A Machine Learning-Based Multi-Temporal, Multi-Source Approach for Accurate Grape Yield Estimation

Aron WALSH, Kyle LOGGENBERG, and Albert STREVER, South Africa

Keywords: precision viticulture, yield prediction, Sentinel-2, vegetation indices, remote sensing, machine learning, Random Forest, XGBoost

SUMMARY

Viticulture yield estimation is essential to manage increasing climate-driven variability, enabling growers to optimise production, maintain fruit quality, and reduce economic risk. This study develops a multi-source, multi-temporal remote sensing machine learning framework for grape yield prediction across different spatial scales and vineyard environments. This research investigated five growing seasons (2021-2025) across both the Northern Cape and Western Cape provinces of South Africa, analysing approximately 1000 yield records from table grape vineyards measured in tons per hectare per block. The framework integrates Sentinel-2 satellite imagery with 33 predictor variables, comprising environmental, terrain and climate features. Three systematic experiments evaluated dataset combinations using ensemble machine learning (ML) models (Random Forest and Extreme Gradient Boosting) designed for high-dimensional geospatial data. The framework achieved strong regional performance with Random Forest achieving an $R^2 = 0.65$ and RMSE = 4.53 tons/hectare for the Western Cape (WC), demonstrating reliable predictive capability using freely available satellite data. For the Northern Cape (NC) Random Forest achieved an $R^2 = 0.37$ and RMSE = 6.4 tons/hectare, likely attributable to the compressed 3-month growing window (September to late November/early December) versus the Western Cape's extended 7-month cycle (September to March), providing significantly fewer temporal data for model training. Despite these regional variations in growing seasons, predictions are available 1-3 months before harvest (October for NC and December for WC) enabling proactive yield management preparation across both regions. In comparison to previous literature this study provides a more thorough approach at predicting viticulture yield, focusing on viticulture blocks spanning 250km in NC and 200km in WC. While other studies reported high accuracies (R2 > 0.8), they often focus on single areas/production units or on the association/correlation between vegetation indices and yield rather than the predictive ability. This study fills this gap by comparing the actual vs predicted yield achieved by the ML models. With the drive for high-resolution satellite imagery and the high costs associated, these results highlight the potential of open-source remote sensing combined with machine learning to provide practical early yield predictions for viticulture applications.

A Machine Learning-Based Multi-Temporal, Multi-Source Approach for Accurate Grape Yield Estimation

Aron WALSH, Kyle LOGGENBERG, and Albert STREVER, South Africa

1 INTRODUCTION

Precision viticulture seeks to optimise vineyard productivity through targeted management informed by detailed monitoring of vine performance and environmental conditions (Shawon et al., 2024). Accurate yield estimation is a key objective in precision viticulture as it enables growers to allocate resources effectively, coordinate harvest logistics, anticipate market supply, and identify underperforming areas requiring intervention (Van Klompenburg et al., 2020; Abdel-salam et al., 2024). Traditional yield estimation approaches rely on grower experience, historical production records, and pre-harvest manual sampling. Whilst these methods can provide reasonable estimates for small operations, they are spatially limited, labour-intensive, and become economically prohibitive for large-scale commercial vineyards. Manual sampling captures only a small fraction of the vineyard area, and destructive sampling methods, such as cluster removal for weight measurements, directly reduce harvestable yield when applied extensively (Barriguinha et al., 2021). Furthermore, reliance on historical records fails to account for inter-annual climatic variability, undermining prediction reliability (Pádua et al., 2017; Anastasiou et al., 2018).

To overcome these limitations, remote sensing (RS) technologies provide opportunities for spatially extensive, non-destructive yield monitoring throughout the growing season (Barriguinha et al., 2021). Satellite platforms such as Sentinel-2 offer high temporal coverage (a 5-day revisit time using both Sentinel-2A and 2B satellites) alongside high spatial resolution (10m), making them ideally suited for vineyard-scale analysis (Spoto et al., 2012; Anastasiou et al., 2018). These spaceborne RS technologies enable the calculation of vegetation indices derived from spectral reflectance of both crop and soil. Vegetation indices, such as the Normalised Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI), have been correlated with canopy condition, water status, and productivity in viticultural systems (Bonilla et al., 2015; Sun et al., 2017; Ballesteros et al., 2020). Compared to manual sampling or high-resolution aerial or terrestrial platforms, satellite-based approaches offer substantial cost advantages for large-scale monitoring of vineyard holdings.

The integration of satellite imagery with advanced analytical approaches has further enhanced predictive capabilities (Van Klompenburg et al., 2020). Machine learning algorithms have demonstrated capacity to model complex, non-linear relationships between remotely sensed observations and crop yield across multiple agricultural systems (Van Klompenburg et al., 2020). Commonly used machine learning algorithms in viticulture include: Artificial Neural Networks (ANN) (Ballesteros et al., 2020), Random Forest (RF) (Maimaitiyiming et al., 2019; Andrade et al., 2023), and Support Vector Regression (SVR) (Palacios et al., 2023). Ensemble methods such as RF and gradient boosting approaches mitigate the potential for overfitting through their aggregation of multiple decision trees (Breiman, 2001; Chen et al., 2015), thus showing promise for capturing multivariate interactions amongst spectral, climatic, and topographic predictors.

