Determining the Climatic Drivers for Wine Production in the Côa Region (Portugal) Using a Machine Learning Approach

Helder Fraga 1,*, Teresa R. Freitas 1, Marco Moriondo 2,3, Daniel Molitor 4 and João A. Santos 1

Abstract: The Côa region in inner-northern Portugal heavily relies on viticulture, which is a cornerstone of its economy and cultural identity. Understanding the intricate relationship between climatic variables and wine production (WP) is crucial for adapting management practices to changing climatic conditions. This study employs machine learning (ML), specifically random forest (RF) regression, to predict grapevine yields in the Côa region using high-resolution climate data for 2004–2020. SHAP (SHapley Additive exPlanations) values are used to potentially explain the non-linear relationships between climatic factors and WP. The results reveal a complex interplay between predictors and WP, with precipitation emerging as a key determinant. Higher precipitation levels in April positively impact WP by replenishing soil moisture ahead of flowering, while elevated precipitation and humidity levels in August have a negative effect, possibly due to late-season heavy rainfall damaging grapes or creating more favorable conditions for fungal pathogens. Moreover, warmer temperatures during the growing season and adequate solar radiation in winter months favor higher WP. However, excessive radiation during advanced growth stages can lead to negative effects, such as sunburn. This study underscores the importance of tailoring viticultural strategies to local climatic conditions and employing advanced analytical techniques such as SHAP values to interpret ML model predictions effectively. Furthermore, the research highlights the potential of ML models in climate change risk reduction associated with viticulture, specifically WP. By leveraging insights from ML and interpretability techniques, policymakers and stakeholders can develop adaptive strategies to safeguard viticultural livelihoods and stable WP in a changing climate, particularly in regions with a rich agrarian heritage, such as the Côa region.

Keywords: random forest; machine learning; winemaking; forecast; productivity; viticulture; SHAP values

1. Introduction

In inner-northern Portugal, the Côa region boasts a rich agrarian heritage with centuries-old traditions. Agriculture plays a pivotal role in the economic, social, and cultural fabric of this region, underpinning its commercial viability and the livelihoods of its inhabitants. Designated as a UNESCO World Heritage Site in 1998 (“Prehistoric Rock Art Sites in the Côa Valley and Siega Verde”), the Côa Valley relies heavily on agriculture as its economic basis. Specifically, grapevines constitute a significant portion, occupying nearly 10% of the total land area. Wine production is a robust source of income for the local population [1]. Additionally, a portion of the Côa region lies within the Douro Demarcated Region,
renowned for its Protected Designation of Origin status and global recognition as a wine-producing brand, particularly Port wine [2]. Other crops, such as olive trees, almond trees, chestnut trees, and grasslands, also contribute to the agricultural sustenance of this region.

It is widely acknowledged that climatic factors exert a significant influence on grapevine physiology, phenology, and productivity, profoundly shaping the agricultural suitability of a given winemaking region [3,4]. Considering the unique climatic conditions in the area, characterized by warm, dry summers and wet, moderately cold winters typical of Mediterranean-like climates, grapevines tend to thrive in this environment [5]. The complex interplay of climatic factors on grape yields remains an area of ongoing research with numerous challenges and uncertainties [6]. Variability in climate conditions, interactions among plant–soil–atmosphere continuum, and the long-term effects of climate change all contribute to the complex nature of this issue [7]. As our understanding evolves, it becomes increasingly crucial to address these uncertainties and improve predictive models to effectively adapt agricultural practices to a changing climate and ensure global food security [8]. Additionally, under climate change, the potential for the adoption of adaptation and mitigation strategies requires comprehending the impact of climatic factors on crop productivity [9].

Several studies have developed models aiming at predicting grapevine yield or wine production [10,11]. While these traditional methods have provided valuable insights, they are not without limitations, such as relying on linear relationships between predictors and predictands, which does not always apply. Machine learning (ML) holds significant promise in enhancing our understanding of how climatic factors affect crop parameters [12], including those of grapevine quality [2]. By analyzing vast datasets encompassing historical climate records and grapevine yields, ML algorithms can unveil intricate patterns and relationships that may elude traditional statistical approaches [2]. Moreover, these algorithms discriminate complex interactions among predictors, such as climatic factors, providing valuable insights into the multifaceted nature of crop responses to changing environmental conditions [13]. Additionally, ML can facilitate the development of more accurate predictive models for assessing the potential impacts of climate change on agriculture, enabling farmers and policymakers to make informed decisions and implement adaptive strategies [14,15]. Furthermore, ML can aid in the optimization of agricultural practices by offering real-time monitoring and decision support, thereby enhancing crop resilience and sustainability in the face of climatic uncertainties [16]. As machine learning continues to advance, it holds the potential to be an indispensable tool in mitigating the challenges posed by climate change to global food production.

