

AN EFFECTIVE APPROACH FOR TURMERIC GROWTH DETECTION USING MULTILEVEL LINEAR ALGORITHM

S. REVATHY¹, Dr. S. KEVIN ANDREWS², Dr. P. S. RAJAKUMAR³ and Dr. V. N. RAJAVARMAN⁴

¹Research scholar, Department of Computer Science, Dr. M. G. R. Educational and Research Institute, Chennai, Tamil Nadu, India.

²Associate Professor, Department of Computer Applications, Dr. M. G. R. Educational and Research Institute, Chennai, Tamil Nadu, India.

^{3,4}Professor, Department of Computer Science and Engineering, Dr. M. G. R. Educational and Research Institute, Chennai-95, Tamil Nadu, India.

Abstract

Turmeric is a major cultivated crop in India. Some growing factors affect the yield and quality of turmeric production. In this paper, multilevel linear regression method is proposed to predict the growth of turmeric. The proposed method multilevel linear regression is predicting the turmeric growth level from turmeric yield dataset. Turmeric growth can predict through different statistical methods such as Linear Regression (LR), Polynomial Regression (PR), Decision Tree (DT) and Naïve Bayes. These algorithms are performing less due to prediction of turmeric growth based on accuracy and time. These algorithms give huge difference in prediction such as accuracy level and speed. To solve the above problem, turmeric yield data fed to the pre-trained for prediction of turmeric growth. The proposed method multilevel linear regression gives high accuracy prediction compared to other statistical algorithms. Multilevel linear regression method gives high accuracy of about 94% compared to conventional methods.

Keywords: Turmeric yield data, Polynomial Regression (PR), Multiple Linear Regression (MLR)

1. INTRODUCTION

Turmeric is the flowering plant of the ginger family and the rhizome of turmeric is used to cook and medical fields. Native to the Indian subcontinent and Southeast Asia, the plant is a perennial, rhizomatous, herbaceous plant that needs temperatures between 20 and 30 °C and a small quantity of annual rainfall to survive. Every year, turmeric plants are harvested for their rhizomes, some for eating and others for multiplication the following season. The primary component of turmeric, curcumin and it is used to make a bright orange-yellow powder that is frequently used as a colouring and flavouring agent in many Asian cuisines, particularly for curries. The rhizomes can also be used fresh or boiled in water and dried. Turmeric powder has an earthy, mustard-like scent and a warm, bitter, black pepper-like flavour. The Food and Drug administration of the United States, the European Parliament, and the World Health Organization have all given their seals of approval to the bright yellow chemical curcumin, which is produced by the turmeric plant. Although haridra, commonly known as turmeric, has been used for many years in Ayurvedic medicine, there is no reliable clinical proof that taking curcumin or turmeric will effectively treat any illness.

India has the highest number of different Curcuma species between 40 and 45 different species. Thailand has between thirty and forty species. There are various wild species of Curcuma in

other tropical Asian nations as well. Recent research has also revealed that it is difficult to classify *Curcuma longa*; only specimens from South India can be classified as *Curcuma longa*. Despite the fact that the people of Polynesia and Micronesia, in particular, have never interacted with India, they use turmeric extensively as a food and dye. An essential component of Ayurveda, Siddha, traditional Chinese, Unani, and the animistic ceremonies of Austronesian peoples, turmeric has been utilised for ages in Asia. Turmeric was referred to as "Indian saffron" in mediaeval Europe. A perennial herbaceous plant, turmeric can grow as tall as 1 meter. It has rhizomes that are cylindrical, fragrant, yellow to orange, and heavily branching. The leaves are placed in two rows and are alternated. Leaf sheath, petiole, and leaf blade are the three categories. A fake stem is created from the sheaths of the leaves. Stem bracts, which are absent of flowers and range in colour from white to green with occasional reddish-purple tinges, are found at the top of the inflorescence. The upper ends of these bracts are tapering. Only the median stamen of the inner circle is fertile, despite the average corolla lobe being larger than the two lateral.

Problem statement

The advance predicting methods are developed still the turmeric growth detection from turmeric yield data. Linear Regression (LR), Polynomial Regression (PR), Decision tree (DT) and naïve bayes classifier performs less during prediction. These methods give low accuracy, time complexity and low speed. Linear regression has limited variables such as input data and output data. The above problems are solved through the multilevel linear regression method.

