
# Estimating Grape Bunch Yield Using Convolutional Neural Networks and Proximal RGB Imaging in the Brazilian Pampa

**Sandro da Silva Camargo[1,2], Eduarda Menezes da Silveira[1,2], Fernando Infantini Satte Alam Nogueira[1,2], Ana Elisandra Freire Campos[1,2], Victoria Zara Mércio[3]**

[1]Center for Embedded Devices and Research in Digital Agriculture (CEDRA) of SENAI-RS, Brazil

[2]Universidade Federal do Pampa (UNIPAMPA), Brazil

[3]Vinícola Estância Paraizo, Brazil

**A paper from the Proceedings of the**

**17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture**

**13-15 July 2026**

**Porto Alegre, Brazil**

***Abstract.***

*Viticulture has become an increasingly important economic activity in the Pampa biome of southern Brazil, where reliable yield estimation is essential for harvest planning and vineyard management. Recent advances in computer vision and deep learning provide new opportunities for rapid, non-destructive, and scalable yield prediction using low-cost imaging systems. This study evaluated the feasibility of deep learning models for estimating grape cluster weight, a key yield-related trait, from proximal RGB images. A dataset comprising 50 grape bunches was assembled, with seven images acquired per bunch under varying acquisition distances to simulate realistic field conditions. After image acquisition, the fresh weight of each bunch was measured and used as the reference variable for supervised learning. Three convolutional neural network architectures (ResNet18, EfficientNetV2, and MobileNetV3) were trained and compared, with emphasis on lightweight models suitable for practical deployment in precision viticulture. All evaluated models showed satisfactory predictive performance, with coefficients of determination ($R^2$) ranging from 0.719 to 0.776. ResNet18 achieved the highest accuracy, yielding a correlation coefficient of 0.892 between predicted and observed cluster weights ($R^2$ = 0.776, MAE = 0.06 kg, RMSE = 0.08 kg). EfficientNetV2 and MobileNetV3 also produced competitive results, demonstrating the potential of computationally efficient architectures for agricultural applications.*

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture. EXAMPLE: Camargo, S. S. et al. (2026). Estimating Grape Bunch Yield Using Convolutional Neural Networks and Proximal RGB Imaging in the Brazilian Pampa. In Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture and Brazilian Congress on Precision and Digital Agriculture.

*The findings indicate that RGB imagery combined with deep learning can effectively capture visual features associated with grape cluster weight, enabling accurate predictions even from a relatively small dataset. Although additional validation across seasons, cultivars, and environmental conditions is required, the proposed methodology represents a low-cost and non-invasive approach for supporting yield estimation and decision-making in precision viticulture, particularly in emerging wine-producing regions.*

***Keywords.***
*Artificial Intelligence, Deep Learning, Productivity Estimation, Neural Networks*

# Introduction

Accurate yield estimation is a critical component of modern viticulture because it directly influences harvest logistics, labor allocation, winery planning, and economic decision-making. Reliable pre-harvest information enables growers to optimize vineyard management practices, improve resource allocation, and reduce uncertainties throughout the production chain. However, grapevine yield is characterized by substantial spatial and temporal variability resulting from the interaction of cultivar genetics, environmental conditions, management practices, and vineyard heterogeneity. Consequently, traditional yield assessment methods based on manual cluster counting, destructive sampling, or visual field inspections are often labor-intensive, time-consuming, and impractical for large-scale commercial vineyards. These limitations have motivated the development of digital and automated approaches capable of providing rapid, objective, and scalable yield estimates (Olenskyj et al., 2022; Victorino et al., 2022; Orlandi et al., 2025).

Recent advances in Precision Agriculture have accelerated the adoption of remote sensing, proximal sensing, and artificial intelligence techniques for crop monitoring and yield forecasting. Within this context, computer vision has emerged as one of the most promising technologies due to the increasing availability of low-cost RGB cameras, mobile devices, and unmanned aerial systems. Unlike conventional field sampling, image-based approaches enable non-destructive data acquisition while preserving spatial information across vineyards. Furthermore, RGB imaging systems are significantly more accessible than multispectral or hyperspectral sensors, facilitating practical implementation in commercial production systems. As a result, RGB-based computer vision has become an important research area for vineyard monitoring and digital viticulture applications (Mohimont et al., 2022; Olenskyj et al., 2022; Orlandi et al., 2025).