Nonetheless, ML models, particularly the complex non-linear ones, often operate as “black boxes”, meaning that they produce predictions without providing clear insight into how those predictions are generated [17]. This difficulty arises because these models process complex interactions between predictors, often involving numerous variables and intricate mathematical transformations. While such models excel at capturing intricate patterns in data, their inner workings can be challenging to interpret for humans, hindering our understanding of the relationship between predictors and outcomes and limiting their acceptance by practical users. Consequently, discerning the precise impact of individual predictors becomes difficult, limiting our ability to extract actionable insights from the model’s output [18]. While certain ML algorithms do offer a feature importance output, they often present a simplistic view, failing to reveal any interactions among features. This limitation can hinder a comprehensive understanding of how different variables influence each other within the model, potentially overlooking nuanced relationships that could significantly impact the prediction accuracy, uncover the deeper connections between features, and enhance the interpretability and predictive power of ML models. SHAP (SHapley Additive exPlanations) values offer a promising solution to the interpretability challenge posed by ML models [19]. Leveraging principles from cooperative game theory, SHAP values assign each predictor variable a contribution to the model’s output for a given forecast [19]. These values quantify the impact of each predictor on the model’s
predictions consistently and understandably. By examining SHAP values, analysts can discern the relative importance of different predictors, understand how each influences individual predictions, and identify complex interactions between variables, empowering modelers to make informed decisions based on a deeper understanding of the underlying data patterns [19].

The utilization of high-resolution spatial datasets further enhances the significance of the application of these prediction algorithms. For capturing the environmental variability inherent to regions undergoing environmental and agricultural transformations [20], it is necessary to provide climatic data that is the closest to the location of interest [20]. Nonetheless, the scarcity of weather stations is a limitation, and gridded datasets provide a comprehensive and spatially detailed view of environmental variables [21]. However, these datasets often lack the necessary detailed spatial resolution, particularly in regions marked by steep climatic gradients, such as mountainous terrains, where low-resolution climatic datasets fail to adequately represent crop microclimatic conditions. In the Côa Valley, most vineyards are subject to factors like elevation, slope, and aspect due to the region’s complex topography. Therefore, micro-meteorological effects yield a significant influence on grape growth [22]. Consequently, the diverse microclimatic conditions must be considered for producing dependable prediction model outcomes. In this context, the adoption of high-resolution downscaled data has proven instrumental in facilitating regional and local-scale assessments and promoting seamless integration with machine learning models [23].

The current study evaluates how an ML model can help predict grapevine yield in the Côa region, hinting at the most important climatic variables for this assessment. To achieve this, a random forest (RF) ensemble algorithm was trained [24], tuned, and tested in conjunction with state-of-the-art high-resolution climate datasets. Subsequently, the trained models were used to predict the grapevine yields in the Côa sub-regions. SHAP values were then used to assess the impact of each climatic variable, further enhancing our understanding of how climate affects grapevine productivity.

2. Materials and Methods

2.1. Study Area and Wine Production Data

The Côa region is located in inner-northern Portugal (centered at 40.95° N, −7.57° W), at the border with Spain (Figure 1a). It is a region with a rich viticultural heritage, characterized by steep slopes and terraced vineyards ranging from 100 to 600 m. Viticulture is the main crop, producing exceptional wines recognized globally for their unique terroir and centuries-old winemaking traditions. The main training systems used are cordon and guyot, while the main varieties are Touriga-Nacional, Touriga-Franca, and Tinta-Roriz (syn. Tempranillo), which are red varieties. Harvest is usually in September [25] and is mostly done by hand. The predominant soil types are schist and granite-based. It includes 12 municipalities (henceforth sub-regions) covering completely the Côa river basin, and partially the Douro and Mondego rivers. To assess the grapevine land cover in the Côa region, the COS2018 dataset was used. The region has approximately $28 \times 10^3$ ha of vineyards [26], mostly concentrated in the northern part of this region, being also areas within the Douro Superior winemaking region. The sub-regions of Freixo de Espada à Cinta (FEC), Torre de Moncorvo (TDM), Vila Nova de Foz Côa (VNFC), Mêda (MD), Figueira de Castelo Rodrigo (FCR), Trancoso (TC), and Pinhel (PN) have the largest vineyard areas. South of the PN region, which is an area of higher elevations, few vineyards are found (Figure 1b). Wine production (WP) data, which include all wine types produced from 2004 to 2020, were obtained from the Portuguese “Instituto da Vinha e do Vinho” and analyzed. Therefore, the data used as target variables consisted of 17 years of WP data for each of the 12 sub-regions in Côa, resulting in 204 values.
Figure 1. (a) Location of the Côa Region within mainland Portugal. (b) Wine production of the Municipalities in the Côa Region, along with the vineyard land cover.