Contributions

To predict the turmeric growth, predicting methods plays a vital role. To solve the above problem multilevel linear regression method is proposed.

- (i) To determine the turmeric through proposed method multilevel linear regression method from turmeric yield dataset
- (ii) To detect the high accuracy prediction of turmeric growth through proposed method multilevel linear regression.
- (iii) To prove the accuracy level of turmeric growth and avoid from turmeric disease

2. LITERATURE SURVEY

Turmeric grows in warm and humid climate. Its rhizomes are harvested annually after cultivation. Some growing factors affect the yield and quality of turmeric production. This paper proposes an inexpensive IoT platform for precision farming on turmeric cultivation, which indicates that the turmeric quality is significantly enhanced [1]. Turmeric is an Asian medicinal plant and used for its medicinal properties. Identification of the miRNAs takes part in the developmental process of turmeric through bioinformatics approach [2]. Changes in nutrient management can have impact on turmeric plant growth and product quality. Turmeric plant grown in artificial environments using sensors and WSN technology used to connect the required sensors [3]. The rhizome rot, one of the plant diseases, affects turmeric and it leads to

major downfall in production of the crop to the farmers. Disease detect through tree classifier and linear regression algorithm to forecast the rhizome rot of turmeric [4]. Traditional techniques for detecting turmeric plant sickness necessitated a large quantity of time, large research, and chronic farm monitoring. Artificial Intelligence (AI) used to detect turmeric growth performance in industries [5]. Adulteration of turmeric powders with spent turmeric is an increased demand for in-field analysis during authentication testing for quick decision-making in India. Machine learning used to detect the level of adulteration [6]. Metanil yellow are often mixed to turmeric powder for achieving the attractive yellow colour without change in taste. Machine vision based approach for detection of adulterant with turmeric powder is proposed [7]. NIR spectral data of pure turmeric powder and turmeric powder adulterated with varying degrees of Metanil yellow powder has been analyzed. The classification model was developed using PCADA and SIMCA to discriminate the pure turmeric samples and adulterated samples [8]. RGB colour sensor used to measure amount of Curcuminoids in Turmeric. The measured data of fresh Turmeric sample from the colour sensor in hue saturation lightness compared with amount of Curcuminoids obtained by conventional method of UV-spectrophotometry [9]. Turmeric is a major cultivated crop in the Erode region of south India and farmers find major leaf spot disease. Machine Learning methods utilized to recognize the illnesses in turmeric leaf [10]. Artificial Intelligence is an emerging sector in all fields of works to improve efficiency in agricultural sector and to improve crop yield by identify the disease affection at early stage [11]. The crop harvest is diminished by the disease and it affects the plant are leaf blotch, leaf spot and rhizome rot [12]. Multivariate calibrations including K-nearest neighbor (KNN) and Support Vector Machine (SVM) used to achieve clustering of the turmeric brands principal component analysis (PCA) [13]. The turmeric leaves exposed to diseases like Leaf Spot and Leaf Blotch and processed using k-Means image segmentation. Leaf images textural analysis carried out using GLCM [14]. Turmeric is a common spice used as a vital ingredient in Ayurvedic medicines and food. Conventional methods of identifying adulterants via chemical reactions are inaccurate [15].

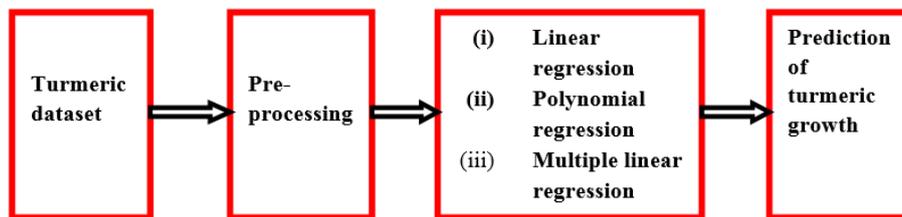
Inference from literature survey

Prediction of turmeric growth has different methods such as Artificial Intelligence (AI), IOT based methods, WSN, support vector machines (SVM), K-nearest neighbour (KNN), linear regression (LR), polynomial regression (PR), decision tree (DT) and naïve bayes. Decision tree gives inaccuracy and unstable to predict the growth level. Naïve bayes only work with independent variables and limited usability during prediction of turmeric growth. Linear regression works with two variables to predict the turmeric growth. Polynomial regression gives less accuracy. To solve the above problems multilevel linear regression method is proposed. The proposed method multilevel linear regression gives high accuracy prediction, good performance, high speed and less time.

