An Application of Deep Learning for Predicting Tomato Growth After Seed Irradiation

Orken Mamyrbayev
Department of Artificial Intelligence, U. Joldasbekov Institute of Mechanics and Engineering, Kazakhstan
morkenj@mail.ru

Waldemar Wójcik
Lublin University of Technology, Poland
waldemar.wojcik@pollub.pl

Sergii Pavlov
Vinnytsia National Technical University, Ukraine
psv@vntu.edu.ua

Keylan Alimhan
Department of Artificial Intelligence, U. Joldasbekov Institute of Mechanics and Engineering, Kazakhstan
keylan@live.jp

Oleksandr Poplavskyi
Kyiv National University of Construction and Architecture, Ukraine
apoplavskyi@gmail.com

Assel Aitkazina
Department of Artificial Intelligence and Big Data, Al-Farabi Kazakh National University, Kazakhstan
aitkazina.aseel@gmail.com (corresponding author)

Larysa E. Nykyforova
National University of Life and Environmental Sciences of Ukraine, Ukraine
profnikiforova@gmail.com

Nurdaulet Zhumazhan
Department of Artificial Intelligence, U. Joldasbekov Institute of Mechanics and Engineering, Kazakhstan
nurdaulet.jj02@gmail.com

Received: 16 June 2025 | Revised: 18 July 2025 | Accepted: 27 July 2025

ABSTRACT

Optimizing growth conditions for tomato crops is essential due to their sensitivity to environmental factors. Pre-sowing seed treatments, particularly irradiation with specific wavelengths, can significantly influence germination and subsequent plant development. This study investigates how five laser wavelengths (red 630 nm, green 530 nm, blue 470 nm, ultraviolet 390 nm, infrared 780 nm) and four exposure durations (15, 30, 45, 60 minutes), along with a control sample, affected two tomato cultivars (Moneymaker and Bull's Heart). A total of 42 treatment combinations were tested using tabular experimental data (numeric input features: cultivar, wavelength, exposure time), with growth outcomes, such as plant height and fruit yield,

recorded. A feedforward neural network with two hidden layers was trained to predict the final height of the plant from the seed treatment parameters. The model achieved a strong predictive accuracy (R^2≈0.92 and MSE≈9.5 cm^2) using an 80:20 train-test data split. This study demonstrates that deep learning can effectively model plant growth responses to physical seed priming and can be used to optimize treatment protocols for improved agricultural outcomes.

Keywords-tomato growth prediction; seed irradiation; deep learning in agriculture; crop yield modeling; germination enhancement; neural network modeling; seed treatment optimization; growth forecasting

I. INTRODUCTION

Tomato (Solanum lycopersicum) is cultivated worldwide as a major vegetable crop, with an estimated global production of around 186 million tonnes in 2022 [1]. The achievement of high tomato yields is crucial for global food security. However, optimizing tomato production remains challenging due to environmental stresses and the lack of predictive tools to guide cultivation decisions. This study addresses the problem of how to improve early growth and yield outcomes through seed-level interventions combined with AI-based prediction models. Key environmental factors, such as water availability, extreme temperatures, and soil salinity, are well known to limit tomato plant growth and productivity [2]. For example, exposure to suboptimal temperatures can drastically impair tomato development, as sustained night temperatures above 21°C or heat waves can inhibit fruit set and reduce yields [3]. Likewise, drought- or nutrient-poor soils and pest/disease pressures often result in suboptimal growth. To overcome these constraints, various cultivation techniques (greenhouse climate control, improved fertilization, breeding of stress-tolerant varieties) have been employed. However, innovative seed-level interventions and advanced predictive tools are emerging as promising strategies to further optimize tomato production. An interesting approach to improve crop performance is seed priming via irradiation. Seeds can be exposed to controlled doses of physical agents, such as lasers or other light sources, before sowing to stimulate physiological changes [4]. In particular, low-level laser irradiation has been reported to act as a biostimulator that can improve seed germination, seedling vigor, and stress tolerance in various plant species [5, 6]. By modulating light wavelength, intensity, and exposure duration, specific photomorphogenic responses can be triggered in seeds, potentially leading to faster germination and more robust growth [4, 7].

Tomato seeds have shown positive responses to such treatments. For instance, exposure to red laser (He-Ne 632 nm) was found to enhance tomato seed germination by approximately 10% and seedling growth by approximately 13% under optimal conditions [6]. Similarly, pre-sowing laser irradiation improved tomato seedling biomass and even final yield by about 26% [6, 8]. These improvements are attributed to laser-stimulated metabolic activation in the seed (e.g., increased activity of growth hormones such as gibberellin and reduced inhibitory abscisic acid levels) [5, 9], as well as enhanced photosynthetic efficiency in the resulting seedlings [9]. Other wavelengths have also been tested in other plants. For example, green laser light significantly increased maize seedling length (+24 cm) and leaf number in a controlled trial, while improving seedling emergence rate from 62.5% (control) to 96% [10]. Such findings suggest that appropriate light-based seed priming could similarly benefit tomato cultivars. However, effectiveness can vary with species and treatment parameters, and excessively long exposures or certain wavelengths may have negligible or even adverse effects. A lack of significant improvement in germination in caper seeds was reported after brief exposure to He-Ne laser, and similar effects were also observed in tomato seeds in some cases [11]. Thus, optimizing the irradiation protocol (wavelength and duration) for each crop and cultivar is critical.