2.2. Climate Data

The use of downscaled climate data with high resolution is of paramount importance as it enables the accurate representation of local microclimatic conditions, especially in regions characterized by topographical complexities, ensuring precision in assessing the impacts of climate change on agriculture and facilitating the development of effective adaptation strategies. Herein, monthly maximum air temperature (TX; °C), minimum air temperature (TN; °C), solar radiation (QQ; W/m²), total precipitation (RR; mm), and relative humidity (HU; %) for the months of January to September (Table 1), were used as predictors in the ML model. As grape harvest in the Côa is usually conducted in September [25], these months were selected taking into account the growing season, including part of the dormancy period. All these variables were obtained from the E-OBS dataset [27] for the historical period (2004–2020) at a ~10 km spatial resolution. However, this spatial resolution may be insufficient for adequately depicting climatic variations within the vineyards of a given region. Therefore, the data underwent a downscaling methodology to a 1 km spatial resolution. This process utilized a multivariate linear regression approach, with latitude, longitude, and elevation serving as independent variables [28]. This approach is similar to well-known databases, such as WorldClim [29] or Chelsa [30]. The monthly climatic data was then obtained for the 12 sub-regions in Côa, and the spatial medium value was used as a representative of each region.
Table 1. List of monthly predictor variables including their acronym.

<table>
<thead>
<tr>
<th>Month</th>
<th>Minimum Air Temperature (°C)</th>
<th>Mean Air Temperature (°C)</th>
<th>Maximum Air Temperature (°C)</th>
<th>Total Precipitation (mm)</th>
<th>Relative Humidity (%)</th>
<th>Solar Radiation (W/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>TN_01</td>
<td>TG_01</td>
<td>TX_01</td>
<td>RR_01</td>
<td>HU_01</td>
<td>QQ_01</td>
</tr>
<tr>
<td>February</td>
<td>TN_02</td>
<td>TG_02</td>
<td>TX_02</td>
<td>RR_02</td>
<td>HU_02</td>
<td>QQ_02</td>
</tr>
<tr>
<td>March</td>
<td>TN_03</td>
<td>TG_03</td>
<td>TX_03</td>
<td>RR_03</td>
<td>HU_03</td>
<td>QQ_03</td>
</tr>
<tr>
<td>April</td>
<td>TN_04</td>
<td>TG_04</td>
<td>TX_04</td>
<td>RR_04</td>
<td>HU_04</td>
<td>QQ_04</td>
</tr>
<tr>
<td>May</td>
<td>TN_05</td>
<td>TG_05</td>
<td>TX_05</td>
<td>RR_05</td>
<td>HU_05</td>
<td>QQ_05</td>
</tr>
<tr>
<td>June</td>
<td>TN_06</td>
<td>TG_06</td>
<td>TX_06</td>
<td>RR_06</td>
<td>HU_06</td>
<td>QQ_06</td>
</tr>
<tr>
<td>July</td>
<td>TN_07</td>
<td>TG_07</td>
<td>TX_07</td>
<td>RR_07</td>
<td>HU_07</td>
<td>QQ_07</td>
</tr>
<tr>
<td>August</td>
<td>TN_08</td>
<td>TG_08</td>
<td>TX_08</td>
<td>RR_08</td>
<td>HU_08</td>
<td>QQ_08</td>
</tr>
<tr>
<td>September</td>
<td>TN_09</td>
<td>TG_09</td>
<td>TX_09</td>
<td>RR_09</td>
<td>HU_09</td>
<td>QQ_09</td>
</tr>
</tbody>
</table>

2.3. Machine Learning Model Runs

In the current study, we employed a robust ensemble learning technique, RF regression [24], to evaluate the feature importance in the empirical modeling of grapevine yield, using the abovementioned monthly climate variables as predictors. The Scikit-learn Python library was used for this approach [31]. RF is a popular ensemble learning technique in machine learning due to its ability to handle large datasets with high dimensionality. The model works by constructing several decision trees with randomized subsets of features and training samples, therefore mitigating overfitting and enhancing robustness [24]. Subsequently, each decision tree independently makes predictions, and the final prediction is determined by aggregating the results.

