Optimizing LED Use in Greenhouse Tomato Cultivation via Plant Biosignals

Master Thesis

Author: Frenk Theodorus Petrus Kester
University: Erasmus University Rotterdam
Faculty: Rotterdam School of Management
Degree Program: Master of Science in Business Information Management
Date of Submission: 24-5-2025
Supervisor: Dr. Yashar Ghiassi-Farrokhfal
Co-reader: Anubhav Goel

1. Introduction

Greenhouse horticulture today has a fundamental contribution to making food production more sustainable, especially in highly populated areas like the Netherlands. These technologically advanced systems enable all-year-round cultivation of crops like tomatoes by providing closely controlled environments that protect plants from harsh weather and enhance growth conditions. But controlling in this way comes at a considerable energy price. The bulk of energy usage in greenhouses involves heating, ventilation, and artificial light (Engler & Krarti, 2021).

Over the past several years, increasing energy costs, climate policy, and consumer pressure to have sustainably produced fruits and vegetables have put pressure on greenhouse operators to maximize energy efficiency without compromising crop yields or quality (Iddio et al., 2020). Consequently, precise horticulture, by using digital sensors, automation and data analysis, is on the rise. Two metrics are central to this trend: energy efficiency and crop performance (Hosseinzadeh & Garcia, 2024).

Most greenhouse horticulture today is controlled by environmental sensor thresholds, such as temperature or CO_2 levels, instead of plant feedback. In avoiding risks such as light underruns or cooling lag, growers tend to apply light or heating conservatively, resulting in inefficiencies in energy consumption (Heuvelink et al., 2025). The absence of real-time plant-actuated feedback poses a significant blind spot.

One innovation that has been seen to meet this challenge is plant electrophysiological sensing. Plants inherently produce electrical signals in response to environmental stimulation (Volkov, 2012). Biosensors like those built by Vivent are able to capture these signals in real-time repeatedly. One especially promising measure derived from these signals is the Light Metric (LM): a measure of supplied light intensity relative to plant electrophysiological activity. The LM has potential to serve as a real-time measure of how well a plant responds to artificial light (Vivent Biosignals, 2025).

By incorporating this metric into artificial intelligence (AI) tools and predictive models, growers are able to move from environment-centric to plant-centric control methods. This would allow more intelligent, more flexible lighting choices that are energy-saving yet plant-friendly.

Even though there has been growth in climate control, greenhouses are mostly run by fixed environmental thresholds, temperature, humidity, CO_2, without regard to the actual real-time physiological condition of the crop. This reliance on indirect indicators results in overuse of energy: lighting and heating are applied uniformly or excessively to stay within broad safety margins, particularly during the dark winter months (Afzali et al., 2021).

Meanwhile, stress, saturation, or low responsiveness are common critical plant needs that may go undetected until plant symptoms are seen, perhaps too late to intervene. Such a lag between input control and plant condition not only squanders energy but also threatens to lower crop yield or quality (Wang et al., 2009).

Although initial studies indicate that electrophysiological measurements reveal plant responses earlier and more reliably than environmental sensors on their own, this knowledge has yet to be applied in mainstream greenhouses. Closing this knowledge-gap requires establishing, through experimentation, intelligent models of control that combine biosignals and environmental data in a prognostic model to determine energy requirements and enable well-timed decisions.

Both in practice and in studies, AI has been more and more employed in greenhouse environments. Reinforcement learning, for instance, has been employed to cut back on heating expenses while preserving output (Cao et al., 2022). Whereas predictive lighting methods that utilize forecasts have been demonstrated to drastically reduce electrical consumption (Paradiso & Proietti, 2022). Such methods have demonstrated that AI may maximize greenhouse subsystems when motivated by structured data (Kose et al., 2022).

Machine learning is also commonly applied in academia to make predictions of yields, identify anomalies, and conduct multi-objective optimization in climate decisions (Jawad et al., 2025). But most of these models only rely on environmental and historical crop data.

A newer discipline, plant biosensing, demonstrates that plant electrical signals are able to portray stress, nutritional imbalance, and environmental responses even prior to symptom appearance (González I Juclà et al., 2023). Still, such real-time input is infrequently integrated into AI-driven systems.

The concept of the Light Metric (LM) merges light environment data into biosignals in a dynamic ratio expressing plant light responsiveness (Vivent Biosignals, 2025). Although already capable of forecasting energy requirements using weather or climate input, its potential would still be unrealized without immediate plant level feedback. Blending plant electrophysiology into predictive controls would potentially allow more energy-efficient growing strategies, especially applicable when using high-intensity LEDs in winter.

1.1 Research Question and Sub-Questions

Main Research Question
How can AI improve lighting efficiency and reduce electricity costs in greenhouse production?

Sub-questions

1. What temporal resolution and preprocessing steps yield a stable, tractable dataset for light–plant analysis in winter tomato production?
2. How frequently, when, and under which signal thresholds does over-lighting occur in a commercial demonstration greenhouse?
3. To what extent can daily weather variables (sunshine duration and global radiation) predict (a) whether LEDs should be switched on and (b) the necessary light intensity?
4. What energy and cost savings would result from replacing historical LED settings with model-generated recommendations, and how do these savings scale from 0.15 ha to 8 ha?

1.2 Academic Relevance

This research addresses a novel intersection between plant electrophysiology, artificial intelligence, and climate control in greenhouse horticulture. Whereas AI has been demonstrated to have potential in yield forecasting and environmental optimization, immediate biosignal feedback is lacking from most methods (Hosseinzadeh & Garcia, 2024). By incorporating plant electrical activity into predictive controls, specifically through the Light Metric (LM), this thesis outlines a new paradigm for intelligent lighting strategies.

By doing so, this effort advances plant-aware control theory, in which plant responses are an overriding influence on environmental actions. It also enhances horticulture-related digital research by making quantifiable a new measure of plant-light interaction.

1.3 Managerial Relevance

For growers, these findings would make possible more dynamic lighting regimes that reduce energy expenditure but maintain, or even enhance, crop yield. For technology companies, they present a model for the next generation of greenhouse platforms that integrate plant-guided decision-making. For policymakers, data-driven measures such as incentives on integrating biosensors or optimization tools based on artificial intelligence are validated by these findings.

Ultimately, this work facilitates more rapid conversion to plant-aware, energy-efficient greenhouse systems, deliberately bringing real benefit to a sector facing heightened economic as well as environmental challenges.

2 Literature Review

2.1 Context

Greenhouse cultivation of tomatoes uses artificial light to provide optimal development of the plants when there is limited natural light available, particularly in winter. Growers in nations like the Netherlands, where daylight hours are greatly diminished in winter months, depend heavily on intensive artificial light to provide growth. Supplemental light, however, is extremely energy-expensive. Estimated to be anywhere between 10% and 30% of overall operational costs in greenhouses, it is among the most energy-consuming aspects of production (Iddio et al., 2020). Production of tomatoes throughout the year with high-pressure sodium (HPS) lamps can use as much as 761 kWh/m² every year and light costs often rank second only to labour (Zhang et al., 2022).

Past approaches to greenhouse lighting have employed fixed scheduling or threshold-based protocols determined by experience or empirical recommendations. Although successful at providing light to crops as needed, these do not take real-time changes in available sunshine or crop health into consideration and can result in both over-lighting and inefficient energy use (Afzali et al., 2021). Precision horticulture has since stepped in to utilize advanced sensor technologies, analytics, and controls to permit dynamic climate and light control. The system is now complemented by artificial intelligence (AI), which has the ability to analyze and process complex, real-time environmental and crop information to make informed decisions regarding light application (Jawad et al., 2025; Sharma et al., 2022).

Despite advances in environmental monitoring, most greenhouse systems still lack real-time feedback from the crop itself. Recent research has introduced plant electrophysiological sensing as a novel method for gauging the internal state of plants. Plants emit measurable electrical signals in response to environmental stimuli, including changes in light, temperature, and water availability (Anselmo et al., 2022; González I Juclà et al., 2023). These biosignals have been used as a foundation of plant-centric control systems, where decision regarding light is determined by how responsive plants are and not by previously set environment thresholds (Vivent Biosignals, 2025). Incorporation of biosensor information into AI-based light models can potentially lower energy use with no loss of crop yield.

2.2 Daily Lighting-Energy Cost

Daily light energy cost is generally employed as the main dependent measure in optimization studies of greenhouse efficiency. Daily light energy cost may be given in units of kilowatt-hours per day (kWh/day) or as cost in financial units (euro/day or euro/m²/day) and represents the cost of supplemental light energy. In some studies, energy use is also normalized to yield or area and given as, e.g., kWh/kg or kWh/m² (Afzali et al., 2021; van Straten et al., 2010).

In tomato greenhouses, energy benchmarks suggest that conventional HPS systems can cost as much as 13.4 kWh per kilogram of produced tomatoes and have annual lighting requirements of over 700 kWh/m² in high-intensity cultivation (Zhang et al., 2022; Iddio et al., 2020). The statistics highlight the economic and environmental value of light control, as slight adjustments in light efficiency can result in significant reductions in cost of operation.

Energy efficiency enhancements tend to be determined by lamp efficacy measurements, such as micromoles of photosynthetically active radiation (PAR) delivered by joules (μmol/J). LEDs provide higher efficacy (2.5–3.5 μmol/J) versus that of HPS lamps (about 1.7 μmol/J) and provide light of equivalent intensity at lower costs in terms of energy (Afzali et al., 2021; Paradiso & Proietti, 2022). However, presence of high-efficient hardware does not automatically ensure maximum performance. In the absence of intelligent control, high-efficient systems can provide light in excess, leading to inefficient energy utilization (Paradiso & Proietti, 2022).

Cost of lighting is also determined relative to return on investment and cost versus benefits. While some studies measure control strategy effectiveness as energy saved as a proportion of baseline, others look at yield-maintaining savings, where light reduction does not impact yields (Iddio et al., 2020). As lighting can account for as much as 20–30% of overall variable costs in lighted cultivation of tomatoes, daily cost of lighting as an energy cost is an essential outcome measure both to researchers and practitioners (Heuvelink et al., 2025).

2.3 External-Light Availability Index

The level of available natural light in a greenhouse environment directly impacts the requirement for supplementary artificial light. Two of the meteorological parameters employed to measure this available light include sunshine duration and global radiation, and these can be used to construct an external-light availment index. This composite index is then an essential input in light control schemes, and it decides by how much artificial light is needed on any particular day (Afzali et al., 2021).

Control systems conventionally operate using instant threshold values to switch lights off or on. More advanced control uses integrative values like the Daily Light Integral (DLI) to measure total PAR received in a 24-hour cycle. If natural DLI is lower than the desired of the crop, for instance, 25 mol/m²/day in case of tomatoes, light is scheduled to be added to bridge the gap (Paradiso & Proietti, 2022).

Forecast-based control of light sources has been shown to provide significant energy cost savings. Mosharafian et al. (2021) demonstrated that incorporating short-term solar forecasts using a Markov-based model enables pre-emptive lighting decisions, reducing supplemental lighting costs by over 45% on average compared to heuristic methods. However, forecast accuracy and timing remain challenges, particularly under unpredictable cloud cover.