The application of computer vision to vineyard yield estimation has evolved from traditional image processing techniques toward deep learning-based approaches. Early studies focused primarily on cluster segmentation, berry counting, and extraction of handcrafted image features, often requiring complex calibration procedures and extensive manual intervention. Although these methods demonstrated the feasibility of image-based yield estimation, their performance was frequently affected by variable illumination conditions, canopy architecture, fruit occlusion, and cultivar-specific characteristics. Occlusion by leaves remains one of the principal challenges in vineyard image analysis because a significant proportion of grape clusters may not be fully visible in field-acquired images. Consequently, the development of robust machine learning methods capable of extracting meaningful yield-related information directly from images has become a major research focus (Mohimont et al., 2022; Victorino et al., 2022).

Convolutional Neural Networks (CNNs) have substantially transformed agricultural image analysis by enabling automatic feature extraction from raw image data. Unlike conventional machine learning approaches that depend on manually engineered descriptors, CNNs learn hierarchical visual representations directly from images, improving their ability to capture complex relationships between canopy characteristics and agronomic variables. This paradigm shift reduces the dependence on handcrafted image descriptors and allows predictive models to learn yield-related patterns directly from complex canopy structures, illumination conditions, and fruit visibility variations. In viticulture, CNN-based methods have been successfully employed for grape cluster detection, phenotyping, disease identification, berry counting, and yield prediction. More recently, end-to-end deep learning approaches have demonstrated that yield can be estimated directly from RGB imagery without requiring explicit object detection or handcrafted feature extraction, reducing annotation requirements and simplifying model deployment (Mohimont et al., 2022; Olenskyj et al., 2022).

The transition from classification tasks to regression-based deep learning models represents an important advancement for agricultural applications. Instead of merely identifying the presence or absence of grape clusters, regression models learn continuous relationships between image characteristics and quantitative variables such as cluster weight, berry number, or final yield. This capability is particularly relevant for commercial vineyard management because it allows direct prediction of agronomically meaningful outcomes. Recent studies have demonstrated that CNN-based regression models can explain a substantial proportion of grapevine yield variability directly from RGB imagery, achieving predictive performance comparable to or exceeding traditional object-counting approaches while requiring substantially less annotation effort. Additionally, image-based regression approaches eliminate several intermediate processing steps commonly required in object counting or segmentation pipelines, potentially increasing operational efficiency and scalability (Mohimont et al., 2022; 2022; Orlandi et al., 2025).

Simultaneously, the rapid development of efficient deep neural network architectures has created new opportunities for agricultural monitoring systems. While deep CNN models often achieve high predictive accuracy, their computational requirements may limit deployment in field conditions where processing resources are constrained. To address this challenge, lightweight architectures such as MobileNet and EfficientNet have been developed to provide improved computational efficiency while maintaining competitive predictive performance. MobileNet employs depthwise separable convolutions to reduce computational complexity and memory requirements, making it particularly suitable for embedded and mobile applications. EfficientNet introduces a compound scaling strategy that balances network depth, width, and image resolution, enabling superior accuracy-efficiency trade-offs compared with conventional CNN architectures. These characteristics have contributed to their growing adoption in agricultural computer vision applications, including crop disease diagnosis, plant stress detection, and image-based phenotyping (Talaviya et al., 2024; Koirala et al., 2023; Wspaniały et al., 2024). Therefore, lightweight architectures such as EfficientNetV2 and MobileNetV3 are particularly attractive for vineyard monitoring systems because they enable the deployment of deep learning models on edge devices, facilitating near real-time yield estimation directly in field conditions.

The emergence of Edge Artificial Intelligence (Edge AI) further reinforces the relevance of lightweight deep learning models in agriculture. Edge AI systems perform inference directly on local devices, reducing dependence on cloud computing infrastructure and enabling near real-time decision support. Such capability is particularly valuable in agricultural environments where internet connectivity may be limited or intermittent. Recent studies have demonstrated the feasibility of deploying deep learning models on embedded platforms for crop monitoring, plant detection, counting, and classification tasks, highlighting the potential of edge-based solutions for scalable agricultural sensing systems. The combination of RGB imaging, lightweight CNN architectures, and embedded processing platforms represents a promising pathway toward operational vineyard monitoring technologies capable of supporting precision management decisions in real time (Ayaz et al., 2022; Koirala et al., 2023; Nguyen et al., 2023). Although Edge AI deployment is not evaluated in this study, the computational efficiency of EfficientNetV2B0 and MobileNetV3Large makes these architectures promising candidates for future implementation in embedded vineyard monitoring systems.