Several studies have combined RS data with machine learning methods for yield prediction in viticulture; however, the reported predictive accuracies differ substantially across platforms and modelling strategies. Satellite-based approaches have shown mixed results: Arab et al. (2021) achieved high prediction accuracy (R = 0.94–0.95) using ANN predictive modelling with Landsat 8-derived NDVI across 31 Afghan vineyards, identifying maximum canopy expansion as the optimal prediction timing, whilst Anastasiou et al. (2018) found more modest performance for satellite-derived estimations (R^2 = 0.33 at harvest) when compared with proximal sensing (R^2 = 0.31 at mid-veraison), with the latter recommended for earlier prediction.

Higher-resolution platforms have generally demonstrated superior accuracy: López-García et al. (2022) achieved an R^2 = 0.98, and root mean square error (RMSE) = 0.21 kg vine^{-1} using vegetation indices from red, green, blue (RGB) and multispectral sensors to characterise intra-plot variability, whilst Palacios et al. (2023) demonstrated R^2 values of 0.54–0.87 for early-season yield prediction across six Spanish grape varieties using computer vision and SVR, with NDVI identified as the most important predictor. Ballesteros et al. (2020) found that machine learning techniques with high-resolution UAV-based remote sensing imagery resulted in more accurate results than linear models (RMSE = 0.5 kg vine^{-1} and relative error [RE] = 12.1%), with more precise yield predictions obtained when images were acquired close to harvest when canopy structure and fruit load are more stable. However, UAV and aircraft platforms, while offering finer spatial resolution, involve higher operational costs, greater logistical complexity, and limited scalability across large commercial operations.

Despite these advances, existing studies have focused primarily on localised geographic extents, often encompassing single vineyard estates (Sun et al., 2017; Anastasiou et al., 2018; Arab et al., 2021; López-Garcia et al., 2022; Palacios et al., 2023; Taylor et al., 2023). This spatial limitation restricts understanding of model transferability across diverse viticultural landscapes and raises concerns about spatial autocorrelation, as studies incorporating only geographically clustered vineyards may overestimate model accuracy due to spatial dependence amongst observations (Ferraciolli et al., 2019). Furthermore, whilst various machine learning algorithms have been applied to viticulture yield prediction, Extreme Gradient Boosting (XGBoost) remains unexplored despite its superior performance in other crop yield predictions (Mariadass et al., 2022; Darra et al., 2023). Critical gaps remain in understanding yield prediction model performance across large geographic scales, multiple commercial varieties, and contrasting climatic regimes over multi-year periods.

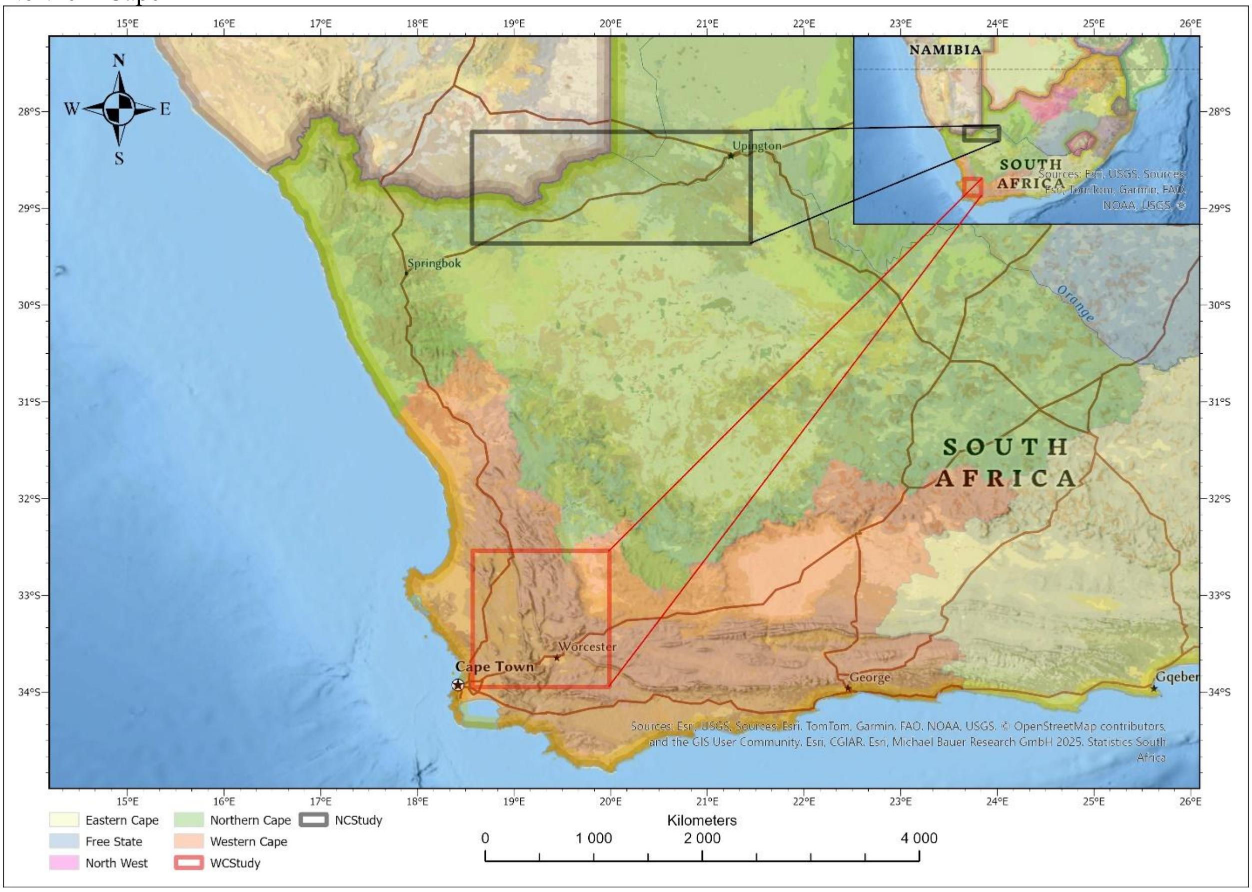
This study, therefore, aims to assess a multisource, multi-temporal, remote sensing-based machine learning framework for grape yield prediction across different vineyard environments. This will be achieved by evaluating RF and XGBoost algorithms for table grape yield prediction across two climatically distinct South African provinces, the Western Cape (Mediterranean climate) and Northern Cape (semi-arid climate), over five growing seasons (2021–2025). The analysis incorporates 27 commercial cultivars across 240 commercial vineyards spanning approximately 900 hectares. To our knowledge, this represents the first multi-regional, multi-cultivar evaluation of XGBoost for vineyard yield prediction at commercial scale across contrasting climatic zones. Overall, the research assesses the operational feasibility of satellite-based yield prediction for large-scale viticultural enterprises and provides evidence-based guidance on algorithm selection and deployment timing under diverse environmental conditions.