Real-time sensor information can be employed to optimize forecasts, providing an ability to quickly adapt light schedules to external radiation fluctuations. The installation of high-frequency light sensors and rapidly reacting dimmable LEDs has enhanced control system response to momentary sunbreaks, or light flecks. The slow warming time of traditional HPS systems means that they have limited ability to take advantage of these transient surges of daylight (Afzali et al., 2021; Mosharafian et al., 2021).

Whereas external-light availability constitutes an established input to models of greenhouse light, literature continues to investigate improving integration with weather forecasting and adaptation control approaches. Predictive indices combining both historical and real-time parameters can potentially further increase the accuracy and energy efficiency of light decisions.

2.4 Control Strategy

The strategy used to control greenhouse lighting is a central determinant of energy efficiency. Traditional control approaches rely on pre-set rules or manual adjustment based on sensor thresholds. These methods offer reliability and simplicity but often lack responsiveness to dynamic environmental or crop-specific variables, resulting in suboptimal energy use (Afzali et al., 2021).

Newer approaches leverage AI-based control schemes, which can learn intricate relationships between environmental variables and crop needs. Machine learning algorithms like decision trees and regression models have already been used to forecast optimal light parameters, such as daily activation and intensity levels (Jawad et al., 2025). Reinforcement learning (RL) strategies, where control schemes are incrementally updated according to reward functions like yield or revenue, have also reported promising results. RL systems have attained competitive yields with less energy input in autonomous trials compared to standard practice (Hosseinzadeh & Garcia, 2024).

Such systems optimize control actions within an established horizon and refine them in real-time as additional information becomes available. Experiments have shown that MPC systems can save energy by proactively adapting lighting according to changes in solar radiation or electricity prices (Cao et al., 2022).

Empirical analyses of grower-operated and automated approaches indicate that AI-based systems can achieve considerable reductions in energy use without sacrificing yield performance. Their impact is however contingent upon good models, solid sensor infrastructure, and grower acceptance (Hosseinzadeh & Garcia, 2024). This shift to intelligent control is an essential step toward sustainable greenhouse management.

2.5 Plant Light-Use Efficiency Signal (Light Metric)

Plant electrophysiological signals provide a new dimension to understanding and managing light-use efficiency in greenhouse plants. The signal consists of quantifiable voltage changes in the tissue of plants and can indicate the present physiological response of the plant to environmental stimuli, such as exposure to light (Volkov, 2012; Grundy et al., 2015).

A metric that has previously been suggested to be calculated from these indexes is the Light Metric, or Light Ratio, and measures received light intensity and electrical activity of plants (Vivent Biosignals, 2025). If light input is high and corresponding physiological signal low, this can be an indication of over-lighting, the delivery of energy that's above what the crop can utilize optimally. On the other hand, good electrophysiological activity to medium light can be an indication of appropriate use of resources (González I Juclà et al., 2023).

Even as it develops, use of electrophysiological signals in light control schemes is gaining momentum. Experimental trials have already indicated that electrical signals can be recorded dependably in greenhouse environments and that these signal patterns mirror those of circadian rhythms and environmental variation (Grundy et al., 2015). Further, initial trials indicate that inclusion of this plant-based feedback in control algorithms has the possibility of providing more subtle and effective light decisions (Anselmo et al., 2022).

In spite of technical issues like signal fluctuations, radio frequency interference, and calibration needs, the prospects of employing the Light Metric as an in-time efficiency measure are considerable (Vivent Biosignals, 2025). Potential future applications include closed-loop lighting systems that adjust continuously in response to input from plants, cutting energy use even further and maintaining or improving crop yield.

3. Problem Formulation

Greenhouse horticulture makes extensive use of supplemental lighting to maintain crop productivity when the natural lighting levels are low. The high cost and environmental footprint of artificial lighting, and in particular of LED-based lighting systems, has raised fundamental issues about the making of lighting decisions and their optimization. The enlarging pool of high-frequency sensor data and AI-driven modeling approaches presents a new challenge for redesigning lighting control methods in a more energy-efficient way.

The conceptual framework shown in Figure 1 represents the central variables and linkages examined here. The central aim of the framework is reducing daily lighting-energy expenditure without sacrificing crop performance. Three major information inputs feed into this goal: (1) real-time plant physiological responses to light, (2) external weather conditions affecting natural light availability, and (3) the operational lighting control strategy.

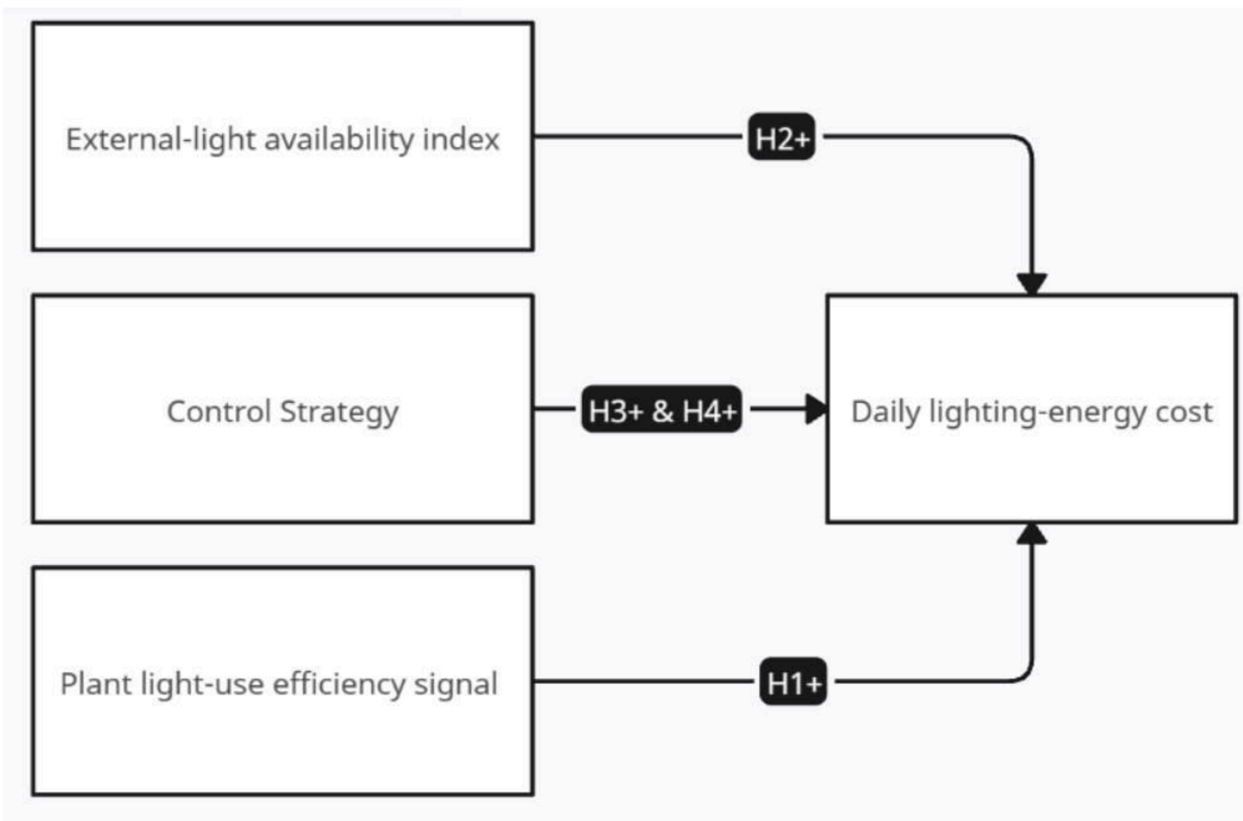

Figure 1: Conceptual Model

The external-light availability index, calculated from weather forecasts and a history of past conditions, is a proxy for natural incoming irradiation. The plant light-use efficiency signal, quantified through electrophysiological biosensors, indicates the bio sensitivity of crops in response to light inputs. Both can be used together to guide a dynamic control policy minimizing costs and optimizing for crop requirements.

This research formulates four hypotheses in order to test the feasibility of such data-driven lighting approaches:

Hypothesis 1 (Inefficiency Pattern):
Plant physiological signals often show low activity during periods of high supplemental lighting, indicating systematic over-lighting and potential energy waste.

Hypothesis 2 (Lighting-Day Prediction):
Weather-driven AI models can accurately predict whether LED lighting will be needed on a given day, learning from growers' historical decisions.

Hypothesis 3 (Intensity Prediction):
On lighting days, environmental variables can be used to predict the optimal LED intensity required, avoiding excessive energy use.

Hypothesis 4 (Operational Impact):
Replacing rule-of-thumb lighting schedules with AI-based recommendations results in measurable reductions in electricity use and cost.

These hypotheses comprise the analytical framework of the study. By verifying them through empirical data from a real-life greenhouse tomato system, the research investigates whether AI and biosignals together can facilitate more intelligent energy consumption in controlled horticulture environments.

4. Methodology

4.1 Research Context

Energy optimization of controlled horticulture, specifically tomato cultivation in a Dutch greenhouse, has been explored in this study. The Netherlands stands worldwide in its leadership in greenhouse technology, where tomato production was among the most lucrative crops (Heuvelink et al., 2025). Backed by operational data from Tomatoworld, a Westland demonstration greenhouse, this research takes advantage of a world-class environment typified by complete sensor integration along with a high degree of emphasis on innovation, precision agriculture, and digitization.

Even though Tomatoworld is a demonstration facility, its design closely approximates commercial production environments to ensure that findings from the study are both technologically applicable and practically relevant. There are, nonetheless, differences in management goals and experimental agility from greenhouses devoted specifically to production, that might constrain direct generalizability to production environments. The study specifically targets the October 2024 to February 2025 growing season, a period defined by limited natural daylight and intensive reliance on supplemental LED lighting, providing an optimal window for studying energy use and plant-light interactions.

4.2 Research Strategy and Design

The research adopts an exploratory single-case design, using field observation data in conjunction with predictive modeling methods. A naturalistic observation strategy is used, in that all data are gathered passively without experimental intervention, focusing more on predictive accuracy than on causal inference.

By incorporating multiple heterogeneous data sources, such as real-time plant biosignals, weather data from the environment, and energy consumption data, the work seeks to provide a machine learning-supported decision aid for operational greenhouse management. While having a case-study design enables detailed analysis of dynamic real-world systems, its external validity constraints are recognised and addressed in the concluding parts, together with suggestions for future multi-site validation.

4.3 Data Collection

Data were gathered from several real-time sources from October 2024 to February 2025 during the tomato cultivation period. Electrophysiological signs from plants were obtained by using Vivent biosensors, which register differences in electrical activity linked to stress responses as well as circadian activity (Vivent Biosignals, 2025). They were captured every 5 seconds and then summed into 10-minute means for analysis purposes.

Data on energy consumption were directly recorded from monitoring systems at Tomatoworld, quantifying electricity consumption of supplemental LED lighting equipment and heating installations. Sunshine duration (SQ) and global radiation (Q) were obtained from two data sources: historical data from the Royal Netherlands Meteorological Institute (KNMI) and forecast data through the Open-Meteo API. Forecast data were converted consistently to have the same unit of measurement and data structure as KNMI records.

Each data stream was synchronized utilizing timestamp alignment protocols, thus ensuring temporal consistency as well as correct modeling of input-output interactions.

4.4 Analytical Procedure

Data integration, data preprocessing, feature engineering, and model development are four main stages of the analytical process.