3. METHODOLOGY

The Figure 1 represents the block diagram of proposed method multiple linear regressions using turmeric yield dataset to predict the turmeric growth level. The turmeric yield dataset has different variables such as rhizome colour, rhizome width, and weight per shoot, lamina length, plant height and number of shoots. These data are pre-processed then compared with different algorithms such as linear regression, polynomial regression and decision tree. These algorithms have some disadvantages to predict the dataset such as accuracy and speed. Therefore, multiple linear regression method proposed. The proposed method multiple linear regression used to predict the turmeric growth level with high accuracy and less time. This accurate prediction used to calculate the period and find diseases on plant.

Fig 1: Shows the Block Diagram of Multiple Linear Regressions



3.1 Linear Regression

Linear regression is used to find the linear relationship between one or more predictor variables and one output variable. Linear regression is frequently used for predictive analysis and modelling. Linear regression is also called as regression, multivariate regression, multiple regression and ordinary least square (OLS). Linear regression classified into two types such as simple linear regression and multiple linear regression. Simple linear regression is used to find the relationship between predictor and response. Predictor is independent variable and response is dependent variable. These variables have statistical relationship not deterministic relationship. One variable can be precisely expressed by the other is called deterministic. Statistical relationship is not accurately expressed the relationship between two variables. The main purpose of simple linear regression is to obtain the best-fit line for given data. Predictor error is small in the best-fit line and this error shows the distance between points to the regression line.

3.2 Polynomial Regression

An nth degree polynomial in x is used to model the relationship between the independent variable x and the dependent variable y in polynomial regression. A nonlinear relationship between the value of x and the corresponding conditional mean of y, denoted $E(y|x)$, can be fit via polynomial regression. Despite fitting a nonlinear model to the data using polynomial regression, the regression function $E(y|x)$ is linear in the unknown parameters that are estimated from the data, making polynomial regression a linear statistical estimation problem. The relationship between the independent and dependent variables is clearly stated by polynomial regression. The isotopes of the sediments are studied using it. It is employed to

research the spread of various diseases among any population. It is employed to research how any synthesis is created.

3.3 Multiple Linear Regression

Multiple linear regression is also called as multiple regression. Multiple linear regression is used to find the relationship between more predictor variables and outcome of response variable. Predictor is independent variable and response is dependent variable. Multiple regression is a statistical technique to find the output. Ordinary least square (OLS) is extending to multiple linear regression because it has more than one predictor variables. Multiple regression models are complex when there are more variables. Multiple linear regression calculates the line of best fit and minimizes the variance of the variables.

Fig 2: Shows Rhizome of Turmeric

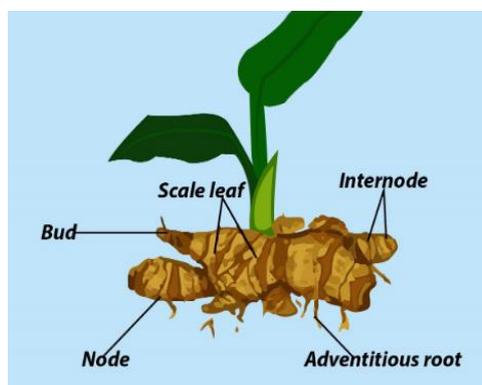


Fig 2 (a)



Fig 2 (b)

4. RESULTS AND DISCUSSION

The turmeric yield data predicted with different types of algorithms such as linear regression, multilevel linear regression, decision tree, naïve bayes and multilevel linear regression. Linear regression has some disadvantages such as sensitive to outliers, looks at the relationship between dependent and independent variables. Polynomial regression overcomes the linear regression disadvantages such analyse and ability to identify outliers but polynomial regression has disadvantage is comes down to the data being used. Decision tree algorithm has some disadvantages such as unstable, inaccurate and over fitting the data. Naïve bayes has few disadvantages such as zero frequency problem, limit usability and independent predictors. To overcome these problems, multilevel linear regression method is proposed. The proposed method multilevel linear regression gives high accuracy prediction, speed and good performance.

