

# Modelling Plant Growth Based on Gompertz, Logistic Curve, Extreme Gradient Boosting and Light Gradient Boosting Models Using High Dimensional Image Derived Maize (*Zea mays* L.) Phenomic Data

Peter Gachoki<sup>1</sup>, Moses Muraya<sup>2</sup>, Gladys Njoroge<sup>1,\*</sup>

<sup>1</sup>Department of Physical Sciences, Chuka University, P.O Box 109-60400, Chuka, Kenya

<sup>2</sup>Department of Plant Sciences, Chuka University, P.O Box 109-60400, Chuka, Kenya

\*Corresponding author: [pkgachoki@gmail.com](mailto:pkgachoki@gmail.com)

Received July 13, 2022; Revised August 19, 2022; Accepted August 28, 2022

**Abstract** Modelling of plant growth is vital for hypotheses testing and carrying out virtual plant growth and development experiments, which may otherwise take a long time under field conditions. Modelling of plant growth has been aggravated by new phenotyping platforms that generate high dimensional data non-destructively over the entire growth time of a plant using a set of camera system. Such platforms generate high-throughput phenomic data, which is complex and constitute many features collected at multiple growth points for the same plant. However, the classical models are limited in that they can only model a single feature at a time. The objective of this study was to apply dynamic plant growth models that could be used to dissect complex relationships between plant growth and development using several modelling strategies. These included sigmoid, light GBM and XGBoost models. The image derived phenomic data was obtained from the Leibniz Institute of Plant Genetics and Crop Plant Research Gatersleben, Germany. The models were fitted using R statistical software and compared based on RMSE, R-squared, AIC and BIC performance metrics. The results showed that the XGBoost (RMSE = 2.1641) and Light GBM (RMSE = 2.7776) performed better than the Gompertz (RMSE = 3.8378) and the logistic function (RMSE = 3.8378) models in modelling maize plant growth. The XGBoost model (RMSE = 2.1641) showed better performance than Light GBM model (RMSE = 2.7776) in modelling maize plant growth. The Gompertz model using plant volume had AIC and BIC values for 139738.3 and 139763.4, respectively. The Gompertz model for plant side area had AIC and BIC values for 98436.15 and 98461.31, respectively. The logistic function model for plant volume had AIC and BIC values for 139749.2 and 139774.4, respectively. The logistic function model for plant side area had AIC and BIC values for 98415.95 and 98441.11, respectively. The Gompertz model and logistic function models showed almost the same performance in modelling maize plant growth. The non-parametric models, the XGBoost and light GBM, were found to perform better than the classical models (Gompertz and logistic functions) in modelling maize plant growth. Therefore, the study recommends the use of XGBoost as a generic model to fit high dimensional and complex phenotypic data in modelling plant growth and prediction of plant biomass yield at different growth points.

**Keywords:** Phenomic data, XGBoost, Light GBM, Gompertz model and logistic function model

**Cite This Article:** Peter Gachoki, Moses Muraya, and Gladys Njoroge, "Modelling Plant Growth Based on Gompertz, Logistic Curve, Extreme Gradient Boosting and Light Gradient Boosting Models Using High Dimensional Image Derived Maize (*Zea mays* L.) Phenomic Data." *American Journal of Applied Mathematics and Statistics*, vol. 10, no. 2 (2022): 52-64. doi: 10.12691/ajams-10-2-3.

## 1. Introduction

Modelling plant growth enables scientists and researchers to test theories and do virtual trials concerning plant growth and development that could have otherwise taken long time under field conditions [1]. The statistical models allow for interpretation of complex regulatory processes contributing to plant growth and development.

Such models are informative instrument that are helpful in research [2]. In addition, growth models help investigators to systematically analyse systems perturbations, develop hypotheses to guide the design of new experimental tests, and ultimately assess the suitability of specific novel traits.

Modelling of plant growth has become a key research area especially in the field of agriculture, environmental science and forestry [3]. This has been facilitated by the sharing of resources and experiences between the mathematicians, biologists and computer scientists. This

modelling of plant growth through integration of knowledge from various disciplines is necessary for advanced research in plant development and growth as well as in simulation studies [3]. Modelling of plant growth such as maize can be important in improving production and yield. This is because modelling can help to understand the dynamics of plant growth and development, and yield responses under different growth conditions [4].

The use of plant growth models to assess genotypes performance across diverse target environments, can help understand suitability of trait variation, and possibly accelerate plant breeding programs. However, the models to support these tasks are still under development and testing. Moreover, currently the phenotyping methods have changed, including use large-scale imaging phenotyping, which derive large, high dimensional and complex data set on all aspect of plant growth and development [5,6,7,8]. Such large and high dimensional data may not be appropriately fitted by existing models [9]. Therefore, the existing model may require an enabling data inference technology in addition to algorithm development with the aim of analytically solving complex and high dimensional data sets [10,11]. Phenotyping has advanced with the application of high throughput phenotyping such as use of automated imaging techniques [8]. This has led to derivation of large quantities of high dimensional data or phenotypic features that could not have been achieved using manual phenotyping in a single run [7]. This called for parallel development of statistical models that can appropriately handle such high dimensional data. Information on such models is limited. Though existing models such as the sigmoid models have been found to predict plant growth with high degree of precision, information on their performance using high dimensional data was still limited.

Models exist that have been used in modelling plant growth and development. However, such models have been used to fit low dimensional data derived using manual measurement [3,12]. The manually collected is limited to few features and the same plant cannot be measured in subsequent growth stages since the plant is destroyed during the data collection. Hence, such models may be inefficient in fitting the high dimensional data that is derived using high-throughput techniques and at different growth stages for the same plant. Some of the existing models for plant growth include the sigmoid models, simple linear regression models, multiple linear regression models and non-linear models such as the quadratic and the cubic models.

Sigmoid models are better suited to individual plant growth and development [5]. These models include the exponential and monomolecular models [13]. These models have been found useful in interpreting individual growth patterns [3]. Linear models have also been developed by expressing the dry weight of the shoot as a function of plant area and plant age [14]. Using the same variables, quadratic and cubic models were also with the linear regression model showing better performance as compared to the fitted non-linear models [14]. Moreover, simple linear regression and multiple linear regression

models had been used to model growth and good performance has been exhibited [5,15].

Despite good performance of the existing growth models, it is worth noting that the models can only accommodate a few number of features. Therefore, the is need to test the performance of these models in case of data with high dimensionality derived using automated phenotyping platforms, such as image derived data. The data from automated phenotyping platforms are quite noisy [16]. Moreover, this high dimensional data has many features, which can perhaps aid in coming up with a more actuality plant growth curve and thus improved prediction. Fitting such data set using the existing classical models, may lead to over fitted models thus affecting their prediction accuracy. Therefore, alongside the already existing models, there is also need to test how non-parametric models such as the XGBoost and light GBM would fit the high dimensional data. Since these non-parametric models had not been used in modelling plant growth [17], there is need to compare their prediction accuracies with the already existing plant growth models.

A study by [18] showed that non-parametric machine learning models handle more data than the other statistical models. This is an important strength that is really useful in handling large dimensional high-throughput data. Some of the most effective predictive models, such as XGBoost and Light GBM continue to be accurate with thousands or millions of additional data features [18,19,20]. Additionally, the non-parametric models have the mechanisms to sort out the variables that contain information relevant to the outcome and those variables that would just add noise to the predictions [21]. Generally, the conventional models that have been so far been applied for growth modeling don't have such inbuilt mechanisms. In extreme situations where the predictor variables are more than the observations, these models results to total failure [22]. In addition, for the existing growth models, the data must meet some underlying assumptions failure to which the prediction accuracy will be low [22].