2 METHODS

2.1 Study Area

This study focuses on table grape vineyards in both the Western Cape and Northern Cape province (Figure 2.1), South Africa. These areas are known commercial regions for wine grapes, table grapes and raisin grapes in South Africa.

Figure 2.1: Study area with vineyards used in this study covering 200km of the Western Cape and 250km of the Northern Cape



2.2 Input Data and Data Preprocessing

This study employed a multimodal modelling approach, integrating heterogeneous data modalities, including multi-temporal satellite wavebands, satellite-derived spectral indices, terrain variables, and climatic data. Together, these predictors represented both direct drivers and proxy indicators of grape yield. This approach aimed to cover the underlying uncertainty in which data combinations provide a greater role in predicting viticulture yield. Prior to model training, several preprocessing steps were applied to ensure data quality and consistency. Monthly median composites of Sentinel-2 imagery were created for each growing season month (September–January) to reduce cloud cover impacts, and all input features including spectral bands, vegetation indices, terrain variables, and climate data, were spatially aggregated to the vineyard block level using zonal statistics. Outliers beyond two standard deviations from the mean were removed, eliminating 27 records and resulting in a final dataset of 985 observations across five growing seasons. Each of the 27 grape cultivars was assigned a unique numerical identifier for model inclusion, and all predictor variables were compiled into a feature matrix where each row represented a unique block-year combination, with monthly values structured as separate columns using numeric suffixes (e.g., NDVI_1 for September through NDVI_5 for January). The response variable (yield in tonnes per hectare) was matched to corresponding block-year combinations, and no feature standardization was applied since the tree-based ensemble methods used, such as Random Forest and XGBoost which are invariant to monotonic transformations of predictor variables.

2.2.1 Yield Data

Yield data were obtained from Karsten Boerdery, covering approximately 902 ha of table grape vineyards across 27 cultivars. Vineyard blocks ranged from 0.23 ha to 9.91 ha in size, with yields recorded at the block level in tons per hectare. Data were available for five growing seasons (2021–2025). Grapes were harvested manually by field workers and transported to an on-site packhouse. At the packhouse, an automated system packed grapes into standardised bins, with each bin weighed to ensure consistent mass. During this process, grapes failing to meet quality standards based on visual assessment were removed and recorded. Final yield for each block was calculated from the total mass of packed grapes and expressed as tons per hectare. The Northern Cape has 24 different cultivars while the Western Cape has 3 different cultivars and 10 of the same cultivars as the Northern Cape.

2.3 Satellite Imagery

Multispectral imagery from Sentinel-2 was acquired using Google Earth Engine (GEE), utilising bands 2, 3, 4, 5, 6, 7, 8, 8A, 11, and 12, ranging from 443-2190nm. All bands were resampled to 10m spatial resolution within GEE to ensure consistent spatial alignment. Monthly median composites were created to achieve two objectives: firstly, to ensure consistent temporal coverage by mitigating challenges associated with cloud cover that may obscure individual acquisition dates; and secondly, to acquire a more stable representation of canopy conditions by smoothing short-term variability and accommodating spatial differences in phenological development across vineyard blocks.. To address cloud contamination, per-pixel cloud masking was applied using a cloud pixel percentage threshold of 30%. Surface reflectance values were used to obtain representative ground reflectance, with atmospheric correction being pre-applied in GEE (Zhang et al., 2021).

2.4 Vegetation Indices

Nineteen vegetation indices, selected based on previous literature, were calculated from the Sentinel-2 imagery to characterise both soil conditions and vegetation health (see Table 2.1). This study expanded on these commonly used indices by assessing their relative importance alongside other vegetation indices reported in crop yield prediction literature. The selected indices capture different aspects of canopy structure, chlorophyll content, water stress, and soil background effects, thereby addressing known spectral complexities in vineyard environments (Matese et al. 2015).