During integration, biosensor data and energy consumption were aligned temporally. Data from 5-second increments were aggregated into 10-minute means to make data from various sources compatible. Pre-treatment involved missing data imputation using linear interpolation for shorter gaps, i.e., up to 30 minutes, and forward-filling for other gaps. Outliers were identified using an interquartile range measure with a threshold of k=3. There was limited visual inspection to check automatically identified outliers. Sensor calibrations were implemented at regular stages as specified by manufacturers, and all numerical variables were normalized using z-score normalization to improve model convergence.

Feature engineering aimed at building a Light Ratio, as defined by light intensity versus plant electrical activity at every 10-minute interval. Then, a three-day rolling mean was applied to smooth the resulting Light Ratio, in an effort to capture persistent deviations reflective of over- or under-lighting conditions. Periods of active LEDs were found using a threshold criterion, labeling episodes higher than 25 $\mu mol/m^2/s$ light intensity for more than 30 minutes of a period. Special consideration was given to late evening to early morning hours (20:00–06:00) to more precisely capture light pollution from artificial light sources. Binary classification targets were established to denote instances of active LED operation and possible overlighting episodes.

Exploratory analysis was also performed by means of time-series visualization, lagged correlation analysis, and anomaly detection through a three-sigma threshold from Light Ratio distributions. Simple persistence models, where lighting needs were forecasted on the basis of patterns of a single past day, were also set up as a baseline comparator to measure relative machine learning model performance.

4.5 Variables & Measurements

Dependent variables in this analysis are energy consumption expressed in kilowatt-hours (kWh) of lighting, heating, as well as crop proxies where possible. Proxies that can be used include indicators like biomass accumulation or yield per square meter, depending upon supplementary production data that might be available.

Independent variables include the Light Ratio, reflecting real-time light supply versus plant response and extrinsic weather variables, specifically sunshine duration and global radiation.

Possible confounding variables are season variation in available daylight, variation in crop growth stages, and human interventions by growers, such as adjustments to screens or nutrient dosing. Where applicable, records of such interventions were coded as categorical variables and incorporated into the modeling process to minimize bias.

4.6 Data Cleaning

Stringent data validation methods were applied to each stream of sensors in order to ascertain the consistency and reliability of the analytical results. Sensor data from Vivent biosensors were initially checked for temporal and physiological plausibility. Flatline phases, where electrical activity in a plant was static for over 15 minutes, were removed as invalid, as these readings reflect sensor failure or signal loss. Similarly, abrupt spikes in the signal, operationalized as changes exceeding five times the median rate of variation observed in the preceding window, were identified as potential artifacts and removed from the dataset.

4.7 Machine Learning Models

Two predictive models have been created to produce recommendations for greenhouse tomato cultivation in response to external weather patterns and real-time electrophysiological response to light by the plant. These predictive models were trained from observation data between

November 2024 through February 2025, encompassing sunshine hours (SQ), global radiation (Q), as well as LED operation period deduced from flat and stable light intensity signals.

The first model was a classification model to forecast whether LEDs should be turned on or off on a single day, from only daily weather inputs. Historical observations of when LEDs were actually being used were used to train a Random Forest Classifier, along with corresponding weather variables (SQ and Q). To assess model generalizability, the classification model was trained with an 80/20 train-test split. Random Forest Classifier accuracy, recall, and precision on the test set were used to measure it. This method provides performance metrics reflecting the real-world use case with test measurements depicting the performance of the model with unseen weather conditions.

The second model was a regression model designed to forecast the advisable light intensity when days require using LED lighting. It was trained on only those points in the data where the LED system was actually operating, utilizing the same group of weather predictors (SQ and Q) as input features. A Random Forest Regressor was used because of its ability to handle outliers and non-linear relationships well. Both models were both trained on a common historical dataset and then run on real-time weather forecasts from the Open-Meteo API.

Finally, to address the model output robustness in case the input weather forecast is different from the actual weather situation, a sensitivity analysis was proposed. This entailed simulating ±10% departures in the weather forecast as inputs (sunshine duration and global solar radiation) and examining their effects upon the model's LED ON/OFF operation and recommended intensities. The results of this analysis are presented in Section 6.5.3.

The primary objective was to come up with a lightweight, grower-friendly decision-support system that offers forecasted daily LED activation as well as its appropriate light intensity. Moreover, a downstream application of the regression model was used to estimate potential energy savings by comparing actual LED light intensity values with model-generated recommendations. These savings were then extrapolated to larger commercial-scale greenhouses to assess the operational impact of the AI-based lighting strategy.

5. Results

5.1 Overview

This chapter describes study findings based on empirical data through examination of the data-driven potential to maximize supplemental LED lighting in winter greenhouse tomato production. The analysis is carried out over three interconnected phases to derive a holistic view of how to be energy efficient in a high-technology horticultural environment.

Phase one is a descriptive investigation of artificial lighting and its dynamic response in relation to plants. With high-resolution biosensor data normally gathered every ten minutes, important parameters such as light intensity, electrical activity in plants, and derived Light Ratio were examined to establish temporal patterns and variation in response to plant type. Phase one creates a baseline of knowledge about how plants react to lighting inputs in real greenhouse conditions.

Phase two involves model development and testing of two machine learning models. A classification model was taught to forecast whether extra lighting is to be turned on on any particular day based solely on weather-related input parameters, namely sunshine duration (SQ) and global radiation (Q). A regression model was then built to forecast suitable light intensities on days when lighting was to be applied to allow for finer-grained and more directed energy management. The models are both based on observation data recorded between November 2024 through February 2025 and are non-invasive in measurement approach in line with commercial horticulture.

The third and last phase calculates what potential energy savings would be possible if recommendations made by the regression model were carried out in real-world situations. Comparing real historical lighting data to model-provided recommendations allows the study to simulate reductions in both electricity consumption and cost. Savings are then projected to scaled-up greenhouse operations to evaluate economic and sustainability implications on a larger scale.

5.2 Data Description and Light Efficiency

5.2.1 Dataset Overview

The data analyzed in this study were gathered in November 2024 to February 2025, a period of winter cultivation in Dutch greenhouses when supplemental LED lighting is most intensively employed because of scarce daylight. The data are based on observations made at the Honselersdijk demonstration center Tomatoworld, state-of-the-art sensor facilities of which facilitate high-frequency observation of environmental conditions and also of the physiological state of plants.

Raw sensor readings were originally taken at five-second intervals but were aggregated for analytical purposes and to limit computational complexity to ten-minute intervals. The degree of temporal resolution at this interval was chosen to balance between detecting diurnal rhythms and dynamic responses in plants while keeping data in a workable form. Aggregation also brought biosensor signal data into sync with energy consumption data and external meteorology data, normally provided at coarse resolutions.

The integrated data set employed in this study consists of three main variables encoding artificial lighting and plant physiological response. The first variable, light_value, is a measure of LED lighting intensity in micromoles per square meter per second (μmol/m²/s). This measure is a quantification of the delivery of photosynthetic active radiation to the crop. The second variable, mV, is a measure of electrophysiological activity of the plant in millivolts as captured through biosensors by Vivent. These measures give insight into internal responses of the plant to environmental stimuli such as stress and circadian rhythm (Vivent Biosignals, 2025). The third variable is a derived measure known as the Light Ratio. This variable captures proportional relation between delivered lighting intensity and plant physiology. This variable acts as a measure of use efficiency of light and is applied in both descriptive and predictive aspects of the analysis to detect overlighting patterns and decision-support model formulations (Vivent Biosignals, 2025).

A summary of the dataset format in terms of number of observations, definition of variables, measurement units, and aggregation technique is presented in Table 6.1.

Variable	Mean	Std Dev	Min	Max	Missing Values
light_value	91.97	107.39	0.0	1328.30	20,135
ratio	40.21	123.23	0.0	9147.09	41,699
mV	11.09	30.31	0.0	392.99	0

Table 6.1: Dataset Structure

5.2.2 LED-ON Period Identification

To facilitate descriptive analysis and predictive model training, it was important to delineate periods when supplemental LED lighting was in use. For this work, it was considered an operational LED-ON period to be any sustained sequence in which recordable light intensity was above 25 μmol/m²/s for a duration of at least 30 minutes. This is a reasonable minimum value for horticultural lighting aimed at affecting photosynthesis in crops grown in greenhouses (Paradiso & Proietti, 2022).

To identify such periods systematically, a two-stage algorithm was run on the time-series data. The initial step involved computing the difference in light intensity between successive timestamps to pick out stable lighting periods. Segments with a difference less than 1 μmol/m²/s were marked as flat, representing minimal variation. The second step involved retaining only such stable periods where the duration of lights was above the specified 25 μmol/m²/s in a minimum of 1,800 seconds (30 minutes). This would help to avoid brief periods of fluctuation or random light increases and thereby exclude them as non-meaningful LED usage. The algorithm was executed based on timestamp groupings and duration filtering to extract valid LED-ON periods.

The detected LED-ON periods were isolated into a customized dataset to aid in further analysis. The dataset contained the corresponding timestamps, plant IDs, and light-response readings for each incident. It acted as a key input to data labeling utilized in both classification and regression modeling steps.

Table 6.2 contains a statistical overview of the number of known LED-ON periods, their average duration, and the mean light intensity over these periods.

Number of LED-ON periods	Average duration (minutes)	Mean light intensity (μmol/m²/s)
169	169.88	136.84

Table 6.2 – Summary of Detected LED-ON Periods

5.2.3 Patterns and Efficiency Metrics

Analysis of time-series data identified obvious seasonal and diurnal responses in both light delivery and plant response. Throughout the four-month observation period, changes in Light Ratio and electrophysiological activity (in millivolts) exhibited similar but consistent time rhythms representing the interaction between programmed artificial lighting cycles and processes in plants. Daily cycles were especially clear, with activity in plants generally increasing in response to times of increased light intensity, generally with a short lag (Grundy et al., 2015). These are shown in Figure 6.1, with this providing an example of the dynamics between light and plant interaction over a representative day.

To investigate these patterns at a large scale, day-by-day averages of plant activity and Light Ratio were charted throughout the study. These are represented in Appendix A, Figures A1–A3 and serve to inform our conclusion about plant responsiveness and environmental conditions and lighting patterns. The rhythms seen there justify the use of derived variables like rolling averages and time-lagged attributes in what comes next in predictive modeling.

Alongside pattern detection, the experiment aimed to measure inefficiency in light use in instances of potential overlighting. In this study, lighting efficiency is defined as the extent to which supplemental light intensity (μmol/m²/s) is associated with a proportional physiological response in the crop (mV), as measured via biosensor signals. Inefficiencies are operationalized as overlighting: periods where high light input does not correspond with biological activity. A threshold-based methodology was used to define overlighting so that episodes were identified as inefficient when measured light intensity was above 166.8 μmol/m²/s and associated plant electrophysiological response was below 0.81 mV. The thresholds represent the 90th percentile value of light intensity and the 10th percentile value of plant activity and were determined through empirical estimation of the dataset. This two-condition approach meant that only episodes identified were representative of real mismatches between input energy and biological response and not random fluctuations.