The existing models needed to be tested alongside newly promising models such as extreme gradient boosting and light gradient boosting. The modelling of plant growth was a regression problem and thus the existing models were compared based on akaike information criterion (AIC), Bayesian information criterion (BIC), root mean squared error (RMSE), mean absolute error (MAE) and r-squared [23]. Smaller values of AIC, BIC, RMSE and MAE shows that a model is a better fit. Larger values of r-squared is an indication that a model is a better fit [23]. For instance, [24] compared the performance of non-linear mathematical models in describing growth. The criterions used were; the coefficient of determination, Akaike information criterion, root mean squared error and Bayesian information criterion. The models fitted were; Richards, Gompertz and logistic model for different animals [24]. The results showed that the Richards models provided a better fit to the experimental data. It was also noted that the different models showed better performance for some animals. In a

study that fitted growth models, the models were compared using the Akaike information criterion and the Bayesian information criterion [25]. The models fitted were; log logistic, weibull, exponential, logistic and Gompertz models. The best models were the logistic and the log logistic.

## 2. Methodology

The data was obtained from Leibniz Institute of Plant Genetic and Plant Crop Research (IPK-Gatersleben),

Gatersleben, Germany. All the analysis was carried out using the R statistical software. The features that used in model fitting were as described in [26]. The features were selected using feature importance statistical technique based on their predictability of the manually collected plant biomass (Table 1).

The selected features (Table 1) were used to fit the logistic curve, the Gompertz model, XGBoost and Light GBM models. A comparison was then made using RMSE, R-squared, AIC and BIC on how the non-parametric models and sigmoid models performed in modelling plant growth and biomass prediction.

**Table 1. Features used in fitting the statistical models**

Feature	Importance 0.05	Importance	Importance 0.95	Permutation error
volume.vis.area090t.norm.mm <sup>3</sup>	1.17192	1.21573	1.24578	4.13907
side.fluo.hull.area.norm	1.17158	1.18725	1.20006	4.04211
volume.vis.iap.norm.px <sup>3</sup>	1.11972	1.13237	1.14195	3.85524
side.fluo.border.length.norm	1.09477	1.11532	1.12343	3.79719
volume.fluo.prism.norm	1.57436	1.64844	1.67001	2.61137
volume.fluo.area090t.norm.mm <sup>3</sup>	1.26338	1.27438	1.28480	2.01880
top.leaf.length.sum.norm skeleton	1.16299	1.26243	1.27748	1.99988
volume.vis.prism.norm.mm <sup>3</sup>	1.21275	1.24389	1.29065	1.97051
top.fluo.area.norm.mm <sup>2</sup>	1.14904	1.16480	1.19361	1.84521
side.fluo.area.norm.mm <sup>2</sup>	1.10517	1.13703	1.14180	1.80123
side.vis.border.length.norm	1.10900	1.13590	1.14617	1.79944
top.fluo.hull.circumcircle.d.norm	1.09831	1.10230	1.10351	1.74621
top.vis.area.norm.mm <sup>2</sup>	1.06629	1.09965	1.10257	1.74201
top.vis.hull.pc2.norm	1.07377	1.08966	1.09113	1.72618
side.fluo.hull.area.norm	1.07207	1.08681	1.09896	1.72166
side.vis.area.norm.max.mm <sup>2</sup>	1.06926	1.07750	1.09406	1.70692
top.vis.hull.area.norm	1.04968	1.05987	1.06964	1.67898
side.height.norm mm	1.04635	1.05916	1.07647	1.67787
volume.fluo.lt.norm.mm <sup>3</sup>	1.05004	1.05562	1.06205	1.67225
top.vis.border.length.norm	1.04194	1.05015	1.07109	1.66358
volume.vis.lt.norm.mm <sup>3</sup>	1.04218	1.04425	1.05265	1.65425
side.vis.hull.pc2.norm	1.01240	1.02102	1.02611	1.61744
volume.vis.iap.norm.px <sup>3</sup>	1.01007	1.01954	1.03539	1.61510
side.leaf.length.sum.norm skeleton	1.01537	1.01930	1.02386	1.61473
side.leaf.width.average.norm skeleton	1.01123	1.01843	1.02451	1.61334
side.fluo.area.norm.max.mm <sup>2</sup>	1.01180	1.01613	1.01844	1.60970
side.vis.area.norm.mm <sup>2</sup>	1.01049	1.01308	1.01946	1.60487
top.leaf.width.average.norm skeleton	1.01138	1.01185	1.01577	1.60292
side.height.norm mm	1.00126	1.01110	1.01368	1.60173
volume.fluo.iap.norm.px <sup>3</sup>	1.00090	1.00833	1.01119	1.59735
volume.vis.iap_max.norm.px <sup>3</sup>	1.00179	1.00470	1.00675	1.59159
volume.vis.area090t.norm.mm <sup>2</sup>	0.99618	1.00390	1.01745	1.59032
top.vis.hull.pc1.norm	0.99807	1.00355	1.00825	1.58977
side.fluo.border.length.norm	1.00159	1.00322	1.00416	1.58925
side.fluo.hull.pc2.norm	1.00095	1.00131	1.00232	1.58623

Importance=score value of the feature in predictability of the manually collected biomass, Importance 0.05 & Importance 0.95=lower and upper limits of the confidence interval for the feature importance scores

Source: [26].

### 2.1. Logistic Growth Curve

Logistic growth curve is an S-shaped (sigmoidal) curve that can be used to model functions that increase gradually at first, more rapidly in the middle growth period, and slowly at the end, levelling off at a maximum value after some period of time [27]. Two models of the logistic growth curve were fitted in this study. The first logistic curve model had the equation;

$$y_1 = \frac{a_1}{(1 + \exp(b_1 \mp c_1 x))}$$

where  $y_1$  = volume,  $a_1$  = the asymptote (the maximum volume that could be attained),  $b_1$  = the displacement on the x-axis,  $c_1$  = the growth parameter (described how quickly the variable  $y_1$  approached the asymptote).

The second logistic curve model had the equation;

$$y_2 = \frac{a_2}{(1 + \exp(b_2 \mp c_2 x))}$$

where  $y_2$  = side area,  $a_2$  = the asymptote (the maximum side area that could be attained),  $b_2$  = the displacement on the x-axis,  $c_2$  = the growth parameter (described how quickly the variable  $y_2$  approached the asymptote).

### 2.2. Gompertz Growth Model

Gompertz model is a sigmoid model that is frequently used with growth data and other data. Two Gompertz models were fitted in this study. The first Gompertz model had the equation;

$$y_1 = a_1 e^{-b_1 e^{-c_1 x}}$$

where  $y_1$  = variable volume,  $x$  = time,  $a_1$  = the asymptote or the carrying capacity obtained by doing the limit of the function when volume tends to infinity (the highest volume that could be attained),  $b_1$  = the displacement on the x-axis,  $c_1$  = the growth rate.

The second Gompertz model had the equation;

$$y_2 = a_2 e^{-b_2 e^{-c_2 x}}$$

where  $y_2$  = side area,  $x$  = time,  $a_2$  = the asymptote or the carrying capacity obtained by doing the limit of the function when side area tends to infinity (the highest side area that could be attained),  $b_2$  = the displacement on the x-axis,  $c_2$  = the growth rate.

### 2.3. Extreme Gradient Boosting Model

Extreme Gradient Boosting (XGBoost) is a tree-based algorithm, which sits under the supervised branch of Machine Learning. In its development, new training data sets were formed by random sampling with replacement from the original dataset, during which some observations were repeated in each new training data set. The observations were weighted and therefore some of them were selected in the new datasets more often. The

XGBoost models fitted for this regression problem had the following general form.

$$y_i = f(x_i) + \epsilon$$

where  $y_i$  = variable being predicted at a specific time point (plant biomass),  $x_i$  = phenotypic features at the specific time point such as; side area, side length, leaf length, top area, side border length, top hull area, side height, volume, side leaf length, side leaf width and top leaf width (Table 1).