Despite recent progress in deep learning for viticulture, important knowledge gaps remain regarding the use of lightweight convolutional neural network architectures for direct grapevine yield estimation under field conditions. Most studies in vineyard computer vision have focused on tasks such as grape cluster detection, segmentation, disease identification, and phenotyping, while relatively fewer investigations have addressed end-to-end regression models that directly predict yield from RGB imagery (Olenskyj et al., 2022; Zhang et al., 2024). Furthermore, although architectures such as MobileNet and EfficientNet have demonstrated promising performance in a wide range of agricultural computer vision applications, their potential for vineyard yield prediction remains insufficiently explored (Jabed & Murad, 2024). In particular, limited information is available on the trade-off between predictive accuracy and computational efficiency when these lightweight models are applied to relatively small agricultural datasets acquired under natural illumination, occlusion, and canopy variability (Jabed & Murad, 2024; Zhang et al., 2024). Addressing these challenges is essential for the development of practical image-based yield estimation systems that can support real-time and resource-efficient vineyard monitoring.

Therefore, this study evaluates the performance of EfficientNetV2B0 and MobileNetV3Large architectures for grapevine yield estimation using RGB images collected in commercial vineyard conditions. Transfer learning and regression-based deep learning approaches are employed to predict grape production directly from image data. The objective is to compare the predictive capability of both architectures and assess their suitability for future implementation in computationally efficient digital viticulture and precision agriculture systems.

## Material and Methods

### 2.1 Study Area and Experimental Material

The study was conducted using grape bunches collected from commercial vineyards located in the Pampa Biome of southern Brazil, an emerging wine-producing region characterized by approximately two decades of commercial viticultural development. The region has experienced rapid expansion of premium wine grape cultivation due to favorable edaphoclimatic conditions and increasing investment in precision agriculture technologies.

A total of 50 grape bunches were selected from a single vineyard population for image acquisition and productivity assessment. The bunches were sampled to represent the natural variability in size, morphology, compactness, and weight commonly observed under commercial production conditions within that population. After image acquisition, each bunch was individually weighed using a precision digital scale, and fresh bunch weight (kg) was adopted as the target variable for model development.

### 2.2 Image Acquisition

RGB images were acquired using a digital camera with a native resolution of 1280 × 720 pixels. To increase variability in image characteristics and simulate realistic field acquisition conditions, each bunch was photographed from seven distinct distances: 30, 45, 60, 75, 90, 105, and 120 cm, according the protocol used in Barbole & Jadhav (2023).

The image acquisition protocol was designed to capture variations in apparent bunch size, perspective, and spatial resolution that may occur during practical deployment of image-based yield estimation systems. As a result, the final dataset comprised 350 RGB images corresponding to 50 individual bunches.

Each image was associated with the measured fresh weight of its respective bunch, generating a supervised learning dataset suitable for regression tasks.

### 2.3 Dataset Preparation

The dataset consisted of 350 RGB images obtained from 50 grape bunches, with seven images acquired for each bunch at different camera-to-target distances. Because multiple images corresponded to the same biological sample, dataset partitioning was performed at the bunch level rather than at the image level to prevent information leakage. Consequently, all images associated with a given bunch were assigned exclusively to either the training, validation, or testing subset.

The dataset was randomly divided into 70%, 15%, and 15% of the bunches for training, validation, and testing, respectively. This procedure resulted in approximately 35 bunches for model training, 7–8 bunches for validation, and 7–8 bunches for independent testing.

Images were resized according to the input requirements of each CNN architecture and normalized using the mean and standard deviation values associated with the corresponding ImageNet pretrained weights. Data augmentation was applied exclusively to the training subset and included random horizontal flipping, rotations of up to ±15°, random scaling (±10%), and brightness adjustments. These operations were selected to improve model robustness while preserving biologically meaningful visual characteristics.

## 2.4 Deep Learning Models

Three convolutional neural network architectures were evaluated: ResNet-18, EfficientNetV2-B0, and MobileNetV3-Large. These architectures were selected because they represent different trade-offs between predictive performance and computational efficiency while remaining suitable for deployment on embedded and edge-computing platforms.

Transfer learning was employed using weights pretrained on the ImageNet dataset. For each architecture, the original classification layer was replaced with a regression head composed of a fully connected layer producing a single continuous output corresponding to grape bunch fresh weight (kg).