Table 1 shows the parameters such as rhizome colour, rhizome width, weight per shoot, lamina length, plant height and number of shoots and data of turmeric yield.

Tab 1: Parameters of Turmeric Yield

Parameters	PC1	PC2	PC3	PC4	PC5	PC6	PC7	pc8
Rhizome colour	-0.024	-0.042	-0.452	0.066	0.422	0.151	-0.477	-0.195
Rhizome width	0.438	-0.619	0.279	0.121	-0.036	-0.016	0.027	-0.013
Lamina length	0.855	0.172	-0.062	0.215	-0.080	-0.021	0.066	0.09
Plant height	0.903	0.112	0.024	0.212	-0.002	0.049	-0.045	0.148
Number of shoots	0.803	0.263	0.016	-0.217	-0.021	0.22	-0.144	0.016
Weight per shoot	0.825	0.132	0.114	-0.339	0.033	0.035	-0.167	-0.014

In **Table 1**, rhizome colour, rhizome width, lamina length, plant height and number of shoots are input parameters and Weight per shoot is an output parameter. Weight per shoot increase during increases of input parameters and these input parameters give high regression value with high accuracy. **Table 2 and 3** shows the output of polynomial regression and multiple linear regressions. **Table 4 and 5** shows the analyse variance of polynomial regression and multiple linear regression. **Table 6** shows the regression values of proposed method.

Tab 2: Polynomial Regressions for Turmeric Dataset

Coefficient	Estimate	Standard Error	t-statistic	p-value
β_0	-0.1178	0.1029	-1.1451	0.304
β_1	0.834	0.3203	2.6038	0.048
β_2	2.1051	0.7145	2.9463	0.032

In **Table 2**, turmeric growth predicted through polynomial regression by using rhizome width and weight per shoot from turmeric yield data. Rhizome width and weight per shoot have some parameters such as estimation, standard error, t-statistic and p-value for turmeric yield data. Standard error represents the average distance from the regression line of turmeric yield dataset. This standard error of turmeric yield data gives precious prediction through polynomial regression. Weight per shoot has less standard error (0.3203) through polynomial regression. Error of turmeric yield data calculate through polynomial regression equation,

$$\text{Weight per shoot} = 2.1051 * \text{Rhizome width}^2 + 0.834 * \text{Rhizome width} - 0.1178 \dots\dots (1)$$

Polynomial regression gives 33% error values and it gives 67% accuracy. The t-statistic represents the coefficient divided by its standard error of turmeric data. Rhizome width has better t-statistic value (2.9463) for polynomial regression. The p-value of turmeric data indicates the slope of regression line. P-value shows the significance linear relationship between dependent and independent variables. Rhizome width has good p-value (0.048) for polynomial regression.

Tab 3: Multiple Linear Regressions for Turmeric Dataset

Predictor	Coefficient	Estimate	Standard Error	t-statistic	p-value
Constant	β_0	-0.0509	0.0847	-0.6006	0.6091
Rhizome colour	β_1	-0.0977	0.2438	-0.4006	0.7275
Rhizome width	β_2	0.1303	0.3284	0.3966	0.73
Number of shoots	β_3	1.0219	0.3781	2.7028	0.114
Lamina length	β_4	-0.3676	1.1917	-0.3085	0.7869
Plant height	β_5	0.3183	1.4859	0.2142	0.8502

In **Table 3**, turmeric growth predicted through multilevel linear regression by using rhizome colour, rhizome width, and weight per shoot, lamina length, plant height and number of shoots variable from turmeric yield data. Rhizome colour, rhizome width, weight per shoot, lamina length, plant height and number of shoots variables have some parameters such as estimation, standard error, t-statistic and p-value for turmeric yield data. Rhizome colour has less standard error (0.2438) through multiple linear regressions. Error of depression data calculated through multiple linear regression equation,

$$\text{Weight per shoot} = -0.0509 - 0.0977 \cdot \text{Rhizome color} + 0.1303 \cdot \text{Rhizome width} + 1.0219 \cdot \text{Number of shoots} - 0.3676 \cdot \text{Lamina length} + 0.3183 \cdot \text{Plant height} \dots\dots\dots (2)$$