The training data was divided into subsets and the final prediction was a weighted sum of all the decision tree functions. In fitting the XGBoost models, samples were selected sequentially. For instance, the first sample was selected and a decision tree was fitted. The model then picked the examples that were hard to learn and using them and a few others selected at random from the training data set, a second model was fitted. Prediction was then made using the first and the second models. The model was then evaluated and hard examples were picked together with other randomly selected examples from training set and another model was fitted. The process of boosting algorithms continued up to a number  $n$ . In other words, the first model is fitted using the original training set. For example, a simple regression model;

$$y = f(x) + \epsilon,$$

if the error was too large, the solution could have been to add more features, use another algorithm, tune the algorithm or look for more training set. However, if the error was not white noise and it had a relationship with the output, the second model was fitted as  $\epsilon = f(x) + \epsilon_1$ . The process continued  $n$  –times and the final model was;

$$\epsilon_n = f_n(x) + \epsilon_{n-1}$$

The final step involved adding these models together with some weighing criteria where the weights =  $\alpha$ 's which yielded the final function that was used for prediction.

### 2.4. Light GBM Model

In Light GBM, the base learners were generated sequentially in such a way that the present base learner was always more effective than the previous one. This kind of boosting tried to optimize the loss function of the previous learners by a new adaptive model that added weak learners in order to reduce the loss function. The Light GBM models fitted for this regression problem had the following general form.

$$y_i = \beta x_i + \epsilon$$

where  $y_i$  = variable being predicted at a specific time point (plant biomass),  $x_i$  = phenotypic features at the specific time point such as; side area, side length, leaf length, top area, side border length, top hull area, side height, volume, side leaf length, side leaf width and top leaf width (Table 1). The fitting process of the Light GBM model was similar to the one outlined in section 2.3.

## 2.5 Models Validation and Comparison

The models were validated by using the cross validation statistical methods. In cross-validation, models are evaluated by training several models on subsets of the available input data and evaluating them on the complementary subset of the data. In this study 10-fold cross-validation method was used. In the 10-fold cross-validation, the input data was split into 10 subsets of data. The models were trained on all but one (10 - 1) of the subsets, and then evaluated on the subset that was not used for training. This process was repeated 10 times, with a different subset reserved for evaluation (and excluded from training) each time. The models were also compared statistically using the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), the root mean squared error (RMSE) and the mean absolute error (MAE). The values of the RMSE,  $R^2$  and MAE showed how the models performed in prediction of the manually collected biomass. The AIC determined the relative information value of the model using the maximum likelihood estimate and the number of parameters (independent variables) in the model. The formula for AIC is given by;

$$AIC = 2K - 2\ln L$$

where  $K$  is the number of features used and  $L$  is the log-likelihood estimate (the likelihood that the model could have produced the observed  $y$ -values. The smaller the AIC value meant that the better the model fit.

The BIC is a method for scoring and selecting a model. The BIC statistic was calculated as;

$$BIC = -2 \times LL + \ln(N) \times k$$

$LL$  is the log-likelihood of the model,  $N$  is the number of examples in the training dataset, and  $k$  is the number of parameters in the model. Based on BIC, more complex models had larger score and in turn, were less likely to be selected.

Root Mean Square Error (RMSE) was the standard deviation of the residuals. Residuals were a measure of how far from the regression line data points are. The RMSE was calculated as;

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

A low RMSE value indicated that the simulated and observed data were close to each other showing a better

accuracy. Thus lower the RMSE meant that the better the model performance.

Mean Absolute Error (MAE) measured the average magnitude of the errors in a set of predictions, without considering their direction. The MAE values were computed as;

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_j - \hat{y}_j|$$

A good MAE was relative to a specific dataset. It was a good idea to first establish a baseline MAE for a dataset using a naive predictive model. A model that achieved a MAE better than the MAE for the naive model had a better accuracy.

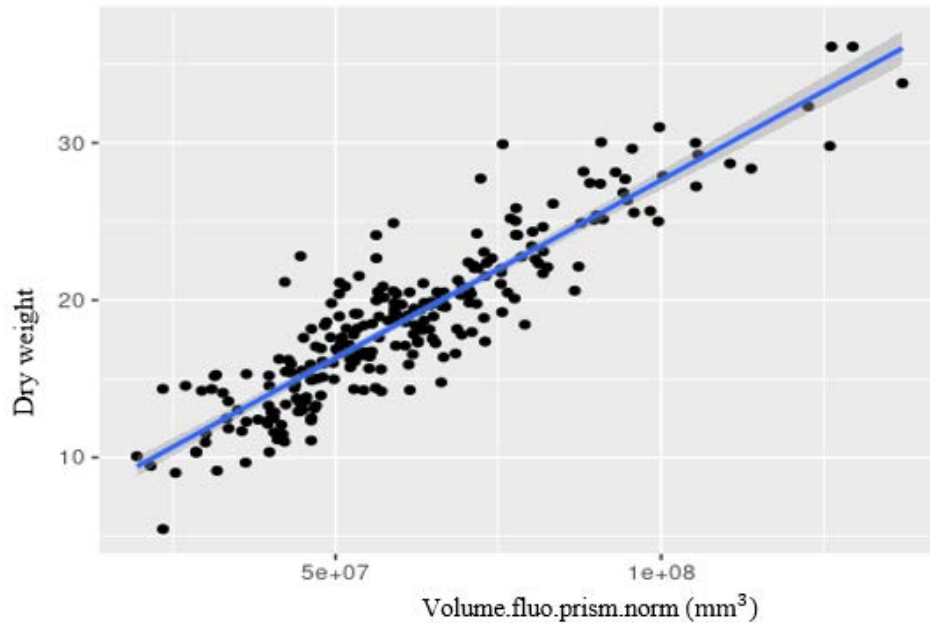
## 3. Results and Discussion

### 3.1. Preliminary Analysis

The preliminary analysis involved fitting a linear model between manually measured plant biomass (dry weight) and some selected phenotypic features such as volume, side height and side area. This was to establish if there was any relationship between the manually collected biomass and the plant phenotypic features from the image derived data. From the fitted model, the selected image features were significant predictors of plant biomass [ $p < 0.05$  Table 2]. The fitted models showed different strengths in predicting the plant biomass (Table 2). The model that was fitted using plant volume and side area showed the best results in terms of adjusted R-squared (Table 2). This showed that the manually collected plant biomass was correlated with the image derived phenotypic features. Additionally, there was also linear relationship between plant biomass and image derived phenotypic features such as plant volume. This suggested that plant biomass could be predicted using the features from the high throughput image derived data. The findings of this current study are in agreement with that of [28] who applied a logistic model for prediction of maize yield under water and nitrogen management and the results showed that the model predicted the maize yield during the growing season with an acceptable accuracy. The findings of this current study are also in agreement with those of [29] who showed that the logistic model is good at estimating the above-ground biomass from the plant height.

**Table 2. Linear Models fitted using the plant volume, side area and side height image derived phenotypic features**

Feature	Model	estimate	std.error	p - value	Multiple R-squared	Adjusted R-squared
Volume	Intercept	5.003e+00	4.430e-01	<2e-16	0.8127	0.8119
	volume.fluo.prism.norm	2.264e-07	6.915e-09	<2e-16		
Side area	Intercept	-2.654e+00	8.276e-01	0.00152	0.7382	0.7171
	side.vis.area.norm	4.710e-05	1.785e-06	< 2e-16		
Side height	Intercept	-0.0806252	1.1228351	0.943	0.5419	0.5400
	side.height.norm	0.0146723	0.0008584	<2e-16		



**Figure 1.** Diagrammatic representation of the linear relationship between manually measured plant dry weight biomass and image derived plant volume phenotypic feature

### 3.2. Fitted Extreme Gradient Boosting Models

The extreme gradient boosting models were fitted using image derived plant phenotypic features extracted from feature importance (Table 1) at various days after sowing (DAS). The values of the RMSE,  $R^2$  and MAE showed how the models performed in prediction of plant biomass from image derived phenotypic features (Table 3). On the basis of the RMSE, the best model was the model that attained the least value of the RMSE [(2.1641) Table 3]. The model was with the least value of RMSE (model at 36 DAS) also had the highest value of  $R^2$  (0.8292) (Table 3). The  $R^2$  value (0.8292) of the model showed that 82.92% of the variation in plant biomass was accounted for by the used selected

image derived phenotypic features. On the basis of the mean absolute error (MAE) metric, the model that had the best performance was the model outputted the least value of the MAE (Table 3). These findings are consistent with those of [23], who evaluated the fitted XGBoost model using R-squared, RMSE, and MAE performance metrics. In their analysis, the fitted XGBoost model delivered the best results [23]. Using the R-squared, RMSE, and MAE performance criteria, it was also discovered that the XGBoost model outperformed Artificial Neural Networks and Support Vector Regression [30]. [31] compared performance linear model, random forest, support vector regression, XGBoost, LASSO regression, and ensemble technique using R-squared, RMSE, and MAE performance measures, with the XGBoost model yielding the best results.