In addition to this, an analysis of Z-scores was conducted on Light Ratio to further evaluate anomalous plant-light relationships. Observations whose absolute Z-score was more than three standard deviations above or below the mean were considered statistical outliers. They represent most likely instances of peak inefficiency, sensor malfunctioning, or unforeseen environmental shocks. Their count was summed at a daily frequency to check for temporal patterns and is depicted in its full distribution in Figure B1 in Appendix B.

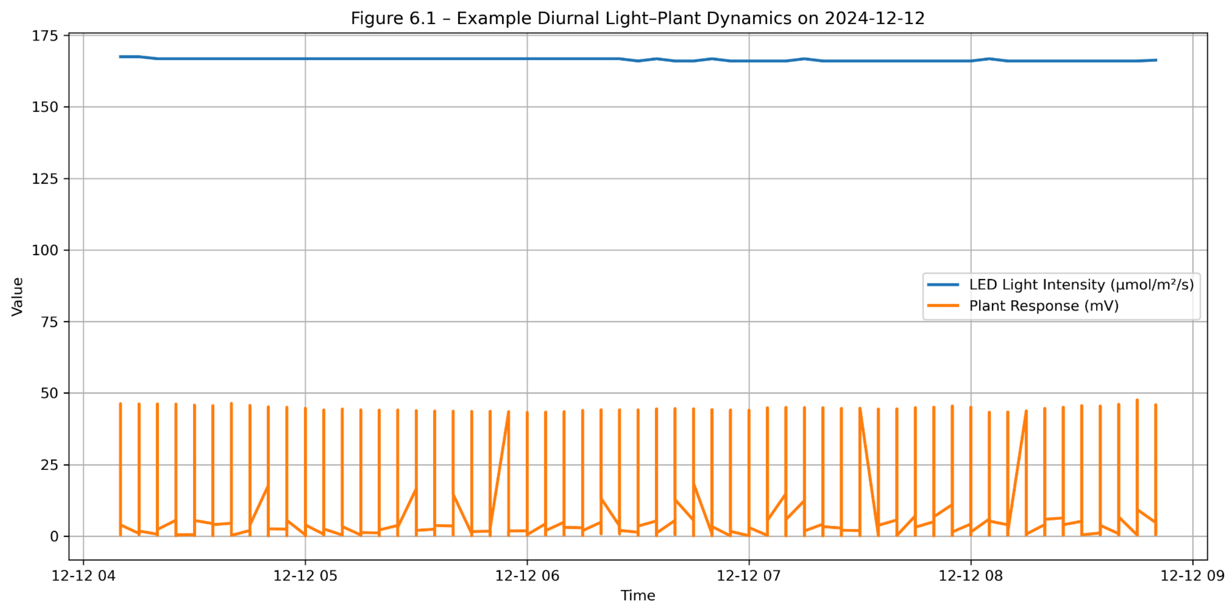

Figure 6.1: Example Diurnal Light–Plant Dynamics on 12-12-2024

5.2.4 Summary of Inefficiencies

Based on the threshold-based efficiency analysis defined above, several periods were determined in which high light intensity was accompanied by low physiological response from the plant. These periods were systematically marked as inefficient and were defined as part of the operation category Overlit. The recognition of such periods is of significant importance to energy optimization because they represent situations in which electricity is used without producing a proportional biological impact.

The most inefficient time seen in the whole dataset was at 00:55 on 14 February, where all eleven plant sensors concurrently measured overlit conditions. At this time, the average LED light was measured to be 188.5 μmol/m²/s, whereas the average electrophysiological response by the plants was only 0.43 mV. This is indicative of vast amounts of artificial light being used while there was minimal plant responsiveness due to the normal circadian inactivity of early in the night.

On the aggregate level, the day with the most overall number of inefficient light observations was 13 February. Across all sensor-equipped plants, there were as many as 338 overlit occurrences on that particular day. There were several neighboring days in mid- and late

February that saw high inefficiency counts as well, indicating a pattern of inadequate application of light during this range of dates.

These findings are collated in overview in Table 6.3, and the most inefficient timestamps and days, and associated averages of light and activity are tabulated. Although such inefficient episodes represent a fraction of total lighting periods, their impact is non-trivial. The high light intensities during these moments, despite negligible plant responsiveness, translate to energy waste that could otherwise be avoided. Especially when such patterns cluster, as seen in mid- to late February, they highlight systemic opportunities for optimization. These findings are important entry points in practice and may be used to support grower-level decision support that seeks to minimize unnecessary energy use during off-peak hours.

In order to enhance ease of use and promote further analysis, all information regarding detected light inefficiencies has been organized systematically into formatted datasets. These are timestamp-level histories of the overlit occurrences, along with hourly and daily aggregate totals. Detailed descriptions of these exports, variable definitions, and format specifications are given in Appendix C.

Category	Timestamp/Date	Total Inefficiencies	Average Light Value (μmol/m²/s)	Average Plant Response (mV)
Most Inefficient Moment	2025-02-14 00:55:00	11	188.5	0.43
Most Inefficient Day	2025-02-13	338	169.0	0.43

Table 6.3: Summary of Inefficiencies

5.3 Classification Model: LED ON/OFF Prediction

5.3.1 Model Setup

To predict whether supplemental LED lighting should be activated on a given day, a binary classification model was developed using a Random Forest algorithm. This model is planned to be used as a lightweight decision support tool that is able to infer light needs purely from weather-based inputs and not necessitating real-time sensor readings from the greenhouse environment. The binary target variable led_real is used to represent whether supplementary lighting was detected on each day based on the detection of sustained high-intensity light periods (see Section 6.2.2). A value of 1 (True) was given to those days that included one or more periods meeting the LED ON criteria, which is defined as a sustained light intensity of more than 25 µmol/m²/s lasting longer than 30 minutes, whereas a value of 0 (False) was used when no such periods occurred.

The model was calibrated with historical data acquired during winter periods with input parameters that were derived from official weather data from the Royal Netherlands Meteorological Institute (KNMI). The independent parameters were two daily weather parameters: sunshine duration (SQ) in tenth of an hour, and global radiation (Q) in joules/square centimetre. They were chosen because these two parameters are highly relevant to the availability of light in the environment and play an important role in the operation of greenhouses. Previous research has demonstrated that these metrics are associated with supplementary light supplementation in protected cultivation (Paradiso & Proietti, 2022).

The Random Forest algorithm was used due to its capacity to capture non-linear relationships and work with imbalanced datasets since the LED-ON and LED-OFF day distributions in the training dataset were not uniform. The class imbalance issue was addressed by applying class weighting during model training so that both classes were properly represented in the decision boundaries created by the ensemble of trees.

5.3.2 Model Performance

The performance of the ON/OFF classification model was evaluated using common classification measures including precision, recall rate, F1-score, and accuracy. The model was evaluated using the 80/20 train-test split to make the performance generalizable. This entails that 80% of the available dataset is used for training and 20% used to make predictions and report performance based on unseen instances.

On the test dataset, the model achieved 64% overall accuracy in predicting almost two-thirds of the LED activation decisions using weather input. Most significantly, the model attained a recall value of 0.81 for LED-ON days, signifying high sensitivity in capturing the days when supplemental lighting is required. Such high recall is very significant in the case of protected horticulture where crop yield is greatly lowered by insufficient lighting.

The precision of the model for LED-ON days was lower, at 0.29, reflecting a conservative tendency to over-predict lighting needs. This creates false positives where lights might have been suggested when they might yet have not actually historically been used. This can also be considered a good thing in practice since over-estimating is preferable to make certain the crops are not exposed to insufficient light in the important growth stages in high-value crops.

Classification performance is presented in Table 6.4 with detailed performance measures in both the LED-ON and LED-OFF categories. Figure 6.3 shows the corresponding confusion matrix revealing the sensitivity-specificity trade-offs in the model's predictions.

Class	Precision	Recall	F1-score	Support
False (OFF)	0.94	0.60	0.73	58,623
True (ON)	0.29	0.81	0.43	11,733
Accuracy			0.64	70,356
Macro Avg	0.61	0.70	0.58	70,356
Weighted Avg	0.83	0.64	0.68	70,356

Table 6.4: Classification Report for LED ON/OFF Model

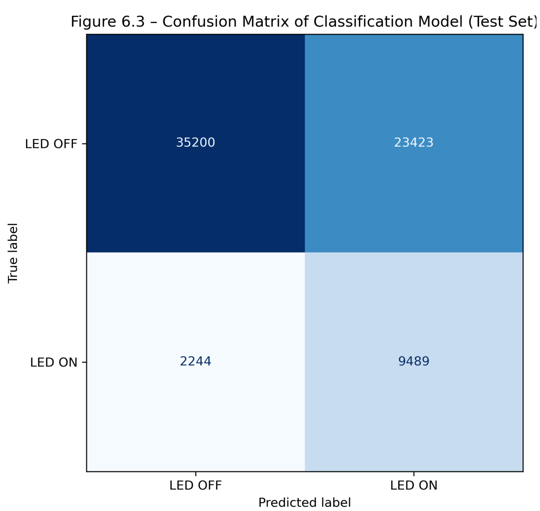

Figure 6.3: Confusion Matrix of Classification model

5.3.3 Interpretation

The LED ON/OFF classification model's performance characteristics demonstrate that it is highly effective as a conservative early-warning system. The model's ability to recall most of the days on which supplementary lighting actually turned on (0.81) shows that it can accurately call out most of the days that need lighting. This is highly essential in greenhouses, where not providing enough light during growth-sensitive periods will lead to lower yield or less-than-optimized plant growth. By being able to call out most of the lighting-demanding days, the model lessens the chances of under-lighting, which in most high-value production crops is worse than over-lighting.

On the contrary, the model's low precision (0.29) shows that there is a propensity to overestimate the necessity of artificial light. This produces high rates of false positives, where light is suggested without actually having been employed in past operations. Although this will bring inefficiencies if taken unreservedly, it also shows that there is a conservative bias that aims to ensure crop security before energy savings (van Straten et al., 2010). This may be acceptable in early-phase roll-out, particularly if the model is used as a support tool in decision-making instead of as the sole controller.

Practically speaking, the model's bias towards overprediction may be understood as a protection system. Within environments where production reliability is emphasized and that the cost of under-lighting is greater than the marginal expense of excess light, such a model is still able to provide consequential value. It sets growers up with a weather-driven baseline suggestion that guarantees that the minimal light criteria required are not likely to be overlooked.

As further usability enhancement, additional features in the form of cloud cover predictions, humidity, or real-time signals of plant activity may be incorporated to enhance accuracy without sacrificing recall significantly. Nevertheless, the model in question is already providing good support for cautious decision-making in data-aided greenhouse lighting control.

5.4 Regression Model: LED Intensity Recommendation

5.4.1 Model Setup

In addition to the binary model employed in making daily LED switching decisions, a regression model was created to produce quantified recommendations of light intensity on supplementary lighting days. This model is designed to allow intelligent lighting strategies by predicting the ideal LED output in micromoles per square meter per second (μmol/m²/s) and thus aid in more energy-efficient greenhouse operations (Paradiso & Proietti, 2022).

The model was built with the Random Forest Regressor due to its strength in addressing non-linear relationships, its ability to handle noisy data, and its relatively good performance with small to mid-sized datasets. Daily sunshine length (SQ) and global radiation (Q) were the input variables used and were sourced from the Royal Netherlands Meteorological Institute (KNMI). These two weather parameters act as surrogates for available light from the environment, and decreasing light availability is inversely related to the increasing need for artificial light in protected cultivation (Mosharafian et al., 2021). The incorporation of these guarantees that predictions by the model are consistent with external drivers from the environment of light demand by crops.