During fine-tuning, pretrained feature extraction layers were initialized with ImageNet weights, whereas the regression head was initialized using Xavier uniform initialization. Network parameters were subsequently optimized through supervised learning using the measured bunch weights as target values.

## 2.5 Model Training

Model development was implemented in Python using the PyTorch framework. The Mean Squared Error (MSE) loss function was adopted because of its suitability for continuous regression problems.

Training was performed using the Adam optimizer with an initial learning rate of $1 \times 10^{-4}$ and a batch size of 16 images. Models were trained for a maximum of 100 epochs. Early stopping was applied using a patience value of 10 epochs based on validation loss to reduce overfitting and improve generalization.

The model presenting the lowest validation loss was retained for final testing. Hyperparameter optimization was conducted through preliminary experiments evaluating different learning rates, dropout configurations, and batch sizes. The final configuration was selected according to validation performance.

## 2.6 Performance Evaluation

Model performance was evaluated using an independent testing dataset not used during training or hyperparameter tuning. Predicted bunch weights were compared with observed values measured using a precision digital scale.

Performance was quantified using the coefficient of determination ($R^2$), Pearson correlation coefficient (r), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE):

$$R^2 = 1 - \Sigma(y_i - \hat{y}_i)^2 / \Sigma(y_i - \bar{y})^2 \quad (1)$$

$$MAE = (1/n)\ \Sigma\ |y_i - \hat{y}_i| \quad (2)$$

$$RMSE = \sqrt{[(1/n)\ \Sigma(y_i - \hat{y}_i)^2]} \quad (3)$$

where $y_i$ represents observed bunch weight, $\hat{y}_i$ represents predicted bunch weight, and $n$ is the number of observations.

Scatter plots of measured versus predicted values were generated to visually assess model behavior and identify potential systematic prediction biases.

### 2.7 Computational Environment

All experiments were conducted using Python 3.11, PyTorch 2.x, NumPy, Scikit-learn, and OpenCV libraries. Model training and evaluation were performed on a workstation equipped with an NVIDIA GPU supporting CUDA acceleration.

The complete processing pipeline was designed to support future deployment in embedded and edge-computing environments, enabling near real-time image-based grape yield estimation under field conditions.

## Results and Discussion

The distribution of grape cluster weights exhibited a clear positive skew, with values ranging from 0.141 to 1.009 kg and a mean weight (0.390 kg) slightly higher than the median (0.358 kg), indicating the presence of relatively heavier clusters in the dataset. The interquartile range (0.285–0.486 kg) showed that most samples were concentrated within the lower and intermediate weight classes, while only a limited number of observations occupied the upper end of the distribution. This imbalance is consistent with the observed concentration of samples around smaller cluster weights and may partially explain the reduced predictive accuracy for larger clusters. Furthermore, the Shapiro-Wilk test revealed a significant departure from normality ($W = 0.876$, $p < 0.001$), confirming that the target variable does not follow a Gaussian distribution. Such a distribution pattern is common in agricultural production datasets, where biological and environmental factors often generate asymmetric weight distributions. The predominance of lighter clusters may have influenced the learning process, favoring model performance within the most represented weight range while reducing accuracy for less frequent, heavier clusters.

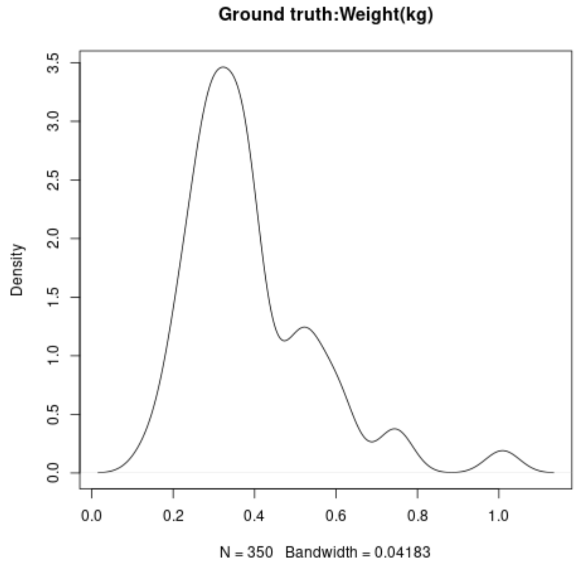


**Fig 1. Distribution of grape cluster weights used for model training and evaluation.**

Figure 2 presents the relationship between observed and predicted grape cluster weights obtained using the three convolutional neural network (CNN) architectures evaluated in this study: ResNet18, EfficientNetV2, and MobileNetV3. The dashed red line represents the ideal prediction scenario (1:1 relationship), while the coefficient of determination ($R^2$) quantifies the proportion of variance in cluster weight explained by each model.