Table 2.1: Outline of vegetation indices used in this study along with their equations

Vegetation Index	Abbreviation
Atmospherically Resistant Vegetation Index	ARVI
Bare Soil Index	BSI
Enhanced Vegetation Index	EVI
Enhanced Vegetation Index 2	EVI2
Ferric Oxides	-
Ferrous Iron	-
Green Normalised Vegetation Index	GNDVI
Green-Red Vegetation Index	GRVI
Laterite	-
Modified Soil Adjusted Vegetation Index	MSAVI
Normalised Difference Moisture Index	NDMI
Normalised Difference Red Edge Index	NDRE
Normalized Difference Salinity Index	NDSI
Normalised Vegetation Index	NDVI
Optimised Soil Adjusted Vegetation Index	OSAVI
Ratio Vegetation Index	RVI
Soil moisture index	S2WI
Soil Adjusted Vegetation Index	SAVI
Soil Moisture Index	SMI

2.5 Terrain and Climate Variables

Terrain variables were derived from the 5m Stellenbosch University Digital Elevation Model (SUDEM) provided by Geosmart. SUDEM was resampled to 10m spatial resolution to match the Sentinel-2 imagery, providing finer topographic detail than freely available alternatives such as the 30m Shuttle Radar Topography Mission (SRTM) DEM. Nine terrain variables were calculated from the resampled DEM, including: aspect, elevation, flow direction, hillshade, slope, roughness, Topographic Position Index (TPI), Terrain Ruggedness Index (TRI), and Topographic Wetness Index (TWI). These variables were computed using GDAL and used as spatial predictors to represent micro-scale environmental variability within the vineyard (add refs that justify the selection of these variables).

Climate data were obtained from on-site weather stations maintained by Karsten Boerdery, with measurements recorded at each production unit. The dataset included key meteorological variables: relative humidity (%), air temperature (°C), wind speed (m/s), solar radiation (W/m²), and precipitation (mm). Hourly measurements were aggregated to average, minimum, and maximum values to achieving a daily summary statistic, after which were subsequently used to calculate monthly summary statistics. To capture the full range of climatic variability affecting vine physiology and fruit development, summary statistics were calculated for each variable at monthly intervals over the five growing seasons (2021–2025). Specifically, minimum, maximum, and mean values were derived for each month during the five-month period (September through January). By accounting for these summary statistics, this approach captures factors that may disproportionately influence yield outcomes compared to monthly averages alone. In cases where data gaps occurred, measurements from the nearest weather station (within 10 km) were used to maintain temporal continuity.

2.6 Experimental Design

To determine the optimal model configuration for grape yield prediction, three experiments were designed. Each experiment varied the inclusion of spectral data (raw bands and vegetation indices), terrain variables, and climate variables, allowing for assessment of their individual and combined predictive value (see Table 2.2). All experiments utilised data from five growing seasons (2021–2025) spanning the critical phenological period (September–January), with machine learning models trained and validated for each month using an 80/20 train-test split. For all experiments, monthly summary statistics (minimum, maximum, mean, standard deviation, and median) were calculated per block for each vegetation index and climate variable, whilst terrain variables were represented by their median values due to their spatial stability within vineyard blocks

Table 2.2: Experimental Design

Experiment	Description
Experiment 1	Combining 19 vegetation indices with terrain and climate variables
Experiment 2	Terrain and climate variables only
Experiment 3	Sentinel Bands, vegetation indices, terrain, and climate variables

2.7 Machine Learning Models

Two supervised regression algorithms were implemented to model and predict vineyard yield: Random Forest (RF) and Extreme Gradient Boosting (XGBoost). Both methods are ensemble techniques capable of capturing non-linear relationships and feature interactions. Random Forest is an ensemble method that constructs multiple decision trees during training and outputs the mean prediction across all trees (Breiman, 2001). RF was selected for its robustness to overfitting, ability to handle high-dimensional data without feature scaling, and inherent provision of feature importance metrics. The model was implemented using the scikit-learn library in Python with initial hyperparameters of 500 estimators, maximum depth of 9, minimum samples split of 2, and minimum samples per leaf of 1.

Extreme Gradient Boosting (XGBoost) is a gradient boosting framework that sequentially builds decision trees, with each subsequent tree correcting errors made by previous trees (Chen et al., 2015). XGBoost was selected because it has demonstrated superior performance in crop yield prediction literature across multiple agricultural systems (Van Klompenburg et al., 2020), yet remains underexplored in viticulture applications. XGBoost is known for its high predictive accuracy, computational efficiency, and capacity to model complex non-linear relationships through its gradient-based optimisation. The model was implemented using the xgboost library in Python with default hyperparameters of 500 estimators, maximum depth of 6, learning rate of 0.1, subsample ratio of 0.8, and column subsample ratio of 0.8.

Hyperparameter tuning was performed using grid search with 5-fold cross-validation on the training set, optimising for R^2 score. For Random Forest, the grid search evaluated 192 combinations across the following parameter ranges: number of estimators (100, 300, 500), maximum depth (10, 20, 30, None), minimum samples split (2, 5), minimum samples per leaf (1, 2), maximum features ('sqrt', None), and bootstrap (True, False). For XGBoost, the grid search evaluated 1,296 combinations across: number of estimators (100, 300, 500), maximum depth (3, 6, 10), learning rate (0.01, 0.1, 0.2), subsample ratio (0.8, 1.0), column subsample ratio (0.8, 1.0), minimum child weight (0.5, 1, 3), gamma (0, 0.1), L1 regularisation (0, 0.1), and L2 regularisation (1, 10). The dataset was split into 80% training and 20% testing for all experiments. A random seed of 1234 was set for both algorithms to ensure reproducibility across all experiments. Models were trained separately for each experimental configuration, allowing for comparison of predictive performance across different feature combinations.