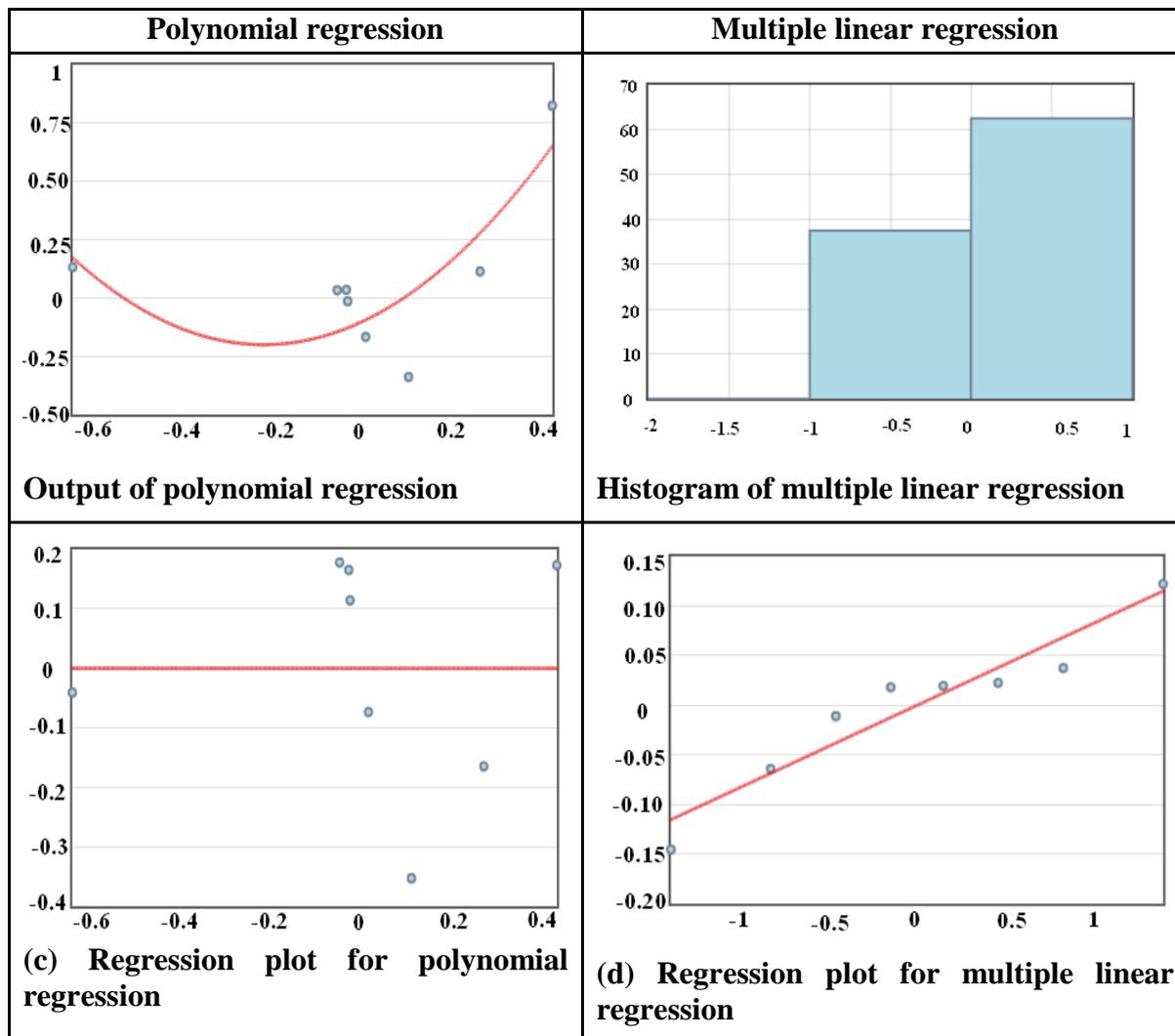
Multiple linear regression gives 6% error values and it gives 94% accuracy. The t-statistic represents the coefficient divided by its standard error of turmeric data. The rhizome colour has better t-statistic value (-0.4006) for multiple linear regression. The rhizome colour has good p-value (0.7275) for multiple linear regression. Compared to other methods multiple linear regression gives the best standard error, t-value, p-value and high accuracy.