All data were pre-processed by normalization and detrending for each specific sub-region. Normalization and detrending are standard pre-processing techniques often applied when modeling datasets [32]. Normalization ensures that all variables are on a comparable scale (e.g., higher production in some regions than others), which is crucial, as it prevents variables with larger scales from dominating the analysis. Detrending is commonly used to remove trends or patterns in the data that are not relevant to the relationship being studied (e.g., an increase in production due to an increase in vineyard area), such as seasonality or long-term trends, which can introduce bias and confound the analysis. A similar approach was used in several other studies, e.g., in Fraga et al. [33]. Following these pre-processing steps, we employed a modeling approach that involved training the RF model using multiple train–test splits, using a standard 80–20 train–test split, ensuring that 80% of the data were utilized for training and 20% for testing. This was replicated 1000 times with different train–test splits (random initialization seed). This partitioning scheme allows for robust model assessment while maintaining a balance between training and testing data for reliable performance evaluation across different regions. This methodology aims to assess the robustness and generalization performance of the model by repeatedly partitioning the dataset into separate training and testing subsets, and we can obtain more reliable estimates of its effectiveness across various data distributions and scenarios. This approach helps mitigate the potential bias introduced by a single train–test split and provides a more comprehensive evaluation of the model’s predictive capabilities.

2.4. Predictor Influence on Model Outcomes

Analysing SHAP (SHapley Additive exPlanations) values for an RF regression model provides crucial insights into the impact of each feature on individual predictions. SHAP values quantify the contribution of each feature to the model’s output for a particular instance. By examining these values, we can understand the relative importance of different features and how they influence model predictions. This enables us to interpret the model’s decision-making process more effectively and identify which features have the most significant influence on its output. Furthermore, SHAP values can help identify potential interactions between these, offering deeper insights into the underlying relation-
ships within the data. Moreover, this analysis enhances our understanding of the model’s behavior and can inform decisions related to feature engineering, model refinement, and problem understanding.

To integrate SHAP values with the RF model, we employed the SHAP Python library [19], a widely used tool for computing SHAP values for tree-based models. After training the RF model on the training dataset, we proceeded to generate SHAP values. This process involved passing the trained RF model and the testing dataset to the SHAP library, which internally performed the necessary computations to derive SHAP values for each instance in the testing dataset. Once computed, the SHAP values provided insights into the contribution of each feature to the model’s predictions. These values encapsulate the impact of individual features on the predicted outcome, facilitating a deeper understanding of the model’s decision-making process.

3. Results
3.1. Wine Production in the Côa Region

The WP data show a fluctuation over the 17 years in the analysis (Figure 2). The overall time series shows a decreasing trend of $-8 \times 10^3$ hl yr$^{-1}$. The lowest year recorded was 2008, with a large decrease from the previous year, which was mostly attributed to atypical mildew problems. After that, steady increases in production are visible until 2017. Table 2 provides an overview of the average WP ($\times 10^3$ hl) and production trends over the period 2004–2020, as well as the vineyard area (ha) for the municipalities in the Côa region. Notably, there is substantial variation in average production, trends, and vineyard area across the municipalities. Nonetheless, this should not impact our modeling approach since all data were pre-processed. FEC and TM exhibit relatively low average production levels, with 2471 and $2417 \times 10^3$ hl, respectively, followed by VNFC and MD with $9157$ and $8076 \times 10^3$ hl of average production, respectively. However, these last municipalities also experienced notable negative production trends over the period, indicating a decline in production levels. In contrast, FCR, TC, and PN demonstrate substantial high production levels, but also with pronounced negative trends, suggesting constraints. Some municipalities like Almeida, Celorico da Beira, Guarda, Sabugal, and Penamacor report average production levels below 1000 thousand hl, reflecting potentially smaller-scale production activities. In general, the data highlight the diverse landscape of WP in the Côa region, with varying levels of production and trends observed across different municipalities.