**Table 3. Gradient Boosting Models for plant phenotypic Features derived using Feature Importance**

day	RMSE	Rsquared	MAE	RMSESD	RsquaredSD	MAESD
11	4.6884	0.1460	3.6923	0.3791	0.0608	0.3203
13	4.6813	0.1457	3.6577	0.3153	0.0573	0.2686
15	4.2397	0.2945	3.2349	0.2378	0.0559	0.2150
18	4.2117	0.3072	3.3200	0.2814	0.0661	0.2349
20	4.1556	0.3207	3.3552	0.1841	0.0594	0.1529
22	3.0956	0.6249	2.4436	0.3017	0.0873	0.1802
24	2.8594	0.6773	2.2935	0.1550	0.0511	0.1292
26	2.7723	0.6952	2.2064	0.1982	0.0679	0.1609
28	2.5828	0.7375	2.0788	0.1691	0.0455	0.1526
30	2.4349	0.7634	1.9143	0.1965	0.0486	0.1596
32	2.2838	0.7918	1.7914	0.1721	0.0380	0.1471
34	2.4331	0.7674	1.9183	0.1973	0.0355	0.1855
36	2.1641	0.8292	1.7254	0.1345	0.0348	0.1298
38	2.1679	0.8104	1.7356	0.1459	0.0331	0.1267
40	2.2867	0.7907	1.7615	0.1075	0.0401	0.0740
42	2.3691	0.7747	1.8284	0.1817	0.0571	0.1455

RMSE-Root Mean Square Error, RMSESD- Standard Deviation of RMSE, MAE- Mean Absolute Error, MAESD- Standard Deviation of MAE.

### 3.3. Fitted Light Gradient Boosting Models

The light gradient boosting models were fitted using the image derived plant phenotypic features extracted from feature importance (Table 1). From the results the best model on the basis of the RMSE was the model that attained the least value of the RMSE [(2.767341) Table 4]. The values of the RMSE showed how the models performed in prediction of plant biomass from image derived phenotypic features. On the basis of the MAE metric, the model that had the best performance was the model that outputted the least value of the MAE [(MAE=2.012106) Table 4]. This findings are similar to those of [32] Kopitar *et al.* (2020) who showed that Light GBM outperformed random forest and generalized linear models when their performance was evaluated using RMSE and MAE performance metrics. In a similar study, RMSE and MAE performance criteria showed that the light GBM outperformed the adaBoost model for regression [33]. A study by [34] showed that the light GBM outperformed neural networks and decision trees when the performance was evaluated using RMSE and MAE performance criteria.

**Table 4. Light Gradient Boosting Model for plant phenotypic features selected using Feature Importance**

RMSE	MAE	day
5.005373	3.869566	11
4.918358	3.815482	13
4.406645	3.393446	15
4.612141	3.558899	18
4.415014	3.452459	20
3.685053	2.824935	22
3.482178	2.646189	24
3.302421	2.503556	26
3.724924	2.806156	28
2.767341	2.012106	30
2.898839	2.116751	32
2.777760	2.034543	34
3.122751	2.287803	36
3.696688	2.752755	38
3.589284	2.617610	40
3.138000	2.280933	42

RMSE-Root Mean Square Error, MAE- Mean Absolute Error.

### 3.4. Fitted Gompertz Models

The gompertz models were fitted using the image derived plant volume and side area phenotypic features. These are the features that attained the best feature importance score (Table 1). The Gompertz model fitted using the plant volume phenotypic feature had parameters estimates of  $3.520762e+08$  ( $8.305924e+07$ ),  $1.055582e+02$  ( $2.897641e+01$ ) and  $3.348069e-01$  ( $3.620450e-02$ ) (Table 5). The values in parenthesis represents the standard errors of the estimates. These parameters were significant at 5% level of significance (Table 5). In the model output, the parameters were represented by symbols Asym, b2 and b3. The symbol Asym represented the asymptote or the carrying capacity obtained by doing the limit of the function when volume tends to infinity. This simply showed the highest plant volume that could be attained. The symbol b2 represented the displacement on the x-axis. The symbol b3 was the growth rate. The log likelihood, the AIC, BIC and the deviance statistic values for this model are: -69865.1, 139738.3, 139763.4 and  $3.97917e+17$ , respectively (Table 6). The results showed that the log-likelihood value for the model can range from negative infinity to positive infinity [35]. The log-likelihood, AIC and BIC values for a given model are mostly meaningless, but they are useful for comparing two or more models [35].

The Gompertz model fitted using image derived plant side area phenotypic feature had parameters estimates of  $7.882705e+06$  ( $2.460224e+06$ ),  $2.747763e+01$  ( $2.719820e+00$ ) and  $5.416621e-01$  ( $3.035230e-02$ ) and its parameters were significant at 5% level of significance (Table 7). The values in parenthesis represents the standard errors of the estimates. The estimates represented the maximum volume that could be attained using the model, the displacement on the x-axis and the growth rate respectively. The log likelihood, the AIC and BIC values for this model are: -49214.0, 98436.15, 98461.31 and  $1.251596e+13$ , respectively (Table 8). The log likelihood, the AIC and BIC values were used for models comparison. These values were very useful in comparing the different models. The Gompertz growth curves showed that growth is slowest at the start and end of a given time period (Figure 2 & Figure 3).

**Table 5. Fitted Gompertz Model using volume plant phenotypic feature**

term	estimate	std.error	statistic	p.value
Asym	$3.520762e+08$	$8.305924e+07$	4.238857	0.000023
b2	$1.055582e+02$	$2.897641e+01$	3.642900	0.000273
b3	$3.348069e-01$	$3.620450e-02$	9.247659	0.000000

**Table 6. Performance metric for the Gompertz model using volume plant phenotypic feature**

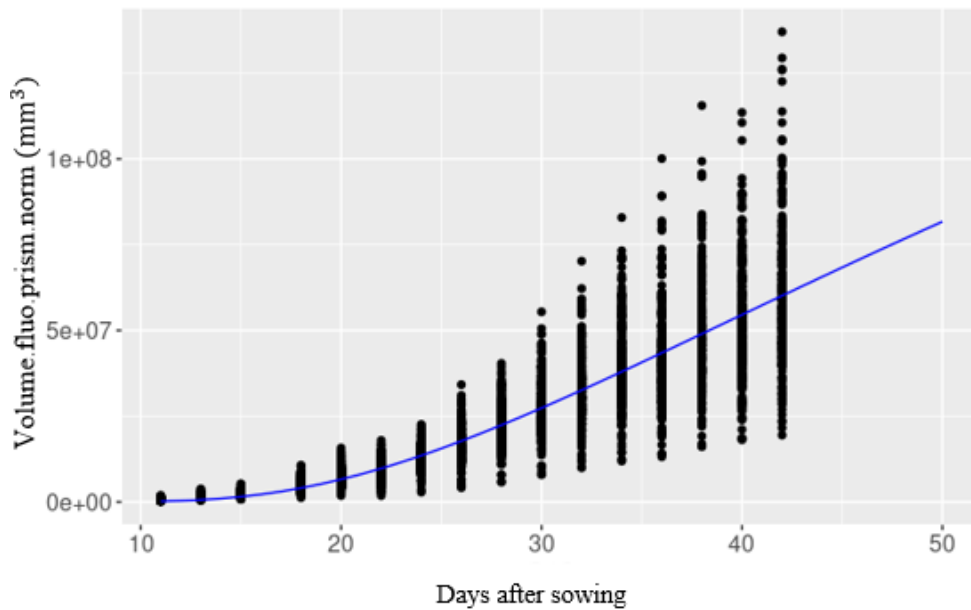
sigma	logLik	AIC	BIC	deviance	df.residual	nobs
9997702	-69865.1	139738.3	139763.4	$3.97917e+17$	3981	3984