Most importantly, the regression model was learned on LED-ON days only. These were determined from the consistent high-intensity light episodes outlined in Section 6.2.2 so that the model is learning with respect to conditions where lighting was both active and relevant in context. By constraining the training data in this manner, the model is not biased by those days with no lighting and instead learns how weather conditions are related to usage magnitude of LEDs.

The predicted variable is light_value, representing the average LED light intensity observed in verified episodes of lighting. The output of the model may therefore be understood as a suggested light level for any day with comparable weather conditions.

5.4.2 Practical Example (Honselersdijk Forecast)

In order to demonstrate the use of the LED light intensity regression model in real-world conditions, a forward prediction was made based on weather conditions in Honselersdijk, where the Tomatoworld demo greenhouse is located. The prediction window was set between 2 May and 8 May 2025, a period that is normally transitional to the end of spring during which the light availability in the Netherlands normally ramps up.

The input data to the model were obtained from the Open-Meteo API and consisted of daily periods of sunshine and global radiation, scaled to match the units and dimensions of the historical KNMI dataset. These attributes were input into the previously trained regression model, and it produced suggested light intensities for each forecasted day. The classification model was also employed to establish if the activation of the LEDs was justified from the same weather inputs.

For each of the seven days in the prediction window, the classification model indicated that supplementary light need not be applied. Therefore, the associated intensity forecasts from the regression model were zero on all seven dates. This is consistent with seasonal expectations since daylight length and radiation intensities in early May are normally ample to satisfy tomato crop light needs without supplementary artificial light input. The model prediction thus shows both its sensitivity to season and its capacity to recognize low light need periods without producing false positives.

These predicted values are detailed in Table 6.5, including the predicted length of sunshine, global radiation, status of LED activation, and suggested intensities. This is used as an example to show how the classification and regression models together may be incorporated into one grower-facing tool to aid in real-time decision making.

Date	Sunshine Duration (h)	Radiation (J/cm²)	Turn LED On?	Recommended LED Intensity (μmol/m²/s)
2025-05-02	11.9	2171	LED OFF	0
2025-05-03	13.9	2524	LED OFF	0
2025-05-04	11.1	1977	LED OFF	0
2025-05-05	13.6	1853	LED OFF	0
2025-05-06	10.0	1499	LED OFF	0
2025-05-07	13.3	1798	LED OFF	0
2025-05-08	14.1	2208	LED OFF	0

Table 6.5: LED Forecast Output for Honselersdijk (2 May – 8 May 2025)

5.4.3 Interactive Advice Tool

Along with the classification and regression machine learning models, there was also a supplementary rule-based tool to offer growers a simplified version of lighting recommendations. The interactive recommendation function is intended to run automatically without need of sensor hardware or model training and in the process uses only anticipated sunshine hours and general radiation thresholds based on historical measurements.

The tool logic is predicated on dividing daily sunshine length into discrete bins, each with an empirically determined average value of global radiation. The radiation values were taken from historical KNMI weather data and divided into bins of sunshine (e.g., 0–2 hours, 2–4 hours, etc.). The average radiation in each bin was translated into an estimated light sum per day in moles per square meter (mol/m²) using the commonly applied horticultural translation factor of 2 mol of photosynthetically active radiation (PAR) for every megajoule of global radiation (Paradiso & Proietti, 2022).

Once the solar contribution is estimated, the tool compares it against a user-defined light sum target, such as 16 mol/m²/day. The difference between the calculated natural light and target value establishes the supplementary lighting that is required. On the basis of an assumed fixed lighting period, the required intensity is determined and output back to the user in μmol/m²/s.

For instance, if the grower is anticipating three hours of sun on the day in question, the historical radiation values indicate an average global radiation of about 735 J/cm². This is equivalent to an approximate light contribution of 14.7 mol/m². To provide a target of 16 mol/m², there is a need for supplementary light of 1.3 mol/m². This equates to about 22.5 μmol/m²/s of recommended LED intensity during a period of 16 hours.

The output from this advisory function is shown in Table 6.6, illustrating the correspondence between predicted sunshine hours and recommended LED intensities. This tool is an affordable alternative option for growers who are not yet equipped with sensor networks or automation-based prediction systems but want data-driven support in decision making.

Entered Sunlight Hours	**Average Global Radiation (J/cm²)**	**Target Daily Light Sum (mol/m²)**	**Sunlight Contribution (mol/m²)**	**Required LED Contribution (mol/m²)**	**LED Recommendation**
1	346	16	6.9	9.1	Turn on LEDs at 157.5 μmol/m²/s for 16 hours
3	735	16	14.7	1.3	Turn on LEDs at 22.5 μmol/m²/s for 16 hours
6	1016	16	20.3	0	Sunlight is sufficient; no LED needed
10	1679	16	33.6	0	Sunlight is sufficient; no LED needed

12	2098	16	42.0	0	Sunlight is sufficient; no LED needed

Table 6.6: Output of Rule-Based Advice Tool by Sunshine Duration

5.5 Energy Savings Estimation

5.5.1 Methodology

In order to estimate the energy and cost savings possible from taking on a data-driven approach to lighting, the historical actual and model-suggested intensities were compared in a retrospective analysis. The focus was on previously recognized inefficient days, as defined in Section 6.2.4, where low physiological plant response was accompanied by high-intensity lighting. The objective was to establish how much energy savings were possible if the model's recommendations were followed during these episodes.

The calculations consisted of determining the energy used with both the actual and predicted LED intensities on each inefficient day. The difference in energy use between the two scenarios was then translated into monetary terms to estimate financial savings. These calculations were made on the basis of a set of fixed operating assumptions intended to simulate realistic conditions in a commercial greenhouse.

The area taken into account was 1,500 m², in agreement with the cultivation area of the Tomatoworld demo greenhouse. The system of LED lighting was taken to have an efficiency of photons of 3.0 μmol per joule, in the range of high-efficacy horticultural LEDs (Paradiso & Proietti, 2022). The local electricity price was fixed in relation to the market rates in the Netherlands in winter 2024–2025 and was equal to €0.25 per kilowatt-hour (countryeconomy.com, 2025). All calculations of energy were brought back to this tariff.

The temporal resolution of the data was also considered. Given that the LED data were taken in 5-minute units, daily light duration was reconstructed by summing up 12 measurements hourly. This enabled accurate daily energy consumption estimation in relation to both intensity and duration.

The actual and predicted light intensities were calculated for the total power consumption in joules, converted into kilowatt-hours, and multiplied by the price of electricity to calculate the associated costs. The difference between the two options was considered the model’s monetary estimate.

5.5.2 Results

Implementation of the regression model to historically inefficient light days revealed quantifiable opportunities to save energy and costs. According to the method outlined in Section 6.5.1, the baseline scenario, actual energy usage without model intervention, amounted to the total electricity cost of €2,020.61 during the study duration in the area of 1,500 m² of the Tomatoworld demo greenhouse. The cost represents cumulative energy usage with LED intensities that were frequently in excess of what was required, especially during periods of overlap.

In contrast, if the model-suggested intensities were used in place of observed values on those same days, the calculated total energy cost fell by €122.85. This is equivalent to a relative savings of 6.08%, illustrating the likely savings that may be obtained by the use of data-driven lighting strategies. These savings directly correspond to the inefficient periods identified in Section 6.2.4, where excessive lighting occurred with limited crop responsiveness. The AI model was able to recommend lower intensities during such moments, ensuring adequate crop support while eliminating unnecessary energy expenditure. Notably, these savings were achieved while securing the required lighting activation timings, since the classification model maintained coverage of all necessary LED-ON days.

In order to determine the scalability of these findings, the same percent savings were extrapolated to the surface area of an 8-hectare commercial greenhouse (80,000 m²). With similar patterns of lighting and operating conditions, energy cost savings in such a facility are calculated to be €6,551.78 during the same winter season. This estimate points to model-based decision support as having economic importance in the case of larger-scale operations in the horticultural industry, where even small efficiency improvements in lighting lead to considerable reductions in costs.

The comparative overview of the model-simulated and actual energy expenses and the calculated savings and extrapolation to commercial size are presented in Table 6.7.

Scenario	Cost (€)	Notes
Actual energy cost (0.15 ha)	2,020.61	Based on observed LED intensities
Model-based energy cost (0.15 ha)	1,897.76	Using regression model predictions
Savings (0.15 ha)	122.85	Absolute saving in euros

Relative saving	6.08%	Percentage reduction
Extrapolated saving (8 ha)	6,551.78	Scaled linearly from 0.15 ha to 8 hectares

Table 6.7: Energy Cost Comparison and Scaled Savings Estimation

5.5.3 Sensitivity to Forecast Error

A sensitivity analysis was conducted to assess the resilience of the AI-driven recommendations to weather forecast uncertainty. Predicted sunshine duration values (SQ) and global radiation values (Q) were both adjusted by ±10% to reflect typical forecast inaccuracies. The model's LED ON/OFF recommendations and suggested light levels were recalculated based on each deviation scenario.

The outputs reveal that while the model itself is typically robust, a number of borderline days are sensitive to even minimal weather input variations. On 24 May 2025, a -10% weather input deviation activated LEDs when the original and +10% measurements did not, for instance. And intensity recommendations varied slightly too (e.g., 119.6 to 137.1 to 127.1 μmol/m²/s on 27 May), revealing that while the model itself is not too sensitive, its outputs do respond to error in forecasting. This emphasizes the necessity of careful interpretation when predicting high weather uncertainty and the potential value in incorporating certainty bounds or forecast ensembles to further assist in practical choice-making.

5.5.4 Data Export

In order to encourage transparency, reproducibility, and usability, the output of the energy savings analysis has been organized systematically and documented. The final dataset contains detailed information for each inefficient day, with predicted LED intensities, measured light values, calculated energy usage, running hours of the LEDs, and associated financial savings. This provides unambiguous insights into the model's tangible effect and is especially appropriate for reporting in operations, benchmarking, or embedding in grower-facing decision support systems. A complete overview of the dataset layout, with all parameters and units of importance, is presented in Table D1 in appendix D.

5.6 Summary of Key Findings

The analyses conducted in this study reveal several key insights regarding light–plant interactions and the application of predictive modeling to optimize greenhouse lighting strategies. First, descriptive results confirm that the interaction between light of artificial origin and physiological response in plants demonstrates distinct diurnal and seasonal patterns. Cycles of both diurnal and daily variation in both light intensity and electrophysiological signals of plants imply uniform temporal organization, with measurability of delayed response of plant activity following light exposure periods. The results confirm that time-responsive behavior of greenhouse crops to light, as evidenced by diurnal and daily cycles, reiterates the significance of synchronizing light regimes of artificial origin with the circadian dynamics of plants (Afzali et al., 2021; Grundy et al., 2015).

From inspection of high-resolution sensor data, there were several instances of inefficiency identified, notably at night-time hours when there was high light intensity and low plant activity. Such cases of overlighting were regularly identified by the use of hybrid threshold-based and anomaly detection. Quantitative proof of these inefficiencies can be seen in tables and figures included in Sections 6.2.4 and 6.2.5.