Table 2. Average wine production ($\times 10^3$ hl) and trend ($\times 10^3$ hl) from 2004 to 2020, along with the vineyard area (ha) in each of the municipalities within the Côa region.
Figure 2. Wine production (WP) data on the Côa region (upper left chronogram with linear regression trend line, Lt) from 2004 to 2020, along with the corresponding WP empirical distributions for the seven outlined sub-regions (box plots). The central box represents the interquartile range (IQR), which spans from the first quartile (Percentile-25) to the third quartile (Percentile-75). The horizontal line inside the box signifies the median value (Percentile-50) of wine production for each sub-region. The whiskers extending from the box indicate the range of the data, excluding outliers (Percentile-10 and 90).

3.2. Climatic Characteristics

An analysis of the climatic factors used in the modeling assessment for an annual period in the Côa region reveals a distinct north–south gradient, with the cooler values concentrated in the north. TN ranges from potentially as low as 5.1 °C in the north to 9.1 °C in the south (Figure 3a). Similarly, TG temperatures exhibit a spread between 10.1 °C and 16.1 °C (Figure 3b). This pattern is also shown in TX, which may exceed 22.1 °C in the south and fall below 16.1 °C in the north (Figure 3c). Conversely, RR demonstrates a northward and inward decrease in precipitation, potentially reaching as low as 380 mm, while in the center-west, namely in Guarda, RR reaches values of 850 mm (Figure 3d). This suggests that the innermost regions in Côa present dryer climatic characteristics, while the southern regions present hotter climates. HU appears to be a relatively similar pattern to RR, with variation in both the east–west and north–south directions, likely ranging from 65% to 74%. QQ also presents a similar pattern to those of the temperatures, with higher values exceeding 190 W/m² concentrated in the southern areas, while the north may receive around 177 W/m². This shows that the Côa region, although a relatively small region in Portugal, exhibits significant spatial variability.
southern regions present hotter climates. HU appears to be a relatively similar pattern to RR, with variation in both the east–west and north–south directions, likely ranging from 65% to 74%. QQ also presents a similar pattern to those of the temperatures, with higher values exceeding 190 W/m² concentrated in the southern areas, while the north may receive around 177 W/m². This shows that the Côa region, although a relatively small region in Portugal, exhibits significant spatial variability.

Figure 3. (a) Annual minimum air temperature (TN; °C), (b) annual mean air temperature (TG; °C), (c) annual maximum air temperature (TX; °C), (d) annual precipitation sum (RR; mm), (e) relative humidity (HU; %) and (f) solar radiation (QQ; W/m²). Calculated from 2004 to 2020 from the E-OBS dataset.

3.3. Output Metrics

The results from the RF 1000 runs show slight variability (Figure 4a) with a coefficient of determination (R²) ranging from 0.5 to 0.95 (minimum–maximum). The median R² value (P50) across all runs was 0.75, with the 25th percentile (P25) at 0.68 and the 75th percentile (P75) at 0.79. This indicates a relatively short range of predictive performance across the 1000 different train–test splits, with a significant portion of runs achieving strong R² values with high explanatory power. These findings emphasize the effectiveness of the RF approach in capturing and explaining the underlying patterns in the data across various regions, thereby contributing to a more nuanced understanding of the predictive modeling outcomes. Regarding the associated errors, the MAPE (Mean Absolute Percentage Error) obtained a median value of 19% (Figure 4b), which is acceptable. The spread of the residuals is relatively narrow, which suggests that the model is capturing most of the variability in the data. The RF approach not only demonstrates consistent performance but also
provides valuable insights into the nuances of the dataset, thus enhancing the overall predictive modeling process. This reinforces its utility as a powerful tool for uncovering and interpreting complex relationships within the data.

![Figure 4. Boxplots representing the (a) coefficient of determination (R-squared) and (b) the mean absolute percentage error (MAPE) over the 1000 model runs.](image)

3.4. Significance of Climatic Variables

Figure 5 allows an inspection of the impact of different features on model predictions. The features are listed on the left side of the plot, sorted from the most important features at the top. The SHAP value of a feature for a particular data point represents the contribution of that feature to the model’s prediction for that data point. Positive SHAP values indicate that the feature value increased the model’s prediction, while negative SHAP values indicate that the feature value decreased the model’s prediction. The color of each dot on the plot corresponds to the feature value, with blue indicating a low feature value (e.g., low levels of precipitation) and red indicating a high feature value (e.g., high levels of precipitation). In this case, the most important features are RR_04, RR_08, and HU_08, either positively or negatively.