2.8 Evaluation and Validation

Model performance was evaluated using three standard regression metrics used in literature (Van Klompenburg et al. 2020): the coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE). R^2 measures the proportion of variance in yield explained by the model, with values closer to 1 indicating better explanatory performance. RMSE quantifies the magnitude of prediction errors in the same units as the response variable (tonnes per hectare), with lower values indicating better performance. MAE provides the average absolute difference between predicted and observed yields, offering a more interpretable measure of typical prediction error that is less sensitive to outliers than RMSE. All metrics were calculated on the test set to assess model generalisability and predictive accuracy across different feature combinations and algorithms. Additionally, R^2 was calculated on the training set to monitor model performance during training and detect potential overfitting by comparing training and test set performance.

3 RESULTS AND DISCUSSION

This study evaluated three experimental configurations to predict table grape yield using RF and XGBoost across five growing seasons (2021–2025) in the Northern Cape and Western Cape provinces of South Africa.

3.1 Algorithm Comparison

Figures 3.1A and 3.1B present the observed versus predicted yield for the best-performing RF models in the Western Cape and Northern Cape, respectively. The Western Cape model demonstrated substantially stronger predictive performance compared to the Northern Cape with tighter clustering around the 1:1 line. Figures 3.2A and 3.2B show the corresponding results for the best-performing XGBoost models in each region, with the Western Cape again exhibiting substantially better performance than the Northern Cape.

RF outperformed XGBoost across both regions. In the Western Cape, RF achieved a maximum test R^2 of 0.65 whilst XGBoost achieved a maximum test R^2 of 0.59. In the Northern Cape, RF achieved a maximum test R^2 of 0.37, whilst XGBoost achieved a maximum test R^2 of 0.29. For the Western Cape, RF outperformed XGBoost in all three experiments. Similarly for the Northern Cape, RF outperformed XGBoost in all three experiments, with both algorithms achieving similar performance ($R^2 = 0.21$) in Experiment 1.
Training performance differed substantially between algorithms. XGBoost training R^2 values frequently reached 0.99, whilst RF training R^2 values ranged from 0.78 to 0.91. The gap between training and test performance was larger for XGBoost (training-test difference up to 0.84) compared to RF (training-test difference up to 0.59).

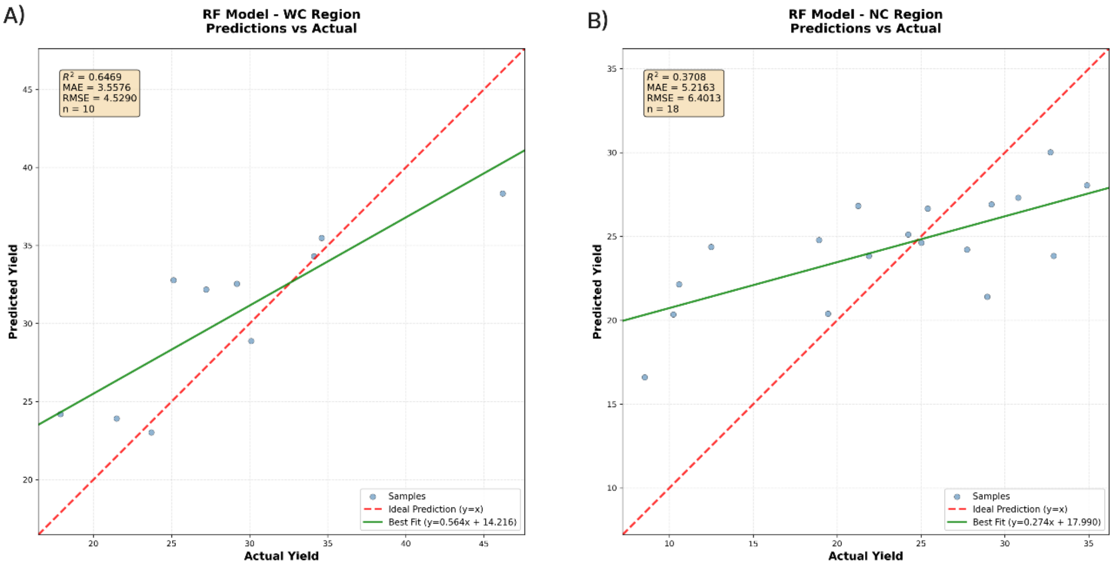


Figure 3.1: Actual versus predicted yield for the best-performing RF model in A) the Western Cape (Experiment 2, 2025 January cumulative). Test $R^2 = 0.65$, RMSE = 4.53 t/ha, MAE = 3.56 t/ha. And B) the Northern Cape (Experiment 2, 2021 October). Test $R^2 = 0.37$, RMSE = 6.40 t/ha, MAE = 5.22 t/ha.