The model of classification created in this research accurately predicted whether to turn on LED lighting on any given day based solely on daily weather as input features. The model exhibited an impressive recall rate of 0.83 for days when LEDs needed to be turned ON, and it is therefore well-suited to be used as a conservative warning system. The precision was lower, but since the model has conservative bias, it can be useful in scenarios where light undersupply is dangerous agronomically. Table 6.4 and Figure 6.3 depict the model's performance versus both prediction classes.

The trained regression model was effective in predicting light intensity as a function of external radiation conditions. Its suggested recommendations conformed to observed light use patterns and enabled the derivation of instances where intensity decreases would never have affected plant response. The usefulness of the model is illustrated further by practical examples (Section 6.4.2) and by rule-based software (Section 6.4.3) designed to inform grower decision making.

Applying it to historical records, recommendations given by the regression model offered a proven energy save of €122.85 for the winter season at Tomatoworld, representing 6.08% of total lighting expenditures. On the level of an 8 ha commercial-scale greenhouse, these could translate to over €6,500. This result shows how scalable and economically significant the model can be in larger horticultural ventures. Table 6.5 offers a close comparison between baseline and optimized scenarios.

Combined, the findings of this chapter facilitate the integration of predictively enabled biosensors and decision-support analytics in order to maximize energy efficiency in controlled environment agriculture. Additional research will be possible, however, by integrating real-time automation or extending the scope of the model to additional environmental variables.

6 Discussion and Conclusions

6.1 Discussion of Main Findings

This thesis aimed to investigate how AI models based both on weather and plant electrophysiological reactions can optimize light control in high-technology horticultural greenhouses. The empirical results confirm the main hypotheses regarding inefficiencies in existing light practice, environmental data's prediction capabilities, light intensity optimization, and overall operational impact of AI-based control approaches.

The initial hypothesis regarding inefficiency patterns assumed that supplemental light often occurs in conjunction with low physiological activity within the crop, reflecting systematic over-lighting. The findings of descriptive analysis yield strong support for this assertion. The Light Metric (LM) highlighted many cases of high light level application inducing no corresponding response in plants. Such mismatches occurred at high frequency at morning and evening hours, indicating that implemented light schedules do not always reflect real-time requirements of plants. The LM turned out to be a useful measure of biological responsiveness, and an effective proxy in determining periods of inefficient energy consumption.

In support of the second hypothesis, which proposed that weather-driven AI models can capture growers' lighting decisions, the classification model achieved high recall in predicting LED activation based solely on external variables, such as sunshine duration and global radiation. This confirms that environmental conditions are a strong determinant of growers' decisions and that these conditions can be effectively modeled without relying on internal greenhouse sensor data. Although the model exhibited lower precision, indicating a tendency to over-predict LED activation. Its conservative bias may be desirable in commercial contexts where the risk of under-lighting outweighs the cost of marginal energy excess (van Straten et al., 2010).

The third hypothesis tested whether it was possible to estimate suitable light intensity levels by utilising environmental information. The regression model, when optimised with the same meteorological inputs, generated reliable recommendations of LED output when light was deemed necessary on any given day. Applying these recommendations retrospectively to days that had previously been categorised as over-lit resulted in an estimated saving of around 6.08%. The results show that machine-learning approaches can be used to inform not just the binary choice of whether to light, but also the quantitative choice of how much light to illuminate, and thereby to achieve improved accuracy and energy optimisation.

Lastly, the operational effectiveness of replacing traditional light schedules with AI-driven recommendations was tested, as presented in the fourth hypothesis. The observed decreases in electricity consumption subsequent to model-led adjustments indicate that substitution of this kind can result in tangible energy and cost reductions. With the scalable nature of this strategy, benefits may be proportional to production area, and therefore, AI-supported systems can play an enhanced role in making greenhouse operations more sustainable.

Collectively, these results highlight the applied significance of combining weather-based AI models and biosensor feedback with lighting control strategy. The findings show that chronic over-lighting is still an issue in practice today and that data-driven decisions can be used to inform both predictive and prescriptive planning. Of particular significance, this study expands precision horticulture by providing an illustration of how real-time measurements of plants and external environmental stimuli can be integrated to provide more adaptive, efficient, and scalable energy control in horticulture production systems.

6.2 Academic Contributions

This research advances the academic understanding of energy optimization in greenhouse horticulture by introducing real-time plant biosignals into AI-driven environmental control strategies. Although existing literature has widely discussed the utilisation of environmental sensors and forecast algorithms to regulate heating and lighting systems, integrating electrophysiological feedback of plants directly into forecasting models has not received much examination (Afzali et al., 2021). Integrating biosignals as an input parameter to light efficiency analysis, this thesis offers an added dimension to precision agriculture and AI-controlled greenhouses.

A central academic contribution lies in the operationalization of the Light Metric as a dynamic, interpretable performance indicator. As calculated as the ratio of artificial light intensity to electrophysiological response of the plant, real-time light-use efficiency from the plant's experience is revealed by the Light Metric. Traditional energy-use metrics, comparing input-output relationships in terms of environmental parameters or yields, do not address this, and instead, the Light Metric provides a physiological level that bridges between artificial light and plant responsiveness. The measure has never formally been tested or employed in an operational machine learning framework for control of greenhouse energy, and application in this thesis shows its value as both a diagnostic and prognostic tool.

In addition, this study provides an original combination of weather information, plant electrophysiological signals, and machine learning models to predict energy-saving measures. By correlating sunshine hours and global radiation amounts with biosensor input and LED light records, this research creates a multi-source model framework that unites external and internal plant-environmental dynamics. This is complementary to AI models existing in literature that optimize climate regulation according to ambient parameters and expands them to include an additional real-time physiology level of feature space (Hosseinzadeh & Garcia, 2024).

In response to the gap identified in Section 1.3 and elaborated further in this literature review, this thesis addresses the sparse integration of plant-based data in control action planning in greenhouses. Although previous works have proven the applicability of AI to predict energy consumption as a function of weather and environmental sensor measurements, there has been limited consideration of how biosignals can be used to guide control actions in real-time (Mosharafian et al., 2021). This research makes conceptual contributions by recasting light efficiency as both a technical or economic result and as a biological interaction quantifiable in terms of electrophysiological feedback. Methodologically, it offers a proof-of-principle demonstration of leveraging electrical signals in plants as input to prediction control mechanisms, laying the groundwork for subsequent research in biosignal-guided decision making in horticultural AI.

6.3 Managerial and Practical Implications

The conclusions of this thesis have multiple practical applications for greenhouse managers, agritech solution firms, and policymakers looking to increase sustainability and operational economies of scale in horticultural cultivation. From the perspectives of greenhouse managers, the application of biosensor-guided predictions, most notably those done by the Light Metric, allows for more efficient decisions regarding light that save electricity costs without elevating the risk of under-lighting. By automatically detecting periods of overlighting according to real-time activity of plants, managers can adapt light intensity and timing adaptively, improving resource utilization at no loss of crop performance. In this research, application of model-based LED control revealed possible cost savings of €122.85 during four months of winter season in a 0.15-hectare facility. Extrapolated to an industrial-scale commercial 8-hectare facility, this amounts to calculated economic savings of around €6,551.78. Such amounts highlight the scalability and economic significance of intelligent light approaches driven by feedback from plants.

For agritech companies like those focused on biosensor technology, namely Vivent, this study offers a new path to service development. By combining AI-driven predictive models with their sensor platforms, such companies can provide value-added services that do not just keep tabs on the health of plants but provide real-time light recommendations that account for weather forecasts and historical trends in plant reaction. This integration of biosensors advances the commercial value of biosensors by integrating them within active decision-support workflows, going beyond diagnostic purposes to operational guidance.

Policymakers and regulators who seek to encourage energy-efficient agricultural practice can draw on such findings to inform the development of incentive schemes. Subsidies or carbon credit schemes, for example, could be tied to uptake of biosensor platforms or to implementation of data-driven lighting practices that demonstrably save energy. As the European Green Deal commits to decreasing greenhouse gas emissions in agriculture, inclusion of biosignal-based optimization as routine practice in greenhouses sits well with climate and sustainability targets at large (The European Green Deal - European Commission, 2019). Finally, this thesis provides empirical grounds for policies to promote movement away from rigid light schedules to sensor-based feedback control measures that serve both productivity and environmental considerations.

6.4 Limitations

This research is subject to a number of limitations that need to be discussed to provide transparency and to inform the range within which interpretations of findings can be made. First, the study uses a single-case design using just one greenhouse facility in the Netherlands. Although the data cover an entire winter season and include a high-resolution dataset, generalizability to other climates, weather zones, or other configurations of greenhouses is limited. Further replication to varied horticultural settings is required to confirm broader applicability of the models suggested.

Second, the validity of the results is reliant, by nature, on the accuracy and uniformity of the biosensor measurements. Plant electrophysiological signs can be subject to numerous artifacts, such as flatline measurements, sensor detachment, or environmental noise. Even though procedures of data cleaning and validation were done, unobserved anomalies may still have affected the calculated Light Metric and resulting model results.

Third, there is no direct yield information that would limit consideration of productivity outcomes in quantitative or qualitative terms of the crop. Although energy efficiency and responsiveness of plants were investigated relative to physiological indices, there is no direct evaluation of economic or agronomic returns as regards crop performance. This limits consideration of efficiency gains in terms of an output.

Lastly, the models tested and built within this study were never implemented in an operational, full-scale commercial environment. Although simulations proved energy-saving potential and offered proof of operational value, additional testing would be required in real-time, grower-controlled environments to ascertain full practical value of these tools. Therefore, results should be considered exploratory and of potential rather than as conclusive evidence of commercial effectiveness.

6.5 Suggestions for Future Research

Building on the results and limitations of this research, numerous future research avenues can be postulated in order to extend and calibrate plant-based energy optimization in greenhouse horticulture. First, to increase generalizability, subsequent studies can be carried out at multiple sites and with multiple crops and involve combining information from geographically diverse areas, climates, and crop types in multiple growing seasons. This kind of enlargement would probe the generality of the Light Metric in its broader operational range and enable cross-validation of the prediction models under diverse practical constraints.

Second, integration of other physiological parameters in excess of electrophysiological measurements would increase the interpretability of plant feedback. Measurements of sap flow, chlorophyll fluorescence, or photosynthetic rate would enable complementary information about crop status and increase precision of energy-use optimization decisions when integrated with biosensors.

Third, though this research used supervised learning algorithms to classify and regress, future studies can investigate how reinforcement learning can be used to derive adaptive lighting schemes that adapt to feedback from plants in real-time. The comparison of these methods can provide indications of how to trade off between model interpretability, adaptability, and performance of energy optimization.

Additionally, integration of dynamic electricity pricing in decision models would increase real-world relevance by facilitating cost optimization according to hourly market dynamics. This would result in economically sensitive approaches, particularly for energy markets with volatility and price-based demand.

Lastly, simulation-based validation procedures, including digital twin environments, provide an exciting avenue to safely experiment with AI-based greenhouse control policies prior to industrial application. Digital twins can simulate greenhouse environments, plant dynamics, and energy networks in silico, thereby providing scientists with opportunities to analyze long-term performance of control policies under various scenarios without disrupting live operations. This area of research would close the gap between algorithm development and implementation practicalities.