RR_04, the most important feature, also indicates that higher precipitation values in April tend to result in higher WP. In effect, precipitation during April may allow replenishment of soil water reserves in the period preceding flowering, which in the region typically occurs in May. Conversely, higher precipitation and humidity levels in August (RR_08 and HU_08) have the opposite effect. This could be easily associated with heavy rains close to the harvest season. August is a dry month that usually precedes harvest, and high levels of precipitation can perhaps damage the bunches and grapes, resulting in lower yields. This variable might capture late-season rainfall, which could impact grape maturation and yield. Furthermore, higher humidity levels may decrease yields due to the spread of fungal diseases such as Botrytis bunch rot. The next features, RR_06, RR_06, and RR_09, play uncertain roles. Nonetheless, this highlights the role of precipitation in WP volume, although the negative vs. positive effect is not clear. Subsequently, higher TX_08 and TG_02 seem to favor production. TN_02 may indeed reflect winter temperatures, which could influence vine dormancy and subsequent growth cycles. The climatic variables are difficult to analyze, still, the significance of these variables highlights their strong influence on model predictive performance, emphasizing the importance of climate in understanding and forecasting grapevine yield and WP. Overall, the SHAP summary plot is a useful tool for understanding how a machine learning model makes its predictions.
4. Discussion

The current study presents a comprehensive analysis of the mechanisms underlining WP in the Côa region of northern Portugal, considering the crucial role of climatic variables. The findings address the challenges and opportunities posed by machine learning (ML) algorithms and explanatory analysis, as well as the importance of high-resolution spatial datasets in enhancing predictive accuracy. The observed fluctuations in WP over the 17-year period underscore the dynamic nature of viticulture systems in each sub-region of Côa, and the multifaceted factors influencing grapevine productivity. Despite variations in production levels across these different municipalities, a general decreasing trend in WP is
evident, emphasizing the need for robust predictive models to navigate the complexities of viticultural yield forecasting.

The climatic characteristics of this region exhibit significant spatial variability, with distinct north–south (cooler–warmer) gradients in temperature and west–east (higher–lower) precipitation. These precipitation differences are shaped by the Atlantic influence and distance to the coast. The western areas experience higher precipitation levels due to the proximity to Atlantic moisture-laden air masses, while the eastern regions encounter comparatively lower precipitation, resulting in a distinct west–east precipitation gradient. Such variability underscores the importance of high-resolution climate data in capturing local microclimatic conditions, particularly in regions with diverse topographical features.

The application of RF regression, coupled with SHAP value analysis, offers valuable insights into the relationship between climatic variables and WP in this specific region. The robustness of the RF approach is evident from its consistent performance across multiple train–test splits, with high coefficients of determination ($R^2$) and relatively low bias (MAPE), indicating strong predictive power. The use of SHAP values elucidates the relative importance of different features. Nonetheless, it should be noted that these results are based on the non-linear relationship between features and targets over 1000 runs. In each of these runs, the connections established between features can be slightly different, highlighting the need for an integrated approach.

SHAP values show that the precipitation variables (RR_04, RR_08) emerge as key determinants of WP. Herein, higher precipitation levels in April are associated with increased yields, potentially replenishing soil moisture ahead of the flowering period. Throughout critical phenological stages, including budburst and the subsequent development of shoots and inflorescences, water availability within a moderate range becomes a critical determinant for optimal vine growth [34]. Several studies have investigated the effect of precipitation on grapevine yields, indicating that higher rainfall at early stages is beneficial [35], promoting vegetative growth and berry and wine quality benefits [36], which are in agreement with the present study findings. Adequate water availability during this phase enables proper shoot growth and promotes a successful berry set [37]. However, this delicate equilibrium can be easily disrupted by water stress during these early stages, resulting in suppressed shoot development, compromised flower formation, and reduced berry set [38]. In contrast to present observations, Molitor and Keller [11] observed for Riesling in the cool climate viticulture region of Luxembourg, that pre-bloom precipitation sums were negatively correlated with annual yield. We assume that here (Luxembourg) the negative effect of high precipitation sums (usually associated with relatively low radiation and low temperatures) in the periods of flower formation might negatively impact flower sizes and lead to smaller clusters, while this effect might be absent or over-compensated under the warmer climatic conditions were radiation and temperature are usually not the limiting factors for photosynthesis and assimilation.