Overall, all CNN architectures demonstrated a satisfactory ability to estimate grape cluster weight directly from images, with $R^2$ values ranging from 0.719 to 0.776. The results indicate that image-based deep learning models can capture a substantial portion of the variability associated with grape cluster mass, supporting their potential application for non-destructive yield estimation in vineyards.

Among the evaluated architectures, ResNet18 achieved the highest predictive performance, with an $R^2$ = 0.776, r = 0.8919, MAE = 0.06 g and RMSE = 0.08 g. The scatterplot reveals a relatively strong alignment between predicted and observed values across the evaluated weight range. Most observations are distributed close to the 1:1 reference line, indicating good agreement between predictions and measurements. Nevertheless, some deviations are evident, particularly among heavier clusters, where the model tends to underestimate the highest observed weights. A few overestimated samples are also observed in the intermediate-to-high weight range, suggesting that prediction uncertainty increases as cluster weight increases.

MobileNetV3 exhibited performance comparable to ResNet18, achieving an $R^2$ = 0.761, r = 0.8924, MAE = 0.06 g, and RMSE = 0.08 g. Despite its considerably lighter architecture and lower computational complexity, the model maintained strong predictive capability. The distribution of points indicates good correspondence between measured and predicted values for low- and medium-weight clusters. Similar to ResNet18, however, the model tended to underestimate the largest clusters, as evidenced by observations located below the 1:1 line at the upper end of the weight distribution. This behavior is common in regression models trained on datasets with relatively few extreme observations, where predictions are naturally biased toward the central tendency of the training data.

In contrast, EfficientNetV2 produced the lowest predictive accuracy, with an $R^2$ = 0.719, r = 0.8494, MAE = 0.06 g, and RMSE = 0.09 g. Although the model successfully captured the overall trend between image features and cluster weight, a greater dispersion of points around the ideal line was observed. Prediction errors were particularly noticeable for medium- and high-weight clusters, indicating reduced robustness compared with the other architectures. This result may be associated with the relatively limited size of the dataset, which can restrict the ability of more complex architectures to fully exploit their representational capacity.

A common pattern across all models is the greater concentration of samples in the lower weight range (approximately 0.20–0.50 kg), where prediction accuracy appears higher. As cluster weight increases, the spread of residuals becomes larger, suggesting heteroscedastic behavior. This phenomenon may reflect increased visual variability among larger grape clusters, including differences in compactness, berry density, occlusion effects, and cluster morphology that are not fully captured by RGB imagery alone.

The relatively small differences in $R^2$ among the three architectures suggest that model complexity was not the primary determinant of predictive performance in this application. In fact, the simpler ResNet18 and MobileNetV3 architectures slightly outperformed EfficientNetV2, indicating that lightweight CNNs may offer a favorable balance between accuracy and computational efficiency for vineyard deployment. This finding is particularly relevant for real-time agricultural applications, where inference speed and hardware limitations are often critical constraints.