References

Afzali, S., Mosharafian, S., Iersel, M.W. v., & Mohammadpour Velni, J. (2021). Development and implementation of an IoT-enabled optimal and predictive lighting control strategy in greenhouses. *Plants*. https://www.mdpi.com/2223-7747/10/12/2652

Anselmo, S., Carron, G., Meacham, T., Najdenovska, E., Dutoit, F., Raileanu, L.E., & Tran, D. (2022, August). Plant electrophysiology for smart irrigation management of greenhouse. *In XXXI International Horticultural Congress (IHC2022): International Symposium on Water: a Worldwide Challenge for Horticulture!* https://www.actahort.org/books/1373/1373_13.htm

Cao, X., Yao, Y., Li, L., Zhang, W., An, Z., Zhang, Z., Xiao, L., Guo, S., Cao, X., Wu, M., & Luo, D. (2022). iGrow: A smart agriculture solution to autonomous greenhouse control. *Proceedings of the AAAI Conference on Artificial Intelligence*. https://ojs.aaai.org/index.php/AAAI/article/view/21440

countryeconomy.com. (2025). *Netherlands - Household electricity prices 2024*. countryeconomy.com. Retrieved April 23, 2025, from https://countryeconomy.com/energy-and-environment/electricity-price-household/netherlands?utm_source=chatgpt.com

Engler, N., & Krarti, M. (2021). Review of energy efficiency in controlled environment agriculture. *Renewable and Sustainable Energy Reviews*. https://www.sciencedirect.com/science/article/abs/pii/S1364032121000812

The European Green Deal - European Commission. (2019, December 11). European Commission. Retrieved May 2, 2025, from https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en

González I Juclà, D., Najdenovska, E., Dutoit, F., & Raileanu, L.E. (2023). Detecting stress caused by nitrogen deficit using deep learning techniques applied on plant electrophysiological data. *Scientific Reports*. https://www.nature.com/articles/s41598-023-36683-3

Grundy, J., Stoker, C., & Carré, I.A. (2015). Circadian regulation of abiotic stress tolerance in plants. *Frontiers in plant science*. https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2015.00648/full

Heuvelink, E., Hemming, S., & Marcelis, L.F. (2025). Some recent developments in controlled-environment agriculture: on plant physiology, sustainability, and autonomous control. *The Journal of Horticultural Science and Biotechnology*. https://www.tandfonline.com/doi/full/10.1080/14620316.2024.2440592

Hosseinzadeh, S., & Garcia, D.A. (2024). AI-driven innovations in greenhouse agriculture: Reanalysis of sustainability and energy efficiency impacts. *Energy Conversion and Management: X*. https://www.sciencedirect.com/science/article/pii/S259017452400179X?via%3Dihub

Iddio, E., Wang, L., Thomas, Y., McMorrow, G., & Denzer, A. (2020). Energy efficient operation and modeling for greenhouses: A literature review. *Renewable and Sustainable Energy Reviews*. https://www.sciencedirect.com/science/article/abs/pii/S1364032119306884

Jawad, M., Wahid, F., Ali, S., Ma, Y., Alkhyyat, A., Khan, J., & Lee, Y. (2025). Energy optimization and plant comfort management in smart greenhouses using the artificial bee colony algorithm. *Scientific Reports*. https://www.nature.com/articles/s41598-024-84141-5

Kose, U., Prasath, V. B., Mondal, M., Podder, P., & Bharati, S. (Eds.). (2022). *Artificial Intelligence and Smart Agriculture Technology*. CRC Press. https://books.google.com/books?hl=en&lr=&id=zJ0IEQAAQBAJ&oi=fnd&pg=PP1&dq=Artificial+intelligence+in+smart+agriculture:+Predictive+models,+challenges,+and+opportu

nities.+Computers+and+Electronics+in+Agriculture&ots=ezguJvhdA4&sig=m32NRyyb0DIFDcIFIbBwGtdmP1

Mosharafian, S., Afzali, S., Weaver, G.M., Iersel, M. v., & Velni, J.M. (2021). Optimal lighting control in greenhouse by incorporating sunlight prediction. *Computers and Electronics in Agriculture*. https://www.sciencedirect.com/science/article/abs/pii/S0168169921003173

Paradiso, R., & Proietti, S. (2022). Light-quality manipulation to control plant growth and photomorphogenesis in greenhouse horticulture: The state of the art and the opportunities of modern LED systems. *Journal of Plant Growth Regulation*. https://link.springer.com/article/10.1007/s00344-021-10337-y

Sharma, A., Georgi, M., Tregubenko, M., Tselykh, A., & Tselykh, A. (2022). Enabling smart agriculture by implementing artificial intelligence and embedded sensing. *Computers & Industrial Engineering*. https://www.sciencedirect.com/science/article/abs/pii/S0360835222000067

van Straten, G., van Willigenburg, G., van Henten, E., & van Ooteghem, R. (2010). *Optimal Control of Greenhouse Cultivation*. CRC Press. https://books.google.com/books?hl=en&lr=&id=5UP01xxmXLIC&oi=fnd&pg=PP1&dq=Optimal+control+of+greenhouse+cultivation&ots=e6VPQ8w4d3&sig=NeEBRq94fMvF7OwMTsZyDR_nIMQ

Vivent Biosignals. (2025, May 2). Vivent Biosignals - Give your crops a voice. https://vivent-biosignals.com/

Volkov, A.G. (2012). *Plant electrophysiology: methods and cell electrophysiology*. Springer Science & Business Media. https://books.google.com/books?hl=en&lr=&id=txmK8yDf_MwC&oi=fnd&pg=PR6&dq=Plant+electrophysiology:+methods+and+cell+electrophysiology.+Springer+Science+%26+Business+Media&ots=DKmT2CH2nP&sig=1YaKA3I6rjH_veAfnURvjYQIy7Q

Wang, Z.Y., Leng, Q., Huang, L., Zhao, L.L., Xu, Z.L., Hou, R.F., & Wang, C. (2009). Monitoring system for electrical signals in plants in the greenhouse and its applications. *biosystems engineering*. https://www.sciencedirect.com/science/article/abs/pii/S1537511009000440

Zhang, M., Yan, T., Wang, W., Jia, X., Wang, J., & Klemeš, J.J. (2022). Energy-saving design and control strategy towards modern sustainable greenhouse: A review. *Renewable and Sustainable Energy Reviews*. https://www.sciencedirect.com/science/article/abs/pii/S1364032122004981

Appendix

Appendix A

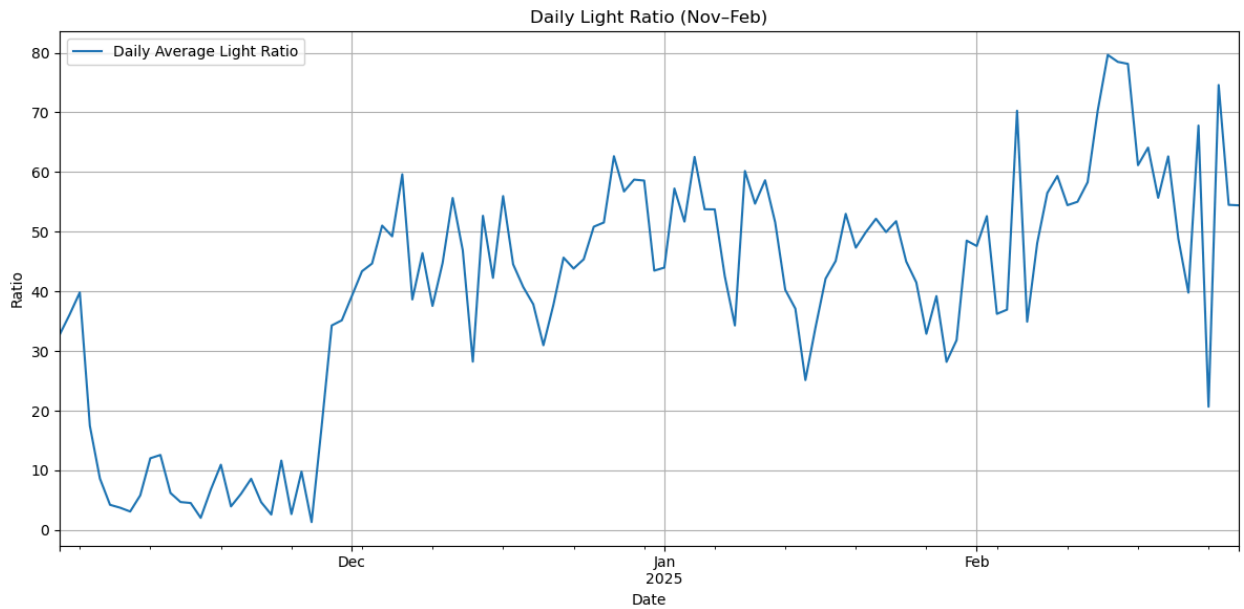

Figure A1: Daily Average Light Ratio (Nov 2024 – Feb 2025)

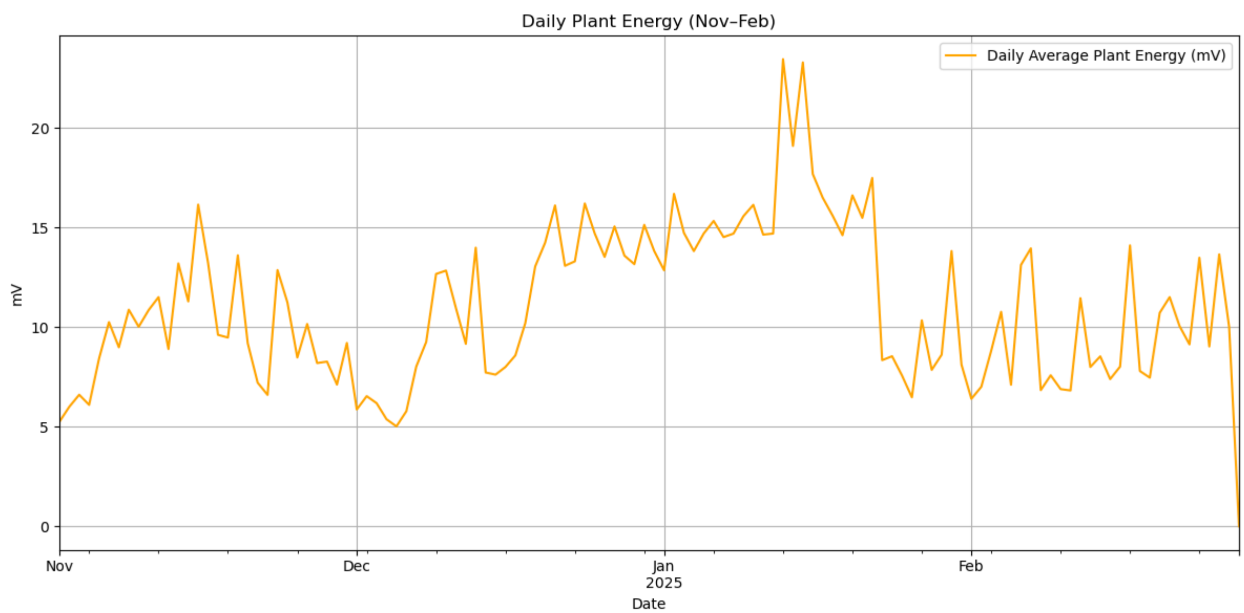

Figure A2: Daily Average Plant Electrophysiological Activity (mV)