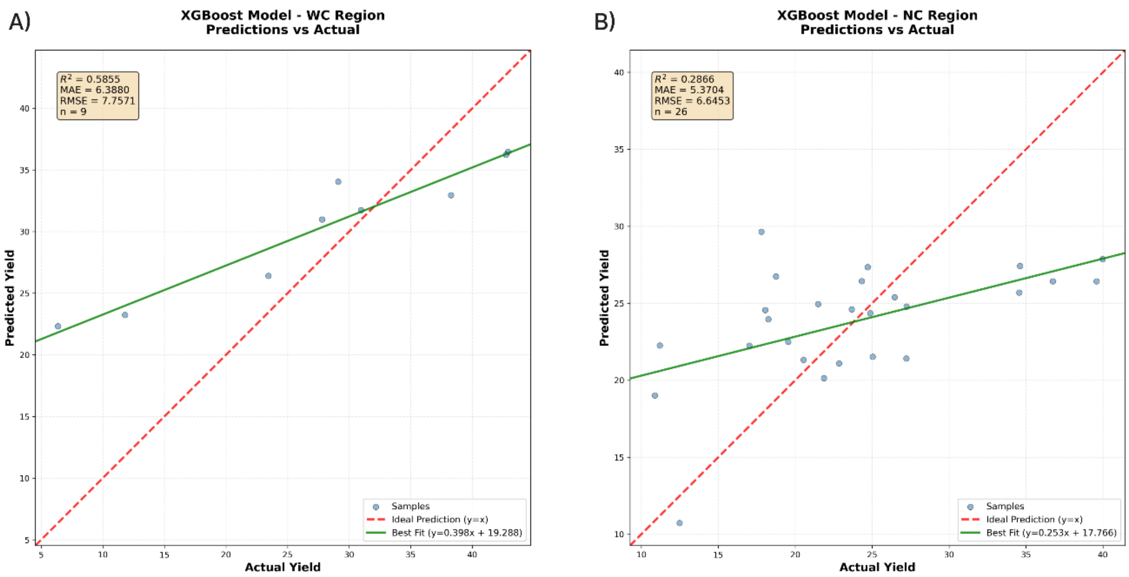


Figure 3.2: Actual versus predicted yield for the best-performing XGBoost model in A) the Western Cape (Experiment 1, 2022 December cumulative). Test R^2 = 0.59, RMSE = 7.76 t/ha, MAE = 6.39 t/ha. And B) the Northern Cape (Experiment 3, 2022 November). Test R^2 = 0.29, RMSE = 6.65 t/ha, MAE = 5.37 t/ha.

3.2 Regional and Temporal Patterns

Temporal patterns varied by region. In the Western Cape, January models demonstrated the strongest performance for RF, with Experiments 1, 2, 3 achieving their highest test R^2 values using January data (ranging from R^2 = 0.33 to 0.65). December models showed strong performance for XGBoost, with Experiments 1, 3 achieving their best results in December (R^2 = 0.28 to 0.59). In the Northern Cape, October and November models dominated across both algorithms, with the majority of the best-performing configurations occurring during these months. Cumulative models outperformed single-month models in select configurations. For the Western Cape, Experiment 2 achieved its highest performance using cumulative data (R^2 = 0.65 for RF, R^2 = 0.46 for XGBoost), whilst single-month models performed better for most other experiments. For the Northern Cape, single-month models consistently outperformed cumulative models across all experiments and algorithms.

These results provide a more realistic assessment of yield prediction capabilities than many published viticulture studies. Whilst studies by Sun et al. (2017), Anastasiou et al. (2018), Arab et al. (2021), López-Garcia et al. (2022), Palacios et al. (2023), and Taylor et al. (2023) demonstrated higher accuracy, they focused on single vineyard blocks or individual production units, potentially suffering from spatial autocorrelation and limited variation in growing conditions (Ferraciolli et al., 2018; Barriguinha et al., 2021). When viticulture blocks in close proximity share similar environmental and management conditions, models may learn location-specific patterns rather than generalisable yield relationships (Ferraciolli et al., 2018).

In contrast, this study encompassed multiple commercial vineyards spanning 200–250km distances within each province and incorporating 27 table grape cultivars. This geographic and genetic diversity introduces substantial heterogeneity in management practices, soil conditions, training systems, irrigation regimes, and pest/disease pressures. Studies conducted at smaller spatial scales or with limited cultivar diversity may report higher R^2 values (often exceeding 0.90), but such results may not reflect the prediction challenges inherent in large-scale commercial operations. Furthermore, while studies done by Anastasiou et al. (2018), and Arab et al. (2021) demonstrate associations between yield and vegetation indices using Pearson correlations but do not perform actual yield prediction with independent test sets. Whilst these studies suggest strong relationships between selected variables and yield, they provide only theoretical expectations based on correlation analysis rather than validated predictive performance on unseen data.

3.3 Regional Differences and Model Generalisability

Model performance differed substantially between the Western Cape and Northern Cape regions. Western Cape models achieved test R^2 values ranging from 0.27 to 0.65, while Northern Cape models achieved values ranging from 0.12 to 0.37. The best-performing Western Cape model ($R^2 = 0.65$) outperformed the best Northern Cape model ($R^2 = 0.37$) by 0.28 R^2 units, representing a 76% improvement in explained variance, with a performance gap that persisted across algorithms and experimental configurations.

