2022.3206399.

[22] Dai, G., Hu, L., Fan, J., Yan, S., Li, R.: A Deep Learning-Based Object Detection Scheme by Improving YOLOv5 for Sprouted Potatoes Datasets. IEEE Access, 10, 85416–85428, (2022). DOI: 10.1109/ACCESS.2022.3192406.
[23] Chen, Z., Wu, R., Lin, Y., Li, C., Chen, S., Yuan, Z., Chen, S., Zou, X.: Plant Disease Recognition Model Based on Improved YOLOv5. Agronomy 2022; 12, 365 (2022). https://doi.org/10.3390/agronomy12020365.
[24] Huang, T., Cheng, M., Yang, Y., Lv, X., Xu, J.: Tiny Object Detection based on YOLOv5. In Proceedings of the 5th International Conference on Image and Graphics Processing, Beijing, China. pp. 45–50 (2022). https://doi.org/10.1145/3512388.3512395.
[25] Benjumea, A., Teeti, I., Cuzzolin, F., Bradley, A.: YOLO-Z: Improving Small Object Detection in YOLOv5 for Autonomous Vehicles. arXiv.arXiv:2112.11798.2023 (2021). doi.org/10.48550/arXiv.2112.11798.
[26] Vidhya, N. P., Priya, R.: Automated On-Tree Mango Fruit Detection and Counting Through Computer Vision. Scope. Vol 13 (4) (2023).
[27] Ngo, T. H. D., Puente, L.: The steps to follow in a multiple regression analysis. In Proceedings of the SAS Global forum. (pp. 22-25) (2012).
[28] Li, Z., Cao, X., Ding, X., Chen, H.: Prediction Model of Multiple Linear Regression Analysis in Grain Production. 5th International Conference on Information Engineering for Mechanics and Materials (ICIMM 2015). DOI 10.2991/icimm-15.2015.23.

**Statements and Declarations**

**Funding**

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

**Competing Interests**

The authors have no relevant financial or non-financial interests to disclose.

**Author Contributions**

All authors contributed to the study conception and design. Data Collection were performed by Dr. Chitra M T, and Dr. Shajeesh Jan P. Material preparation, and Analysis were performed by Vidhya N P, and Dr. Priya R. The draft and final version was collectively contributed by all authors. All authors read and approved the final manuscript.

**Data Availability**

The datasets generated and/or analysed during the current study are available upon request.