Conversely, our results indicate that elevated precipitation and humidity levels in August (RR_08 and HU_08) exert a negative impact on yields, highlighting the vulnerability of grapevines to threats such as late-seaõon heavy rainfall, as well as bunch and grape damage. Moreover, humid conditions during ripening may increase the vulnerability of grapes to Botrytis bunch rot, a fungal infection that negatively affects the quality and marketability of the harvest [39]. This is corroborated by the results of Moriondo et al. [40] and Jones and Davis [41], who explored the relationships between meteorological data and grapevine performances in Chianti and Bordeaux wine regions and found that high rainfall rates during veraison limits harvest quality and this is likely due to the onset of fungal disease triggered by high humidity in pre-harvest time. Additionally, during the final stages of the growing season coinciding with berry ripening, studies suggest that drier climatic conditions can contribute to the production of high-quality wines. Under these circumstances, a reduction in leaf development promotes enhanced water use efficiency, enabling the vine to prioritize resource allocation toward berry maturation [42]. Heavy precipitation during the pollen season tends to have negative effects on yield due to inade-
quate pollen dispersal [43], which is somewhat visible in Figure 5, although not entirely clear, due to the obfuscated results of RR_05 and RR_06 found herein. In sum, the intricate interplay between precipitation and grapevine development highlights the importance of meticulous water management practices in viticulture. Achieving a balance between sufficient water availability and periods of water stress is critical for optimizing vine growth, berry maturation, and, ultimately, the production of premium-quality wines [34].

Regarding the temperature features, the present study indicates that higher temperatures during the growing season seem favorable to higher WP. Several studies, however, use linear models as an attempt to capture the intrinsic relationship between temperature and grapevine yields. Some of them have indeed been used in WP/yield modeling for nearby wine regions, such as Douro/Porto [44,45] and Minho [33] in Portugal and Rías Baixas [46] in Spain. Most of these studies found that warm, stable temperatures after flowering are beneficial for WP, which is in agreement with the results found. For other cool climate viticulture regions, such as Luxemburg, Molitor and Keller [11] came to contrasting results studying the yield of Riesling. Here, post-bloom high temperatures were negatively correlated with annual yield, which suggests the key role played by the cultivars in the climate vs. yield relationship.

Our study underscores the importance of solar radiation for WP. Solar radiation is a critical driver of photosynthesis, the process by which plants convert sunlight into energy for growth. Adequate radiation during the winter months enhances photosynthetic activity, which can significantly contribute to improved crop development and, ultimately, higher yields [47]. Molitor and Keller (2016) [11] also observed that relatively high temperatures after mid-winter were correlated with higher yields. Consequently, our findings are consistent with these lines, as higher values of QQ_02 (solar radiation in February) seem to be an important factor for WP. Nonetheless, studies show that excessive solar radiation during the advanced stages of the growing cycle can lead to problems such as leaf and cluster sunburn, decreased bud fertility, and unbalanced wines with undesirable high alcoholic content and low acidity [48], which is also backed up by our study, as higher values of June solar radiation (QQ_06) negatively impact wine production.

Comparing our results to previous studies in diverse geographical contexts, we identify a common thread of specific climatic variables, such as precipitation, temperature, and solar radiation, which have a consistent and significant impact on QP. While the strength and nature of these relationships may vary by region and crop type, the overarching importance of these climatic factors remains clear. However, it is crucial to acknowledge that local agricultural practices, grape varieties, and soil conditions can modulate the effects of these climatic variables. Therefore, tailoring agricultural strategies to the unique climatic conditions and agricultural context of our region remains a priority. These strategies may include optimizing irrigation schedules [49], using shading materials to protect vines from extreme heat [50], applying mulch to conserve soil moisture [51], and adopting precision viticulture tools to monitor and manage vineyard conditions more effectively.