Several factors likely explain this regional difference. First, the Western Cape exhibits more moderate and stable Mediterranean climatic conditions with winter rainfall and cooler growing season temperatures. Western Cape mean temperatures, humidity, and Normalised Difference Vegetation Index (NDVI) values showed less extreme variability than Northern Cape conditions. These relatively predictable environmental conditions may create stronger and more consistent relationships between remotely sensed vegetation indices and final yield. Additionally, the Western Cape has a substantially longer growing season, extending from September through February or March, compared to the Northern Cape's compressed 3-month window from September to late November or early December. This extended growing period allows vegetation indices to capture a fuller phenological progression from budburst through harvest. The longer accumulation period may also create more stable predictor-yield relationships, as yield formation occurs over a more gradual timeframe with less sensitivity to individual stress events.

In contrast, the Northern Cape experiences more extreme continental climate conditions with summer rainfall, higher temperatures, and greater inter-annual variability. Furthermore, the Northern Cape's compressed growing season—spanning only 3 months from September to late November or early December compared to the Western Cape's 6–7 month window (September to February/March)—concentrates all phenological stages into a shorter timeframe. This condensed growing period provides fewer temporal observation points and may increase the impact of individual stress events, as there is less time for vines to recover from mid-season disruptions. The abbreviated season also limits the temporal signal available for remote sensing-based prediction, potentially contributing to the weaker model performance observed in this region. Northern Cape yields ranged from 23.54 t/ha (2021) to 25.73 t/ha (2025), exhibiting greater year-to-year volatility than Western Cape yields, which ranged from 33.21 t/ha (2021) to 28.50 t/ha (2025). Such environmental instability weakens relationships between mid-season spectral measurements and harvest yield, as late-season stress events or management interventions can substantially alter final productivity.

These regional performance differences have critical implications for operational yield prediction systems. Models developed and validated in one region cannot be assumed to maintain equivalent performance when applied to climatically or geographically distinct regions. Many published studies focus on single vineyards spanning limited geographic extents (<10km). Such models may perform well in localised contexts but fail when applied across diverse terroirs, climates, and management systems.

The 0.28 R^2 unit performance gap between regions demonstrates that a single universal model would substantially underperform region-specific models tailored to local conditions. However, it remains unclear whether training a single model on data from both regions could bridge this performance gap or whether region-specific models will always be necessary. This finding emphasises the importance of incorporating regional climate characteristics, growing season length, and environmental variability when developing operational yield prediction systems for commercial viticulture.

3.4 Future Works and Limitations

This study did not include vineyard management data such as irrigation scheduling, canopy management practices, disease and pest control interventions, and harvest timing decisions. Future studies could incorporate these factors to determine whether they improve predictive accuracy by accounting for management-induced yield variability. Additionally, satellite observations were limited to cloud-free conditions, which varied in availability across the study period. While multi-temporal aggregations partially mitigate this limitation, they may obscure short-term stress events or management effects that occur between observation dates. Future studies could explore higher temporal frequency imagery (e.g., bi-weekly composites) to capture short-term dynamics. Future studies could investigate the spatial and temporal transferability and the impact of spatial resolution on grape yield estimation accuracy.

4 CONCLUSION

This study aimed to develop a multisource, multi-temporal, remote sensing-based machine learning framework for grape yield prediction across different vineyard environments. This aim was achieved by evaluating Random Forest (RF) and Extreme Gradient Boosting (XGBoost) algorithms for table grape yield prediction across two climatically distinct South African provinces (Western Cape: Mediterranean climate; Northern Cape: semi-arid climate) over five growing seasons (2021–2025), incorporating 27 commercial cultivars across 240 commercial vineyards spanning 900 hectares. During this study, critical limitations in existing viticulture yield prediction literature were addressed through the following objectives: (1) develop machine learning models to predict grape yield using Sentinel-2 multispectral imagery; (2) compare predictive performance of RF and XGBoost across diverse geographic and climatic conditions; (3) assess optimal timing for satellite-based yield prediction relative to harvest. By assessing these objectives this study provides value in the operational feasibility of satellite-based yield prediction for large-scale viticultural enterprises.

There are several key findings from these objectives: First, machine learning models successfully predicted grape yield using Sentinel-2 multispectral imagery. Second, RF consistently outperformed XGBoost, achieving the best predictive performance in the Western Cape (R^2 = 0.65, RMSE = 4.53 t/ha, MAE = 3.56 t/ha) and Northern Cape (R^2 = 0.37, RMSE = 6.40 t/ha, MAE = 5.22 t/ha). The substantial performance gap between regions (0.28 R^2 units) raises the question of whether models calibrated in one viticulture region can be used to predict yield in climatically distinct areas. Third, optimal timing was identified as 1-2 months pre-harvest (January for Western Cape, October-November for Northern Cape). Notably,

This study demonstrates that many published viticulture yield prediction studies substantially overestimate operational accuracy. The performance disparity between Anastasiou et al. (2018), who achieved an R^2 = 0.33 and studies reporting R^2 > 0.90 in single-vineyard contexts highlights how study design directly impacts reported metrics. Several studies showcase associations between spectral indices and yield through correlation analysis rather than validated predictive performance on independent test sets, further exacerbated by the predominance of single-cultivar, geographically limited studies that often reflect favourable validation conditions rather than robust model generalisability.

5 REFERENCE

Abdel-salam, M., Kumar, N., & Mahajan, S. (2024). A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning. *Neural Computing and Applications*, *36*(33), 20723-20750.