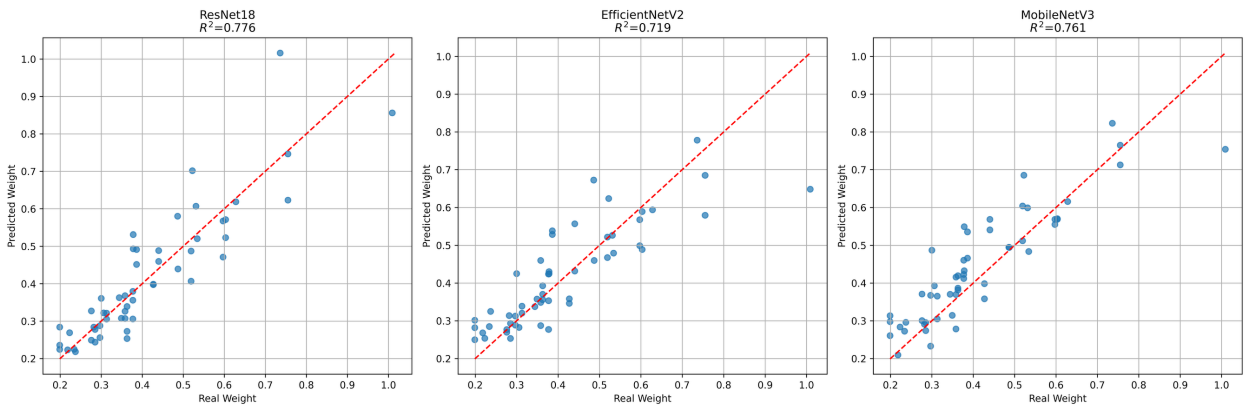


**Fig 2. Relationship between observed and predicted grape cluster weights obtained using ResNet18 ($R^2$ = 0.776), EfficientNetV2 ($R^2$ = 0.719), and MobileNetV3 ($R^2$ = 0.761). The red dashed line represents the ideal 1:1 relationship between measured and predicted values.**

Table 1 presents the predictive performance of the evaluated convolutional neural network architectures for grape cluster weight estimation. All models achieved satisfactory results, with coefficients of determination ($R^2$) ranging from 0.719 to 0.776 and Pearson correlation coefficients (r) above 0.84, demonstrating the feasibility of using image-based deep learning approaches for non-destructive yield assessment. ResNet18 achieved the highest predictive accuracy ($R^2$ = 0.7764), followed closely by MobileNetV3 ($R^2$ = 0.7610), whereas EfficientNetV2 showed slightly lower performance ($R^2$ = 0.7189). Although all architectures yielded the same mean absolute error (MAE = 0.06 kg), ResNet18 and MobileNetV3 exhibited lower RMSE values (0.08 kg) than EfficientNetV2 (0.09 kg), indicating greater robustness in handling larger prediction errors. The relatively small differences among models suggest that increasing architectural complexity does not necessarily translate into superior predictive performance for this application. Notably, MobileNetV3 achieved performance comparable to ResNet18 while requiring substantially fewer computational resources, highlighting its potential for deployment in edge computing environments, mobile applications, and smart sensing platforms. These findings are particularly relevant for precision viticulture, where rapid, non-destructive, and scalable estimation of grape cluster weight can support yield forecasting, harvest planning, labor allocation, and decision-making throughout the production cycle. The results further demonstrate that lightweight CNN architectures may provide an effective balance between prediction accuracy and operational feasibility, facilitating the adoption of artificial intelligence technologies in commercial vineyard management.

**Table 1. Predictive performance of the evaluated convolutional neural network architectures for grape cluster weight estimation, expressed by the coefficient of determination ($R^2$), Pearson correlation coefficient (r), mean absolute error (MAE), and root mean square error (RMSE).**

| Architecture | $R^2$ | r | MAE | RMSE | Model Size |
|---|---|---|---|---|---|
| ResNet18 | 0.7764 | 0.8919 | 0.06 kg | 0.08 kg | 42.7Mb |
| EfficientNetV2 | 0.7189 | 0.8494 | 0.06 kg | 0.09 kg | 77.8Mb |
| MobileNetV3 | 0.7610 | 0.8924 | 0.06 kg | 0.08 kg | 16.2Mb |

From a practical perspective, the obtained results demonstrate the feasibility of using deep learning-based computer vision systems for grape cluster weight estimation in the emerging viticultural regions of southern Brazil. The achieved prediction accuracies are sufficiently promising to support applications such as yield forecasting, harvest planning, and vineyard management. Future studies should investigate larger and more diverse datasets, additional grape cultivars, multi-season observations, and multimodal approaches integrating RGB imagery with depth, hyperspectral, or LiDAR information to further improve prediction accuracy and model generalization.

## Conclusion

This study demonstrated the potential of deep learning-based computer vision techniques for non-destructive grape cluster weight estimation in the emerging wine-producing region of the Brazilian Pampa biome. Three convolutional neural network architectures (ResNet18, EfficientNetV2, and MobileNetV3) were evaluated using RGB images of grape clusters, achieving satisfactory predictive performance across all tested models.

Among the evaluated architectures, ResNet18 achieved the highest predictive accuracy ($R^2$ = 0.7764), while MobileNetV3 delivered comparable performance ($R^2$ = 0.7610) with lower computational complexity, highlighting its suitability for resource-constrained environments. EfficientNetV2 also demonstrated promising results, although with slightly lower predictive performance. The relatively small differences among the models suggest that lightweight architectures can effectively capture the visual features associated with grape cluster weight without requiring highly complex network structures.

