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. Moreover, information on usefulness of these features and their selection criteria is limited. 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 models, 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 image data was imputed using *k* Nearest Neighbours technique. The feature importance, Shapley values and LASSO regression were used to extract the features that were used to fit the models. The Shapley values extracted 25 phenotypic features, feature importance extracted 31 features and LASSO regression extracted 12 features. Of the three techniques, the feature importance technique emerged the best feature selection technique since its features produced the best performing XGBoost with RMSE and R-squared values of 2.1641 and 0.8292, respectively. The suitability of the RMSE and the R-squared was because that the study was a regression problem where the aforementioned tools are used as 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 feature importance for feature selection, whenever high dimensional and complex phenotypic data is involved. More over, the study also recommends the use of XGBoost as a generic model to fit high dimensional and complex phenotypic data in modelling plant growth and to predict plant biomass yield at different growth points.