Anastasiou, E., Balafoutis, A., Darra, N., Psiroukis, V., Biniari, A., Xanthopoulos, G., & Fountas, S. (2018). Satellite and proximal sensing to estimate the yield and quality of table grapes. *Agriculture*, *8*(7), 94.

Andrade, C. B., Moura-Bueno, J. M., Comin, J. J., & Brunetto, G. (2023). Grape yield prediction models: Approaching different machine learning algorithms. Horticulture, 9(12), 1294.

Arab, S. T., Noguchi, R., Matsushita, S., & Ahamed, T. (2021). Prediction of grape yields from time-series vegetation indices using satellite remote sensing and a machine-learning approach. Remote Sensing Applications: Society and Environment, 22, 100485.

Ballesteros, R., Intrigliolo, D. S., Ortega, J. F., Ramírez-Cuesta, J. M., Buesa, I., & Moreno, M. A. (2020). Vineyard yield estimation by combining remote sensing, computer vision and artificial neural network techniques. *Precision Agriculture*, *21*, 1242-1262.

Barriguinha, A., de Castro Neto, M., & Gil, A. (2021). Vineyard yield estimation, prediction, and forecasting: A systematic literature review. Agronomy, 11(9), 1789.

Bonilla, I., de Toda, F. M., & Martínez-Casasnovas, J. A. (2015). Vine vigour, yield and grape quality assessment by airborne remote sensing over three years: Analysis of unexpected relationships in cv. Tempranillo. *Spanish Journal of Agricultural Research*, *13*(2), e0903-e0903.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., ... & Zhou, T. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4), 1-4.

Darra, N., Anastasiou, E., Kriezi, O., Lazarou, E., Kalivas, D., & Fountas, S. (2023). Can yield prediction be fully digitilized? A systematic review. *Agronomy*, *13*(9), 2441.

Ferraciolli, M. A., Bocca, F. F., & Rodrigues, L. H. A. (2019). Neglecting spatial autocorrelation causes underestimation of the error of sugarcane yield models. Computers and Electronics in Agriculture, 161, 233-240.

López-García, P., Ortega, J. F., Pérez-Álvarez, E. P., Moreno, M. A., Ramírez, J. M., Intrigliolo, D. S., & Ballesteros, R. (2022). Yield estimations in a vineyard based on high-resolution spatial imagery acquired by a UAV. Biosystems Engineering, 224, 227-245.

Maimaitiyiming, M., Sagan, V., Sidike, P., & Kwasniewski, M. T. (2019). Dual activation function-based Extreme Learning Machine (ELM) for estimating grapevine berry yield and quality. Remote Sensing, 11(7), 740.

Mariadass, D. A.-L., Moung, E. G., Sufian, M. M., & Farzamnia, A. (2022). Extreme Gradient Boosting (XGBoost) Regressor and Shapley Additive Explanation for Crop Yield Prediction in Agriculture. *2022 12th International Conference on Computer and Knowledge Engineering (ICCKE)*, 219–224. https://doi.org/10.1109/ICCKE57176.2022.9960069

Matese, A., Toscano, P., Di Gennaro, S. F., Genesio, L., Vaccari, F. P., Primicerio, J., Belli, C., Zaldei, A., Bianconi, R., & Gioli, B. (2015). Intercomparison of UAV, aircraft and satellite remote sensing platforms for precision viticulture. Remote Sensing, 7, 2971-2990. https://doi.org/10.3390/rs70302971

Pádua, L., Adão, T., Hruška, J., Sousa, J. J., Peres, E., Morais, R., & Sousa, A. (2017). Very high-resolution aerial data to support multi-temporal precision agriculture information management. *Procedia computer science*, *121*, 407-414.

Palacios, F., Diago, M. P., Melo-Pinto, P., & Tardaguila, J. (2023). Early yield prediction in different grapevine varieties using computer vision and machine learning. *Precision Agriculture*, *24*(2), 407-435.

Shawon, S. M., Ema, F. B., Mahi, A. K., Niha, F. L., & Zubair, H. T. (2024). Crop Yield Prediction Using Machine Learning: An Extensive and Systematic Literature Review. *Smart Agricultural Technology*, 100718.

Spoto, F., Sy, O., Laberinti, P., Martimort, P., Fernandez, V., Colin, O., ... & Meygret, A. (2012, July). Overview of sentinel-2. In 2012 IEEE international geoscience and remote sensing symposium (pp. 1707-1710). IEEE.

Sun, L., Gao, F., Anderson, M. C., Kustas, W. P., Alsina, M. M., Sanchez, L., ... & Post, K. (2017). Daily mapping of 30 m LAI and NDVI for grape yield prediction in California vineyards. Remote Sensing, 9(4), 317.

Taylor, J. A., Bates, T. R., Jakubowski, R., & Jones, H. (2023). Machine-Learning Methods to Identify Key Predictors of Site-Specific Vineyard Yield and Vine Size. American Journal of Enology and Viticulture, 74(1).

Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. Computers and electronics in agriculture, 177, 105709.

Zhang, M., Zhang, M., Yang, H., Jin, Y., Zhang, X., & Liu, H. (2021). Mapping regional soil organic matter based on sentinel-2a and modis imagery using machine learning algorithms and google earth engine. Remote Sensing, 13(15), 2934.