**Table 7. Fitted Gompertz Model using the Feature using side area plant phenotypic feature**

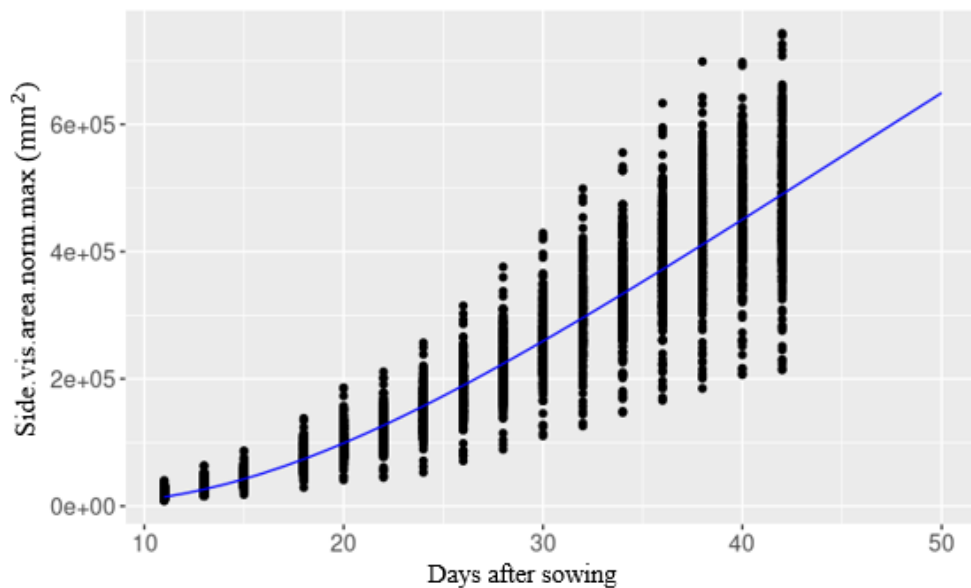
term	estimate	std.error	statistic	p.value
Asym	$7.882705e+06$	$2.460224e+06$	3.20406	0.0013657
b2	$2.747763e+01$	$2.719820e+00$	10.10274	0.0000000
b3	$5.416621e-01$	$3.035230e-02$	17.84582	0.0000000

**Table 8. Performance Metrics for the Gompertz model using side area plant phenotypic feature**

sigma	finTol	logLik	AIC	BIC	deviance	df.residual	Nobs
56070.71	3e-07	-49214.0	98436.15	98461.31	1.251596e+13	3981	3984



**Figure 2.** Gompertz growth curve generated using image derived plant volume phenotypic feature



**Figure 3.** Gompertz growth curve generated using image derived plant side area phenotypic feature

The results showed that plant biomass accumulation in maize followed a sigmoidal growth. The increase is from a fixed point origin (germination point). As the plant develop and produces additional leaves and thus enabling the growth of either more or larger leaves, this creates an acceleration portion of the growth curve. Later in the plant development stage as the plant switch to the reproductive phase, the accumulation of additional plant biomass slows down which ultimately give an S-shaped curve. The dataset for this study did not extend into reproductive phase and thus captured only the first phase of the sigmoidal biomass accumulation plants pattern, hence producing J-shaped curves. This study findings is in agreement with that of [36] who investigated a simple

mathematical model that described the growth of the area and the number of total and viable cells in yeast colonies. The study showed that with low inocula, viable cells showed an initial short exponential phase when the colonies were not visible. This phase was shortened with higher inocula. In visible or mature colonies, cell growth displayed Gompertz-type kinetics. It was concluded that the cells growth in colonies was similar to liquid cultures only during the first hours, the rest of the time they grow, with near-zero specific growth rates, at least for 3 weeks [36]. [37] studied maize dry matter production and macronutrient extraction model as a new approach for fertilizer rate estimation. The study showed that dry matter accumulation followed a sigmoidal model and the



macronutrient content a power model. Biomass accumulation and partitioning in maize has also been found to follow a sigmoidal curve with an exponential phase followed by a linear ending with the senescence phase [38].

[39] did comparison of nonlinear models to describe the feather growth and development curve in yellow-feathered chickens. The study showed that the inflection point for the whole-body feather mass occurred, which was interpreted as the age of maximum feather mass gain as indicated by the change in the increase in feather mass from fast to slow [38]. The findings of this current study

are in agreement with that of [40] who applied Gompertz model to describe the growth of corn. The study showed a slow growth rate at the initial stage, followed by a rapid growth stage to a critical point then the rate of growth began to decline reaching to a stability phase. The findings of this current study are also in agreement with those of [41] who applied Gompertz curve to the growth of tobacco leaves and stem. The study showed that the growth of tobacco leaves and stem was slowest at the start and end of a given time period.

### 3.5. Fitted Logistic Curve Models

**Table 9. Fitted logistic growth curve using volume plant phenotypic feature**

term	estimate	std.error	statistic	p.value
Asym	7.049763e+07	1.642039e+06	42.93298	0
xmid	3.281341e+01	3.418623e-01	95.98429	0
scal	5.805428e+00	1.632803e-01	35.55499	0

**Table 10. Performance metrics for the logistic growth curve using volume plant phenotypic feature**

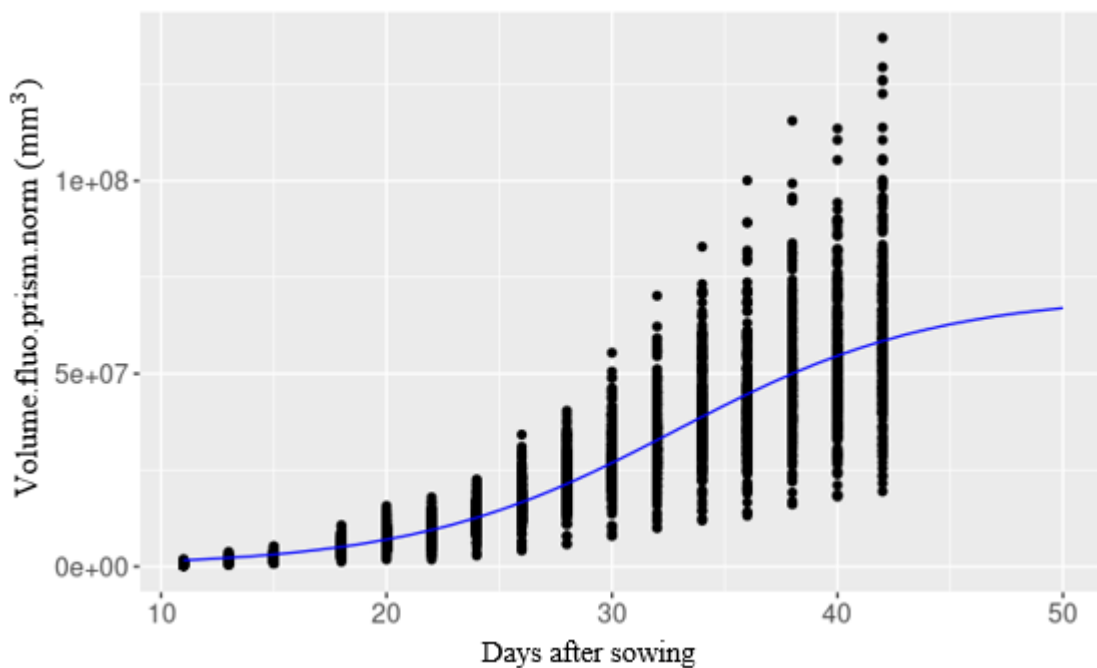
sigma	finTol	logLik	AIC	BIC	deviance	df.residual	nobs
10011471	4e-07	-69870.61	139749.2	139774.4	3.990138e+17	3981	3984