Table 4: Analyse of Variance of Polynomial Regression

Source	df	SS	MS	F-statistic	p-value
Regression	2	0.5474	0.2737	5.2417	0.0593
Residual Error	5	0.2611	0.0522	-	-
Total	7	0.8085	0.1155	-	-

In **Table 4**, df represents degrees of freedom and it defined as sum of regression and residual degrees of freedom. Degrees of freedom used to calculate SS value. Total degrees of freedom (df) of turmeric data is 7. SS represents the sum of squares that used to find the relationship between dependent and independent variables of turmeric data. Total sum of squares (SS) value of turmeric data is 0.8085. MS represents mean squared error value of turmeric data. This mean squared value find through mean of sum of square value. Total mean square error (MS) value of turmeric data through polynomial regression is 0.1155. F-Statistic used to test the hypothesis of slope. F-statistic value of turmeric data is 5.2417. A p-value of polynomial regression is 0.0593. The values of df, SS, MS, F-statistic and p-values are calculated through polynomial regression from turmeric dataset and it gives less accuracy. **Figure 3** shows the output of multiple linear regression for turmeric yield data.

Fig 3: Shows Output of Multiple Linear Regressions



In **Figure 3**, (a) represents the output of polynomial regression and it is plotted between rhizome width and weight per shoot variables of turmeric data. A (b) represents the histogram of multiple linear regression and it is plotted between rhizome colour, rhizome width, lamina length, plant height and number of shoots are input parameters and Weight per shoot is an output parameter variables of turmeric data. This histogram shows the shape of large number of dataset with bar graph. A (c) shows the output of polynomial regression. X-axis contain rhizome width variable of turmeric data and y-axis contain weight per shoot variable of turmeric data. Polynomial regression shows the result as curve line and it gives less accuracy to predict the turmeric growth. A (d) shows the output of multiple linear regressions. X-axis contains above input variables of turmeric data and y-axis contains weight per shoot variable of turmeric data. Multiple linear regressions show the result as linear line and it gives high accuracy to predict the turmeric growth level.

Tab 5: Analyse of Variance of Multiple Linear Regression

Source	df	SS	MS	F-statistic	p-value
Regression	5	0.7654	0.1531	7.1165	0.1278
Residual Error	2	0.043	0.0215	-	-
Total	7	0.8085	0.1155	-	-

In **Table 5**, Total degrees of freedom (df) of turmeric yield data is 7. Total sum of squares (SS) value of turmeric data is 0.8085. Total mean square error (MS) value of turmeric data through multiple linear regression is 0.1155. F-statistic value of turmeric data is 7.1165. The p-value of multiple linear regression is 0.1278. The values of df, SS, MS, F-statistic and p-values are calculated through multiple linear regression from turmeric yield dataset. Compared to other statistical method multiple linear regressions give the best df, SS, MS, F-statistic and p-value for turmeric yield data.

Tab 6: Regression value of proposed method

Output	Polynomial Regression	Multiple Linear Regression
R-squared (R²)	0.6771	0.9468
Adjusted R-squared	0.6233	0.8137

In **Table 6**, multiple linear regression has high regression value as 0.9468 for turmeric dataset.

CONCLUSION

In this paper the proposed method, multiple linear regression predicts the turmeric growth from turmeric yield dataset. Compare to other statistical methods multiple linear regression method produce enhanced output for turmeric yield dataset. Multiple linear regression method gives satisfactory performance, response, regression and clear plots of turmeric yield dataset compared to other methods. Multiple linear regression method gives high accuracy about 94%, less time and speed for turmeric yield dataset. This prediction used to know growth level and avoid from plant disease.