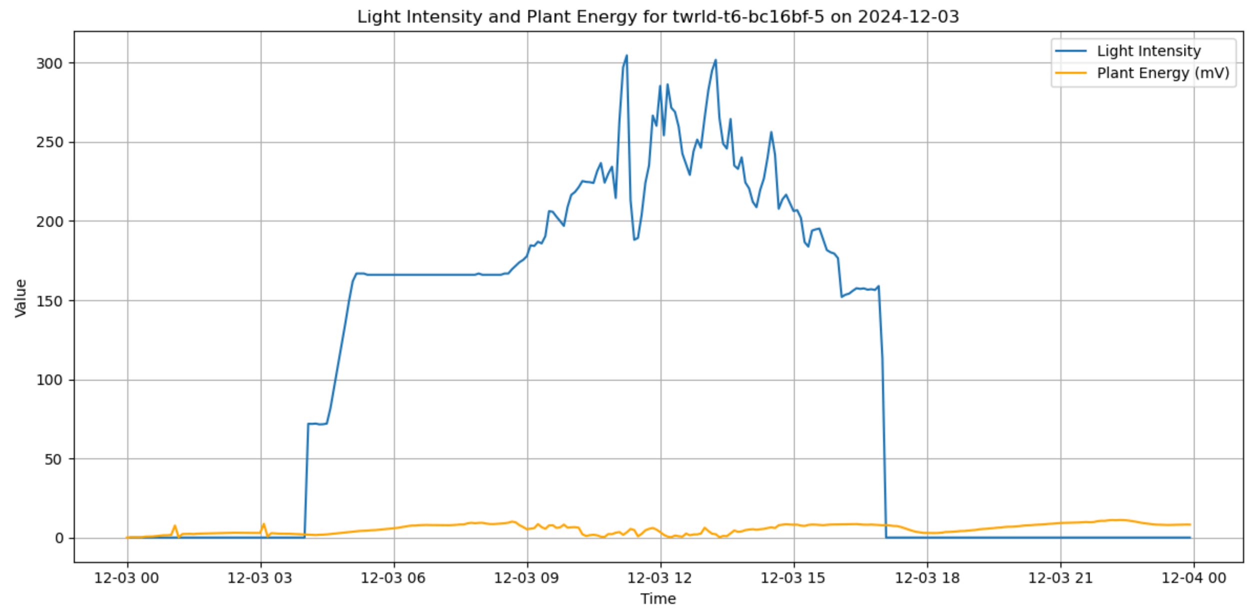

Figure A3 Sample Day Plot: Light Intensity vs. Plant mV

Appendix B

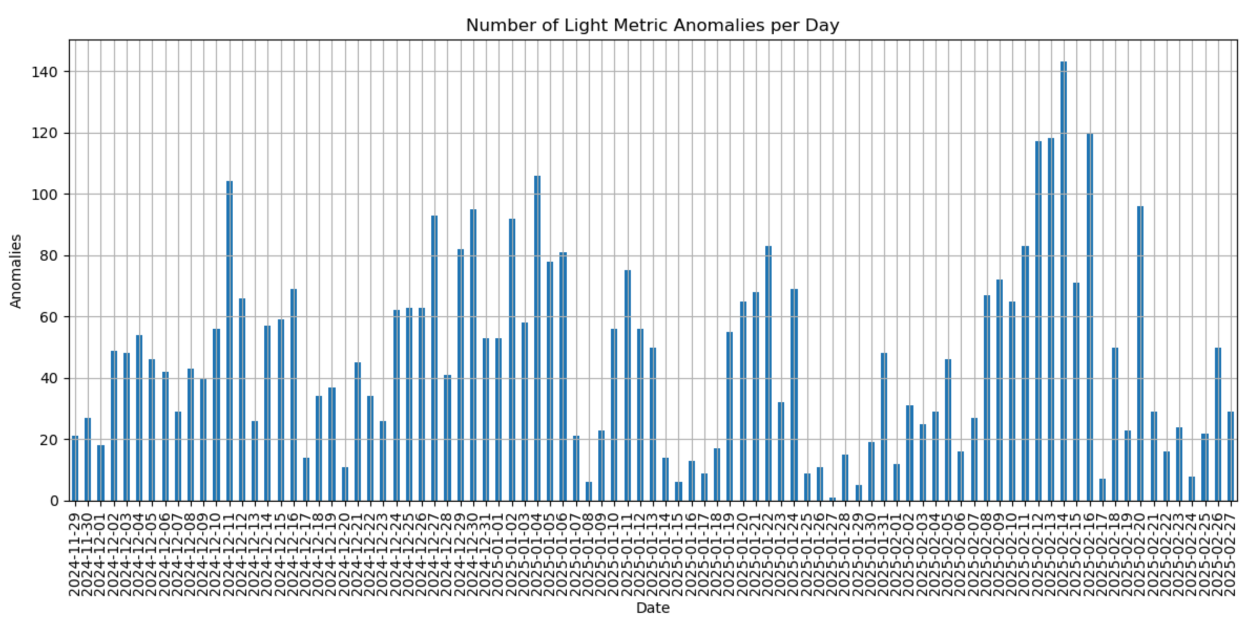

Figure B1: Number of Detected Anomalies per Day

Metric	Value
Average Light Intensity (μmol/m²/s)	145.98
Std Dev Light Intensity	25.45
Average Plant Activity (mV)	8.28
Std Dev Plant Activity	19.33
Average Light Ratio	83.24
Std Dev Light Ratio	152.04
Total LED-ON Observations	179,222
Number of Unique Plants	32
Number of LED-ON Days	91

Table B1: Descriptive Statistics of Detected LED-ON Periods

Appendix C

Timestamp	Plant ID	mV	Light Value (μmol/m²/s)	Ratio	Time
2024-12-10 05:10:00	twr1d-t6-bc168f-7	0.493187	167.5	339.683	05:10:00
2024-12-10 05:15:00	twr1d-t6-bc168f-5	0.622986	167.5	268.866	05:15:00
2024-12-10 05:20:00	twr1d-t6-bc17e2-1	0.346416	167.5	483.522	05:20:00
2024-12-10 05:25:00	twr1d-t6-bc17e2-7	0.775112	167.5	216.908	05:25:00
2024-12-10 05:30:00	twr1d-t6-bc168f-3	0.710575	167.5	233.594	05:30:00
2024-12-10 05:35:00	twr1d-t6-bc168f-2	0.273792	167.5	611.778	05:35:00
2024-12-10 05:40:00	twr1d-t6-bc168f-3	0.702615	167.5	237.178	05:40:00
2024-12-10 05:45:00	twr1d-t6-bc17e2-7	0.783778	167.5	213.788	05:45:00

2024-12-10 05:50:00	twr1d-t6-bc168f-1	0.343786	167.5	544.961	05:50:00
2024-12-10 05:55:00	twr1d-t6-bc168f-7	0.153518	167.5	1091.4	05:55:00

Table C1: First 10 rows of Timestamp-Level Records of Overlit Lighting Events

Date	Time	Total Inefficiencies	Average Light Intensity (μmol/m²/s)	Average Plant Response (mV)	Inefficiency Type
2025-02-14	00:55:00	11	188.5	0.42727	Overlit
2025-02-15	04:05:00	9	181.5	0.482447	Overlit
2025-02-13	03:40:00	9	169.1	0.385819	Overlit
2025-02-13	04:00:00	9	188.5	0.440876	Overlit
2025-02-20	04:55:00	8	181.6	0.556569	Overlit
2025-02-13	04:25:00	8	175.3	0.441165	Overlit
2025-02-13	04:20:00	8	189.5	0.493062	Overlit
2025-02-13	04:30:00	8	175.3	0.480425	Overlit

2025-02-14	00:50:00	8	188.5	0.327578	Overlit
2025-02-14	04:20:00	8	190.3	0.363042	Overlit

Table C2: First 10 rows of Aggregated Inefficiency Summary by Timestamp

Date	Total Inefficiencies	Average Light Intensity (μmol/m²/s)	Average Plant Response (mV)	Inefficiency Type
2024-12-10	50	169.88	0.46292	Overlit
2024-12-11	93	168.16	0.49238	Overlit
2024-12-12	15	167.59	0.59345	Overlit
2024-12-13	13	167.56	0.49856	Overlit
2024-12-14	15	167.50	0.56159	Overlit
2024-12-15	14	167.63	0.49875	Overlit
2024-12-25	51	167.74	0.39088	Overlit
2025-01-01	19	167.57	0.39988	Overlit
2025-01-13	2	221.35	0.53964	Overlit

2025-02-13	338	168.99	0.42639	Overlit
2025-02-14	309	189.41	0.40921	Overlit
2025-02-15	242	179.47	0.47733	Overlit
2025-02-16	174	174.69	0.34934	Overlit
2025-02-20	285	185.24	0.45604	Overlit
2025-02-26	121	191.88	0.44196	Overlit
2025-02-27	82	178.06	0.47961	Overlit

Table C3: Daily Summary of Detected Lighting Inefficiencies

Appendix D

Date	Inefficiencies	Avg Light (µmol/m²/s)	Avg mV	Type	LED Hours	SQ	Q	Predicted Light	Light Diff	% Saving	€ Saving	Actual Cost
2024-12-10	50	169.88	0.4629	Overlit	7	60	144	144.56	25.32	14.91%	22.16	148.64
2024-12-11	93	168.16	0.4924	Overlit	7.42	52	164	163.68	4.48	2.66%	4.16	155.90
2024-12-12	15	167.60	0.5935	Overlit	5	66	166	166.47	1.03	0.61%	0.61	99.45
2024-12-13	15	168.69	0.4518	Overlit	6	56	167	167.92	0.78	0.46%	0.68	147.61
2024-12-14	4	167.55	0.5610	Overlit	5.25	139	166	166.29	1.24	0.74%	0.82	109.94
2024-12-15	4	167.63	0.5616	Overlit	5.33	77	164	164.65	2.88	1.72%	2.28	132.62
2024-12-16	4	167.51	0.5223	Overlit	5.33	68	165	164.86	2.64	1.57%	1.92	122.14
2024-12-17	10	167.51	0.3964	Overlit	5.92	71	165	166.64	0.86	0.51%	0.57	111.81

2024-12-25	31	167.57	0.3999	Overlit	5.92	32	165	165.17	2.46	1.47%	1.83	123.93
2025-01-13	2	221.35	0.5396	Overlit	8.17	70	461	141.10	80.25	36.26%	81.92	225.96
2025-02-14	309	189.41	0.4092	Overlit	6.33	56	653	189.28	0.13	0.07%	0.10	149.95
2025-02-15	242	179.41	0.4737	Overlit	6.17	21	468	178.16	1.31	0.73%	1.01	138.34
2025-02-20	174	174.69	0.3493	Overlit	5	75	904	167.85	6.84	3.92%	4.28	109.18
2025-02-26	121	191.28	0.4419	Overlit	6.92	65	920	190.91	0.37	0.19%	0.32	165.38
2025-02-27	82	178.06	0.4796	Overlit	3.58	17	452	177.60	0.46	0.26%	0.20	79.76

Table D1: Daily Energy Savings from Model-Based LED Optimization on Inefficient Lighting Days