**Table 11. Fitted logistic growth curve using side area plant phenotypic feature**

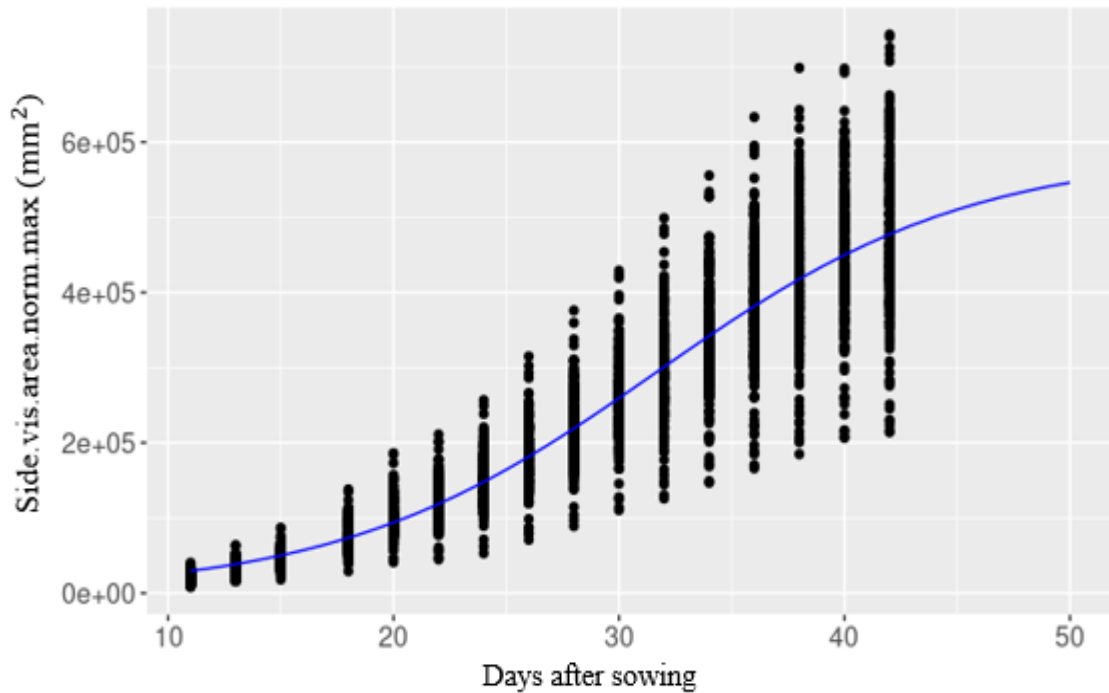
term	estimate	std.error	statistic	p.value
Asym	585775.55336	9784.1058903	59.87012	0
xmid	31.61234	0.2885569	109.55324	0
scal	7.01180	0.1357929	51.63598	0

**Table 12. Performance metrics for the logistic growth curve fitted using side area plant phenotypic feature**

sigma	finTol	logLik	AIC	BIC	deviance	df.residual	nobs
55928.75	1e-07	-49203.98	98415.95	98441.11	1.245267e+13	3981	3984



**Figure 4.** Logistic growth curve fitted using image derived plant volume phenotypic feature



**Figure 5.** Logistic growth curve fitted using image derived plant side area phenotypic feature

The logistic growth models were fitted using the image derived plant volume and side area phenotypic features. These are the features that attained the best feature importance scores. The model fitted using image derived plant volume phenotypic feature had parameters estimates of  $7.049763e+07$  ( $1.642039e+06$ ),  $3.281341e+01$  ( $3.418623e-01$ ) and  $5.805428e+00$  ( $1.632803e-01$ ) and its parameters were significant at 5% level of significance (Table 9). The values in parenthesis represents the standard errors of the estimates. The estimates represented the maximum volume that could be attained using the model, the displacement on the x-axis and the growth rate respectively. The log likelihood, the AIC and BIC values for this model are:  $-69870.61$ ,  $139749.2$ ,  $139774.4$  and  $3.990138e+17$ , respectively (Table 10). The log likelihood, the AIC and BIC values for the logistic curve model fitted using image derived plant volume phenotypic feature were used for models comparison. The model fitted using image derived plant side area phenotypic feature had parameters  $585775.55336$  ( $9784.1058903$ ),  $31.61234$  ( $0.2885569$ ) and  $7.01180$  ( $0.1357929$ ) and its parameters were significant at 5% level of significance (Table 11). The values in parenthesis represents the standard errors of the estimates. The estimates represented the maximum plant volume that could be attained using the model, the displacement on the x-axis and the growth rate respectively. The log likelihood, the AIC, BIC and the deviance statistic values for this model are:  $-49203.98$ ,  $98415.95$ ,  $98441.11$  and  $1.245267e+13$ , respectively (Table 12). The log likelihood, the AIC and BIC values for the logistic curve model fitted using image derived plant side area were used for models comparison. The logistic growth curves for models fitted using volume and side area showed that growth is slowest at the start and end of a given time period (Figure 4 & Figure 5).

The findings demonstrated that maize plant biomass

accumulation follows a sigmoidal growth pattern. These findings are in agreement with those of [42] who using logistic regression model showed that the plant growth is slowest at the start and end of a given time period. A study by [43] that investigated patterns of dry biomass accumulation and nutrient uptake by okra (*Abelmoschus esculentus* L.) under different rates of nitrogen application found out that growth followed a sigmoid curve which was accurately described a logistic equation.

### 3.6. Comparison of the Statistical Power of the Models for Plant Growth

The comparison of the statistical power of the models that were used for plant growth, involved comparison of the prediction power of the classical models against the machine learning models. Further, the two machine learning models were also compared to find out, which had a better prediction power. The comparison metrics that were used in this case were the root mean squared error, R-squared, mean absolute error, Akaike information criterion and Bayesian information criterion. Root mean squared error estimated the accuracy of the forecasting models predicted values versus the actual or observed values while training the regression models. It was used to measure the error in the predicted values since the target or response variable was a continuous number. This criterion was essential in shortlisting the best performing model among different forecasting models that were trained on one particular dataset. The comparison was simply done by comparing the RMSE values across all models and select the one with the lowest value on RMSE (Table 13). The results in Table 13 showed that the extreme gradient boosting model performed better than the light gradient boosting (smaller values of RMSE and MAE in XGBoost than in Light GBM).

**Table 13. Comparison of Statistical Power for Machine Learning Models for modelling plant growth**

day	XGBOOST		LIGHT GBM	
	RMSE	MAE	RMSE	MAE
11	4.6884	3.6923	5.005373	3.869566
13	4.6813	3.6577	4.918358	3.815482
15	4.2397	3.2349	4.406645	3.393446
18	4.2117	3.3200	4.612141	3.558899
20	4.1556	3.3552	4.415014	3.452459
22	3.0956	2.4436	3.685053	2.824935
24	2.8594	2.2935	3.482178	2.646189
26	2.7723	2.2064	3.302421	2.503556
28	2.5828	2.0788	3.724924	2.806156
30	2.4349	1.9143	2.767341	2.012106
32	2.2838	1.7914	2.898839	2.116751
34	2.4331	1.9183	2.777760	2.034543
36	2.1641	1.7254	3.122751	2.287803
38	2.1679	1.7356	3.696688	2.752755
40	2.2867	1.7615	3.589284	2.617610
42	2.3691	1.8284	3.138000	2.280933

On comparison of the Gompertz growth model and the logistic curve model, the two models showed almost the same values of Akaike information criterion and Bayesian information criterion (Table 14). However, these models have a limitation in that they can model only a single phenotypic feature at a time with their independent variable always being time (DAS). A study by [44] that was in line with this study found out that the Gompertz model is the simplest model in the estimation process, while the logistic model is more difficult in the computing process. The Gompertz and logistic models used in the study by [44] had high degree of accuracy with the value of the coefficient of determination of more than 90%. Other Similar findings showed that Akaike's information criteria (AIC) and Bayesian information criterion (BIC) have been used in comparing several growth models [45]. The findings of this current study are in line with those of [45] who showed that Gompertz model and the logistic function yielded a better fit since their values of AIC and BIC were low and that there was no an autocorrelation between the residual values. [40] also showed that a simple sigmoid model is preferred as it is easier to interpret the parameters biologically. The logistic model fitted better in describing the plant height growth compared to Gompertz model, as it yielded coefficient of determination more than 99% [40]. Based on the model, the absolute growth rate tended to be bell-shaped and right-skewed for logistic and Gompertz, respectively [40].