REFERENCES

- 1) Chen, W.L., Lin, Y.B., Lin, Y.W., Chen, R., Liao, J.K., Ng, F.L., Chan, Y.Y., Liu, Y.C., Wang, C.C., Chiu, C.H. and Yen, T.H., 2019. AgriTalk: IoT for precision soil farming of turmeric cultivation. *IEEE Internet of Things Journal*, 6(3), pp.5209-5223.
- 2) N. Singh, S. Srivastava and A. Sharma, "A bioinformatics study of miRNAs and their regulating targets in *Curcuma longa* (turmeric)," 2016 International Conference on Bioinformatics and Systems Biology (BSB), 2016, pp. 1-3, doi: 10.1109/BSB.2016.7552160.
- 3) M. P. M and V. J., "Turmeric Farm Monitoring System using Wireless Sensor Network with ESP32 Module," 2021 Smart Technologies, Communication and Robotics (STCR), 2021, pp. 1-5, doi: 10.1109/STCR51658.2021.9588817.
- 4) V. Janani and N. M. S. Mangai, "Analysis and Classification of Rhizome rot Disease for Turmeric Plant using Artificial Intelligence," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), 2022, pp. 411-414, doi: 10.1109/ICSCDS53736.2022.9760884.

- 5) S. Velmurugan, K. Ramdinesh Reddy, S. G. Rahul, S. Vardhan, D. Subitha and S. K. Vignesh, "A Comparative Computational Analysis of VGG16 and VGG19 in Prediction of Turmeric plant Disease," 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), 2022, pp. 130-134, doi: 10.1109/ICACCS54159.2022.9784970.
- 6) A. R. Behera, H. Suresh, A. Kumar, S. K. Selvaraja and R. Pratap, "Detection of spent turmeric adulteration in powdered *Curcuma longa* using Vis-NIR spectroscopy and machine learning," 2020 5th IEEE International Conference on Emerging Electronics (ICEE), 2020, pp. 1-3, doi: 10.1109/ICEE50728.2020.9776996.
- 7) D. Mandal, A. Chatterjee and B. Tudu, "A machine vision based approach towards identification of adulterant in turmeric powder," 2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), 2017, pp. 6-9, doi: 10.1109/ICRCICN.2017.8234472.
- 8) R. Ranjan, N. Kumar, A. H. Kiranmayee and P. C. Panchariya, "Application of Handheld NIR Spectroscopy for Detection of Adulteration in Turmeric Powder," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 1238-1241, doi: 10.1109/ICACCS51430.2021.9441790.
- 9) S. Supannarach and D. Thanapatay, "The study of using RGB color sensor to measure the Curcuminoids amount in Turmeric (*Curcuma longa* Linn.) And Zedoary (*Curcuma Zedoarie* Rose.) by comparing colors with HSL system," 2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 2008, pp. 529-532, doi: 10.1109/ECTICON.2008.4600487.
- 10) C. N. Vanitha, S. Malathy, P. Shenbagavalli, S. A. Krishna and K. Kavin, "Detecting Turmeric Taphrina Maculans Disease using Machine Learning Algorithms," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2022, pp. 431-436, doi: 10.1109/ICAAIC53929.2022.9793291.
- 11) C. Rajasekaran., S. Arul., S. Devi., G. Gowtham. And S. Jeyaram., "Turmeric Plant Diseases Detection and Classification using Artificial Intelligence," 2020 International Conference on Communication and Signal Processing (ICCSP), 2020, pp. 1335-1339, doi: 10.1109/ICCSP48568.2020.9182255.
- 12) V. Janani, N. M. S. Mangai and R. Naveenkumar, "Analysis on agronomy and recent detection methods of rhizome rot in turmeric and methodology using artificial intelligence," 2021 6th International Conference on Communication and Electronics Systems (ICES), 2021, pp. 1-4, doi: 10.1109/ICES51350.2021.9488985.
- 13) S. Kar, H. Naskar, B. Tudu and R. Bandyopadhyay, "Comparison of Multivariable Techniques for Brand Classification of Turmeric Powders by Near-infrared (NIR) Spectroscopy," 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), 2018, pp. 98-102, doi: 10.1109/ICRCICN.2018.8718705.
- 14) G. Kuricheti and P. Supriya, "Computer Vision Based Turmeric Leaf Disease Detection and Classification: A Step to Smart Agriculture," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 545-549, doi: 10.1109/ICOEI.2019.8862706.
- 15) K. Nallappan, J. Dash, S. Ray and B. Pesala, "Identification of adulterants in turmeric powder using terahertz spectroscopy," 2013 38th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz), 2013, pp. 1-2, doi: 10.1109/IRMMW-THz.2013.6665688.