The findings confirm the feasibility of employing image-based deep learning models as decision-support tools for precision viticulture. Accurate estimation of grape cluster weight can contribute to yield forecasting, harvest scheduling, labor management, and logistical planning, supporting more efficient and data-driven vineyard management practices. Furthermore, the use of RGB imagery offers a low-cost and scalable alternative to traditional destructive sampling methods.

Future research should focus on expanding the dataset to include multiple growing seasons, cultivars, and environmental conditions to improve model robustness and generalization. The integration of additional sensing modalities, such as depth, hyperspectral, or LiDAR data, may further enhance prediction accuracy. In addition, the deployment of lightweight architectures such as MobileNetV3 on mobile and edge-computing platforms represents a promising direction for the development of real-time vineyard monitoring systems.

### Acknowledgments

This work has been partially funded by the project VisAgro, supported by Competence Center for Digital Agriculture Hardware, with financial resources from the PPI IoT/Manufatura 4.0 / PPI HardwareBR of the MCTI grant number 056/2023, signed with EMBRAPII.

## References

Ayaz, M., Ammad-Uddin, M., Sharif, Z., Mansour, A., & Aggoune, E. H. M. (2022). Internet-of-Things (IoT)-based smart agriculture: Toward making the fields talk. IEEE Access, 10, 129551–129580. https://doi.org/10.1109/ACCESS.2022.3224155

Barbole, D. K., & Jadhav, P. M. (2023). GrapesNet: Indian RGB & RGB-D vineyard image datasets for deep learning applications. Data in Brief, 48, Artigo 109100. https://doi.org/10.1016/j.dib.2023.109100

Jabed, M. A., & Murad, M. A. A. (2024). Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. Heliyon, 10(24), e40836. https://doi.org/10.1016/j.heliyon.2024.e40836

Koirala, A., Walsh, K. B., Wang, Z., & McCarthy, C. (2023). MobileNet-based lightweight deep learning models for real-time crop monitoring applications. Smart Agricultural Technology, 5, 100325. https://doi.org/10.1016/j.atech.2023.100325

Mohimont, L., Alin, F., Rondeau, M., Vaillant-Gaveau, N., & Dumur, D. (2022). Computer vision and deep learning for precision viticulture. Agronomy, 12(10), 2463. https://doi.org/10.3390/agronomy12102463

Nguyen, T. T., Pham, X. D., Dang, L. M., Nguyen, T. T., & Duong, T. Q. (2023). Edge-AI-enabled real-time crop disease detection using lightweight convolutional neural networks on embedded devices. Smart Agricultural Technology, 4, 100222. https://doi.org/10.1016/j.atech.2023.100222

Orlandi, G., Matese, A., Ulrici, A., Calvini, R., Berton, A., & Di Gennaro, S. F. (2025). Automated yield prediction in vineyard using RGB images acquired by a UAV prototype platform. OENO One, 59(1), 1–20. https://doi.org/10.20870/oeno-one.2025.59.1.8133

Olenskyj, A. G., Sams, B. S., Fei, Z., Singh, V., Raja, P. V., Bornhorst, G. M., & Earles, J. M. (2022). End-to-end deep learning for directly estimating grape yield from ground-based imagery. Computers and Electronics in Agriculture, 198, 107081. https://doi.org/10.1016/j.compag.2022.107081

Talaviya, T., Shah, D., Patel, N., Yagnik, H., & Shah, M. (2024). Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides. Artificial Intelligence in Agriculture, 9, 1–18.

Victorino, G., Braga, R. P., Santos-Victor, J., & Lopes, C. M. (2022). Overcoming the challenge of bunch occlusion by leaves for vineyard yield estimation using image analysis. OENO One, 56(1), 117–131. https://doi.org/10.20870/oeno-one.2022.56.1.4863

Wspaniały, P., Kamiński, P., & Szymański, J. (2024). EfficientNet-based deep learning models for plant disease recognition under field conditions. Computers and Electronics in Agriculture, 226, 109228. https://doi.org/10.1016/j.compag.2024.109228

Zhang, Y., Li, H., Wang, X., & colaboradores. (2024). A precise grape yield prediction method based on a modified DCNN model. Computers and Electronics in Agriculture, 225, 109338. https://doi.org/10.1016/j.compag.2024.109338