Statistically it was not possible to directly compare the machine learning models and the classical growth models [46]. This is due to the fact that, the growth models can only accommodate one feature at a time with the independent variable always being time intervals [46]. However, in this study, it was shown that having a model that fits just one feature is inadequate to predict plant growth because some information is left out. All the features that had been extracted at feature selection stage could all not be accommodated in the Gompertz or the

logistic function model. To show that there was loss of information which in turn led to a reduced statistical power, an XGBoost model with all the best selected 31 image derived phenotypic features was compared with the XGBoost model that had the only one feature that had used in the developing the sigmoid models (Table 15). The model with one feature showed poorer performance in terms of RMSE, MAE and R-squared as compared to the model with the 31 features. This is a clear indication that that the XGBoost and Light GBM models are better for modelling plant growth as compared to the mostly used sigmoid models. In agreement with this findings is a study by [47] who compared a Gompertz model against a machine learning model with the latter attaining superior results. [48] also showed that the machine learning models such as neural networks and XGBoost produced better performance in terms of prediction accuracy when compared with the sigmoid models such as the logistic model and the Gompertz model.

**Table 14. Models Comparison for the Gompertz and Logistic Curve Models**

	Gompertz Models		Logistic Curves	
	Model 1	Model 2	Model 1	Model 2
sigma	9997702	56070.71	10011471	55928.75
isConv	TRUE	TRUE	TRUE	TRUE
finTol	2E-07	3E-07	4.00E-07	1.00E-07
logLik	-69865.1	-49214.1	-69870.6	-49204
AIC	139738.3	98436.15	139749.2	98415.95
BIC	139763.4	98461.31	139774.4	98441.11
deviance	3.98E+17	1.25E+13	3.99E+17	1.25E+13
df.residual	3981	3981	3981	3981
nobs	3984	3984	3984	3984

**Table 15. Models Comparison for XGBoost with 31 features and XGBoost with 1 feature**

Parameter	Model with 31 Features	Model with 1 Feature
RMSE	2.1641	3.8378
Rsquared	0.8292	0.7644
MAE	1.7254	2.019
RMSESD	0.1345	0.2659
RsquaredSD	0.0348	0.0434
MAESD	0.1298	0.1576

On comparison of the machine learning models, the extreme gradient boosting model showed better performance than the light gradient boosting model (Table 13). These results agree with research findings that have shown that Light GBM is faster and more accurate than Cat Boost and XG Boost using variant number of features and records [49]. [50] also found out that LightGBM can significantly outperforms XGBoost and Semi Global Matching with Neural networks in terms of computational speed and memory consumption. In a study comparing Gradient Boosting Decision Tree Algorithms for prediction Performance, LightGBM algorithm had the best performance of the three with a balanced combination of accuracy, speed, reliability, and ease of use, followed by XGBoost with the histogram method, and CatBoost came last with slow and inconsistent performance [51].

Another study that concurs with this study found out that non-parametric models offer more accurate predictions since they offer a better fit to data than parametric ones [52]. Further, non-parametric algorithms provide a good fit for data since they can fit many forms of a function [52].

## 4. Conclusion

In conclusion, this study found out that the machine learning models performed better than the conventional sigmoid models in explaining plant growth. The extreme gradient boosting model and the light gradient boosting models had the advantage of modelling several phenotypic features at once as opposed to the classical models that could model only one phenotypic feature at a go. Further, the results also showed that extreme gradient boosting model showed better performance when modelling plant growth as compared to the light gradient boosting model. The developed models could also be used to predict plant biomass at different developmental stages.

## Acknowledgments

The authors acknowledge the Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany for making available the data used in this study. The data was obtained with the support of grants from the German Federal Ministry of Education and Research (BMBF) and performed within the ConFed projects (identification number: 0315461C). The authors declare no conflicts of interest.

## References

- [1] Fourcaud, T., Zhang, X., Stokes, A., Lambers, H., & Körner, C. (2008). Plant growth modelling and applications: the increasing importance of plant architecture in growth models. *Annals of Botany*, 101(8), 1053-1063.
- [2] Vos, J., Evers, J. B., Buck-Sorlin, G. H., Andrieu, B., Chelle, M., & De Visser, P. H. (2009). Functional-structural plant modelling: a new versatile tool in crop science. *Journal of experimental Botany*, 61(8), 2101-2115.
- [3] Rahaman, M., Chen, D., Gillani, Z., Klukas, C., & Chen, M. (2015). Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Frontiers in plant science*, 6, 619.
- [4] Dold, Christian. (2018). Climate Change Impacts on Corn Phenology and Productivity. 10.5772/intechopen.76933.
- [5] Chen, D., Neumann, K., Friedel, S., Kilian, B., Chen, M., Altmann, T., & Klukas, C. (2014). Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. *The Plant Cell*, 26(12), 4636-4655.
- [6] Klukas, C., Chen, D., & Pape, J. M. (2014). Integrated analysis platform: an open-source information system for high-throughput plant phenotyping. *Plant physiology*, 165(2), 506-518.
- [7] Junker A, Muraya MM, Weigelt-Fischer K, Arana-Ceballos F, Klukas C, Melchinger AE, Meyer RC, Riewe D, Altmann T (2015) Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems. *Frontiers in Plant Sciences*. 5(770): 1-21.
- [8] Muraya MM, Chu J, Zhao Y, Junker A, Klukas C, Reif JC, Altmann T (2017) Genetic variation of growth dynamics in maize (*Zea mays* L.) revealed through automated non-invasive phenotyping. *The Plant Journal*, 89: 366-380.
- [9] Boyd, D., & Crawford, K. (2011). Six provocations for big data. In *A decade in internet time: Symposium on the dynamics of the internet and society*.
- [10] Unnisabegum, A., Hussain, M., & Shaik, M. (2019). Data Mining Techniques for Big Data, Vol. 6, Special Issue. 10.13140/RG.2.2.25408.07686.
- [11] Blum, A., Hopcroft, J., & Kannan, R. (2020). *Foundations of data science*. Cambridge University Press.
- [12] Neumann, K., Klukas, C., Friedel, S., Rischbeck, P., Chen, D., Entzian, A., & Kilian, B. (2015). Dissecting spatiotemporal biomass accumulation in barley under different water regimes using high-throughput image analysis. *Plant, cell & environment*, 38(10), 1980-1996.
- [13] Karadavut, U., Palta, C., Kokten, K., & Bakoglu, A. (2010). Comparative study on some non-linear growth models describing leaf growth of maize. *Int. J. Agric. Biol.* 12, 227-230.
- [14] Golzarian, M. R., Frick, R. A., Rajendran, K., Berger, B., Roy, S., Tester, M., & Lun, D. S. (2011). Accurate inference of shoot biomass from high-throughput images of cereal plants. *Plant methods*, 7(1), 1-11.
- [15] Flombaum, P., & Sala, O. E. (2007). A non-destructive and rapid method to estimate biomass and aboveground net primary production in arid environments. *Journal of Arid Environments*, 69(2), 352-358.
- [16] Rani, Sushma & Anurag, & Rao, Srinivas. (2019). Study and Analysis of Noise Effect on Big Data Analytics. 8. 5841-5850.
- [17] Xu, Y., Qiu, Y., & Schnable, J. C. (2018). Functional modeling of plant growth dynamics. *The Plant Phenome Journal*, 1(1), 1-10.
- [18] Bratsas, C., Koupidis, K., Salanova, J. M., Giannakopoulos, K., Kaloudis, A., & Aifadopoulou, G. (2020). A Comparison of Machine Learning Methods for the Prediction of Traffic Speed in Urban Places. *Sustainability*, 12(1), 142.
- [19] Kong, Y., & Yu, T. (2018). A deep neural network model using random forest to extract feature representation for gene expression data classification. *Scientific reports*, 8(1), 1-9.
- [20] Putra, F. H. R., Larkin, D., Khanagha, S., & Panza, K. (2019). Paper to be presented at DRUID19 Copenhagen Business School, Copenhagen, Denmark June 19-21, 2019.
- [21] Zhang, Y. F., Fitch, P., & Thorburn, P. J. (2020). Predicting the Trend of Dissolved Oxygen Based on the kPCA-RNN Model. *Water*, 12(2), 585.
- [22] Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European journal of epidemiology*, 31(4), 337-350.
- [23] Shahani, N. M., Zheng, X., Liu, C., Hassan, F. U., & Li, P. (2021). Developing an XGBoost Regression Model for Predicting Young's Modulus of Intact Sedimentary Rocks for the Stability of Surface and Subsurface Structures. *Front. Earth Sci*, 9, 761990.
- [24] Teleken, J. T., Galvão, A. C., & Robazza, W. D. S. (2017). Comparing non-linear mathematical models to describe growth of different animals. *Acta Scientiarum. Animal Sciences*, 39, 73-81.
- [25] Sprouffske, K., & Wagner, A. (2016). Growth curve: an R package for obtaining interpretable metrics from microbial growth curves. *BMC bioinformatics*, 17(1), 172.
- [26] Gachoki, Peter, Moses Muraya, and Gladys Njoroge. "Features Selection in Statistical Classification of High Dimensional Image Derived Maize (*Zea Mays* L.) Phenomic Data." *American Journal of Applied Mathematics and Statistics* 10, no. 2 (2022): 44-51.
- [27] Birmingham, J. R. (2003). On exponential growth and mathematical purity: a reply to Bartlett. *Population and Environment*, 25(1), 71-73.
- [28] Sepaskhah, A. R., Fahandezh-Saadi, S., & Zand-Parsa, S. (2011). Logistic model application for prediction of maize yield under water and nitrogen management. *Agricultural Water Management*, 99(1), 51-57.
- [29] Xiangxiang, W., Quanjiu, W., Jun, F., Lijun, S., & Xinlei, S. (2014). Logistic model analysis of winter wheat growth on China's Loess Plateau. *Canadian Journal of Plant Science*, 94(8), 1471-1479.
- [30] Osman, A. I. A., Ahmed, A. N., Chow, M. F., Huang, Y. F., & El-Shafie, A. (2021). Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal*, 12(2), 1545-1556.

