The timing of cherry blossom bloom is influenced by climatic conditions, particularly temperature accumulation over time. Understanding and predicting bloom dates provide insights into phenological shifts due to climate change. This study employs a hybrid modeling approach, integrating a boosted tree model with temperature-derived growing degree days (GDD) to predict the 2025 bloom dates for multiple locations.

The primary predictor variables include latitude, longitude, altitude, and historical bloom dates. These geographic and climatic features capture site-specific variations that influence bloom timing. Temperature data, specifically daily maximum and minimum temperatures, is used to compute GDD, a widely accepted metric for plant development. GDD is accumulated until a threshold is met, indicating sufficient thermal exposure for blooming. The inclusion of GDD is biologically justified, as temperature is a dominant factor in floral bud development and dormancy break.

A boosted tree model (XGBoost) was selected for its ability to handle non-linearity and interactions among predictors while providing high predictive accuracy. This model was trained on historical bloom dates using latitude, longitude, altitude, and year as features. Predictions for 2025 were generated using the most recent year's geographic and climatic data. However, for locations where temperature data was limited, an alternative method based on GDD accumulation was used. The threshold for bloom initiation was estimated based on past observations, allowing for prediction without requiring direct historical bloom dates. This dual-method approach ensures robustness across varying data availability.

Cherry trees require a period of chilling to break dormancy, followed by sufficient heat accumulation to induce flowering. GDD captures this heat accumulation effectively, making it an essential factor in bloom prediction. The boosted tree model accounts for non-linearities in climatic influences and regional differences in bloom responses. The combination of these methods allows for a more comprehensive prediction strategy, ensuring applicability across diverse climatic conditions.

The model successfully predicts bloom dates for multiple locations, demonstrating the effectiveness of integrating machine learning with biological principles. Locations such as Liestal and Kyoto, which lack complete GDD data, were predicted exclusively through the boosted tree model, while others used a threshold-based GDD approach. This study highlights the potential for machine learning in phenological forecasting and provides a scalable framework for predicting bloom dates under varying climate conditions. Future work could integrate additional climatic variables, such as precipitation and extreme temperature events, to further refine predictions.