As climate variability continues to pose challenges to viticulture, the development of accurate predictive models becomes indispensable for informing adaptive strategies and mitigating risks. ML techniques commonly surpass linear approaches and offer promising avenues for improving our understanding of climate–crop interactions and enhancing predictive capabilities. While these models adeptly capture intricate data patterns, their “black-box” nature impedes the discernment of individual predictor contributions to outcomes, thereby hindering the extraction of actionable insights. Nevertheless, emerging methodologies such as SHAP values offer a promising solution to analytical workflows, allowing modelers and practitioners to comprehend the non-linear model predictions, thereby engendering confidence and facilitating informed decision-making predicated upon a thorough understanding of underlying data dynamics. The combination of these techniques will contribute valuable information for optimizing grapevine yield predictions and understanding the key drivers behind variations in grapevine productivity, ultimately aiding in informed decision-making for vineyard management and agricultural planning.
Some limitations of the current manuscript should be pointed out. The present study covers the period of 2004–2020, which was selected due to the data availability. This timeframe allowed us to analyze a substantial period, providing a robust basis for our conclusions. However, we acknowledge that this period is relatively limited and may not capture longer-term climate variability or emerging trends beyond 2020. Future research could extend the analysis to include more recent data, which could provide additional insights into the ongoing impacts of climate change on viticulture. Furthermore, while our results are specific to the Côa region of Portugal, we believe they have broader implications for viticulture in regions with similar climatic characteristics. However, the generalizability of our findings to other winemaking areas requires careful consideration. Differences in microclimates, cultivars, soil types, and viticultural practices also significantly influence WP.

Under climate change, the projected future warming and drying in the Côa [5] may bring further challenges to local winemakers, which further highlights the need to use advanced tools to forecast crop production, or in this case, WP. In effect, several studies suggest that the winemaking industry should focus on minimizing the impact of yearly yield fluctuations and optimizing production [52]. This issue is especially relevant in scenarios of future warmer climates [53–56]. The modeling techniques applied herein may be used as a valuable tool for WP stabilization, as providing a WP forecast that may improve future planning within the winemaking sector, bringing efficiency gains to the whole value chain.

5. Conclusions

This study underscores the critical role of climatic factors in shaping WP in the Côa region and demonstrates the efficacy of ML models. The integration of high-resolution spatial datasets and advanced modeling techniques represents a significant step toward enhancing agricultural sustainability and resilience in the face of climate variability. By leveraging insights from machine learning and interpretability techniques, such as SHAP values, modelers can withdraw valuable information about non-linear relationships between climatic factors and WP. The utilization of these ML methodologies not only facilitates the identification of the most crucial climatic factors influencing wine production but also enables policymakers and stakeholders to devise tailored adaptation strategies. By leveraging these technologies, stakeholders can plan and adapt their management practices based on the climatic characteristics of the current year. For instance, growers can implement or optimize irrigation schedules based on forecasts provided by these models. Moreover, these methodologies can aid in developing early warning systems for potential extreme weather events (e.g., prolonged drought), allowing for timely interventions and enhanced resilience across the viticultural landscape. These forecasting systems may allow growers, stakeholders, and decision-makers to make informed choices to safeguard viticultural livelihoods and ensure the future sustainability of the sector under a changing climate.


Funding: Authors thank the Portuguese Foundation for Science and Technology (FCT) for UIDB/04033/2020 (https://doi.org/10.54499/UIDB/04033/2020), LA/P/0126/2020 (https://doi.org/10.54499/LA/P/0126/2020) and 2022.02317.CEECIND (https://doi.org/10.54499/2022.02317.CEECIND/CP1749/CT0002). Authors also acknowledge the CoaClimateRisk FCT project (COA/CAC/0030/2019), MONTEVITIS HORIZON project (GA: 101059461), and the WaterQB FCT project (2022.04553.PTDC).

Data Availability Statement: All data used are publicly available.

Acknowledgments: Authors thank the Portuguese Foundation for Science and Technology (FCT) for UIDB/04033/2020 (https://doi.org/10.54499/UIDB/04033/2020), LA/P/0126/2020 (https://
doi.org/10.54499/LA/P/0126/2020) and 2022.02317.CEECIND (https://doi.org/10.54499/2022.02317.CEECIND/CP1749/CT0002). Authors also acknowledge the CoaClimateRisk FCT project (COA/CAC/0030/2019), MONTEVITIS HORIZON project (GA: 101059461), and the WaterQB FCT project (2022.04553.PTDC).

**Conflicts of Interest:** The authors declare no conflicts of interest.

**References**

2. Fraga, H.; Guimaraes, N.; Santos, J. Vintage Port prediction and climate change scenarios. *OENE One* **2023**, *57*, 1–6. [CrossRef]


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.