- [31] Huang, J. C., Tsai, Y. C., Wu, P. Y., Lien, Y. H., Chien, C. Y., Kuo, C. F., ... & Kuo, C. H. (2020). Predictive modeling of blood pressure during hemodialysis: A comparison of linear model, random forest, support vector regression, XGBoost, LASSO regression and ensemble method. *Computer methods and programs in biomedicine*, 195, 105536.
- [32] Kopitar, L., Kocbek, P., Cilar, L., Sheikh, A., & Stiglic, G. (2020). Early detection of type 2 diabetes mellitus using machine learning-based prediction models. *Scientific reports*, 10(1), 1-12.
- [33] Liang, Y., Wu, J., Wang, W., Cao, Y., Zhong, B., Chen, Z., & Li, Z. (2019, August). Product marketing prediction based on XGboost and LightGBM algorithm. In *Proceedings of the 2nd International Conference on Artificial Intelligence and Pattern Recognition* (pp. 150-153).
- [34] Pu, Q., Li, Y., Zhang, H., Yao, H., Zhang, B., Hou, B., & Zhao, L. (2019). Screen efficiency comparisons of decision tree and neural network algorithms in machine learning assisted drug design. *Science China Chemistry*, 62(4), 506-514.
- [35] Abdulhafedh, A. (2017). Incorporating the multinomial logistic regression in vehicle crash severity modeling: A detailed overview. *Journal of Transportation Technologies*, 7(03), 279.
- [36] Rivas, E. M., Gil de Prado, E., Wrent, P., de Silóniz, M. I., Barreiro, P., Correa, E. C., & Peinado, J. M. (2014). A simple mathematical model that describes the growth of the area and the number of total and viable cells in yeast colonies. *Letters in applied microbiology*, 59(6), 594-603.
- [37] Martins, K. V., Dourado-Neto, D., Reichardt, K., Favarin, J. L., Sartori, F. F., Felisberto, G., & Mello, S. C. (2017). Maize dry matter production and macronutrient extraction model as a new approach for fertilizer rate estimation. *Anais da Academia Brasileira de Ciências*, 89, 705-716.
- [38] Connor, D. J., Loomis, R. S., & Cassman, K. G. (2011). *Crop ecology: productivity and management in agricultural systems*. Cambridge University Press.
- [39] Xie, W. Y., Pan, N. X., Zeng, H. R., Yan, H. C., Wang, X. Q., & Gao, C. Q. (2020). Comparison of nonlinear models to describe the feather growth and development curve in yellow-feathered chickens. *Animal*, 14(5), 1005-1013.
- [40] Wardhani, W. S., & Kusumastuti, P. (2014). Describing the height growth of corn using Logistic and Gompertz model. *AGRIVITA, Journal of Agricultural Science*, 35(3), 237-241.
- [41] Ismail, Z., Khamis, A., & Jaafar, M. Y. (2003). Fitting nonlinear Gompertz curve to tobacco growth data. *Pakistan Journal of Agronomy*, 2(4), 223-236.
- [42] Tessmer, O.L., Jiao, Y., Cruz, J.A., Kramer, D.M., & Chen, J. (2013). Functional approach to high-throughput put plant growth analysis. *BMC Syst. Biol.* 7 (Suppl.6): S17.
- [43] Moustakas, N. K., Akoumianakis, K. A., & Passam, H. C. (2011). Patterns of dry biomass accumulation and nutrient uptake by okra (*abelmoschus esculentus*(L.) Moench.) Under different rates of nitrogen application. *Australian Journal of Crop Science*, 5(8), 993-1000.
- [44] Noor, R. R., Saefuddin, A., & Talib, C. (2012). Comparison on accuracy of Logistic, Gompertz and Von Bertalanffy models in predicting growth of new born calf until first mating of Holstein Friesian heifers. *Journal of Indonesian Tropical Animal Agriculture*, 37(3), 151-160.
- [45] Sariyel, V., Aygun, A., & Keskin, I. (2017). Comparison of growth curve models in partridge. *Poultry Science*, 96(6), 1635-1640.
- [46] Levy, J., Mussack, D., Brunner, M., Keller, U., Cardoso-Leite, P., & Fischbach, A. (2020). Contrasting classical and machine learning approaches in the estimation of value-added scores in large-scale educational data. *Frontiers in psychology*, 2190.
- [47] Hansen, B. D., Tamouk, J., Tidmarsh, C. A., Johansen, R., Moeslund, T. B., & Jensen, D. G. (2020, July). Prediction of the Methane Production in Biogas Plants Using a Combined Gompertz and Machine Learning Model. In *International Conference on Computational Science and Its Applications* (pp. 734-745). Springer, Cham.
- [48] Ravnik, J., Jovanovac, J., Trupej, A., Vištica, N., & Hriberšek, M. (2021). A sigmoid regression and artificial neural network models for day-ahead natural gas usage forecasting. *Cleaner and Responsible Consumption*, 3, 100040.
- [49] AlDaoud, E. (2019). Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset. *International Journal of Computer and Information Engineering*, 13(1), 6-10.
- [50] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- [51] Alshari, Haithm & Saleh, Yahya & Odabas, Alper. (2021). Comparison of Gradient Boosting Decision Tree Algorithms for CPU Performance. *Erciyes Tip Dergisi*. 157-168.
- [52] Marques, Armando & Prates, Pedro & Pereira, André & Oliveira, M. & Fernandes, José & Ribeiro, Bernardete. (2020). Performance Comparison of Parametric and Non-Parametric Regression Models for Uncertainty Analysis of Sheet Metal Forming Processes. *Metals*.