Machine Learning (ML) and Deep Learning (DL) in agriculture have advanced rapidly in the past decade, allowing data-driven modeling of complex biological growth processes. In high-value crops, such as tomatoes, accurate yield and growth prediction models are extremely valuable for decision support. Traditional empirical models often struggle to capture the non-linear interactions of genotype, environment, and management. However, DL models can learn patterns from large datasets, making them ideal for crop prediction tasks. In addition, optimization methods, such as Taguchi and TOPSIS, have also been applied in agricultural engineering to improve system efficiency and improve experimental design strategies [12]. Recent studies have demonstrated the effectiveness of DL for diverse agricultural applications, including detecting bruises in plums using PlmNet [13], pear leaf disease classification with PL-DenseNet [14], alfalfa variety identification via custom leaf datasets [15], and crop classification using transfer learning. This study aimed to address the problem of predicting tomato growth responses to seed irradiation treatments using a DL approach to optimize pre-sowing protocols. The contribution of this study lies in the integration of experimental agronomic data with neural network modeling to predict growth based on irradiation parameters and cultivar-specific responses.

II. DEEP LEARNING FOR CROP GROWTH PREDICTION

The advent of DL has revolutionized predictive modeling in agriculture. Unlike rule-based or linear models, deep neural networks learn complex feature representations from data [16], enabling them to capture non-linear interactions between genotype, environment, and management factors that influence plant growth. In the context of tomato cultivation, DL has been applied to tasks ranging from yield prediction to disease detection, with a comprehensive review of crop yield prediction methods in [17]. For tomato yield specifically, both regression-type models and computer vision models have been tested. One approach uses environmental and physiological time-series data as input to RNNs. For instance, a CNN-RNN model was developed to process climate sensor readings and growth observations in a greenhouse and output weekly tomato yield predictions [17]. This model, after training on historical crop data, achieved greater accuracy than traditional statistical models, effectively learning the temporal patterns of tomato development under greenhouse conditions.

Another approach involves image-based yield estimation, where an improved YOLOv4 CNN was combined with a DeepSORT tracking algorithm to count tomatoes from video frames captured by a monitoring robot [18]. By detecting tomato flowers, unripe (green) and ripe (red) fruits with high precision (93-98% across stages), this method could predict yield by tracking the progression of counted fruits. This demonstrates how DL can automate fruit counting, a key component of yield prediction. Similarly, in [19], a faster R-CNN was applied to detect heavily occluded tomato fruits in greenhouse images, achieving an R^2 of 0.87 between the model's fruit count predictions and the ground truth yields. This high accuracy, even under occlusion, indicates the robustness of deep CNNs for plant vision tasks. Beyond yield count, DL has been used to predict continuous growth variables, including plant disease detection. In [20], YOLOv5 was used to detect diseases in tomato plant images, achieving an accuracy of 93%. In [21], a Random Forest (RF) model used features extracted from weekly UAV imagery to predict final tomato yield under control and salt-stress conditions. This study showed that structural image features (such as canopy area) were the most important predictors, and the model could estimate yield in both normal and stress treatments with good accuracy (error rates on the order of 10-15%). Although RF is not a deep neural network, it can serve as a benchmark. Nowadays, CNNs coupled with temporal models are increasingly used to process such imagery data. For example, a Transformer-based segmentation model (SegLoRA) was introduced to simultaneously assess tomato fruit ripeness and counts for yield prediction [17]. By leveraging attention mechanisms, this model segmented images into ripe and unripe fruit masks, leading to precise yield estimates, allowing farmers to forecast harvest timing. The overall trend in the literature is a move toward integrated DL frameworks that can incorporate heterogeneous data (images, sensors, management actions) to predict crop outcomes.

In summary, DL has been proven effective in modeling tomato growth and yield. However, its application in the analysis of experimental agronomic treatments (such as seed irradiation) is still nascent. There is a clear gap in the use of such AI tools to extrapolate the effects of novel treatments between conditions. This work aimed to address this by using a neural network to learn from experimental irradiation data. If successful, this approach could help optimize seed treatment protocols (by predicting outcomes for untested scenarios) and demonstrate a blueprint to marry agronomic experiments with AI modeling for improved decision-making.

III. MATERIALS AND METHODS

This study selected two tomato cultivars, Moneymaker and Bull's Heart, due to their distinct growth habits and popularity in cultivation. Moneymaker is an indeterminate early-maturing cultivar characterized by vigorous vine growth and high yield potential. Plants typically reach 1.5-1.8 m in height under greenhouse conditions and bear clusters of 7-12 medium-sized (90-100 g) red fruits. It is known for its extended fruiting until frost and can yield up to approximately 9 kg per plant under optimal conditions. Bull's Heart is a large-fruited, semi-determinate heirloom variety, notable for its beefsteak-type fruits that can weigh 500-800 g each. Bull's Heart plants have a sprawling growth habit with fewer leaves and typically set 5-6 fruit clusters before terminating growth at approximately 1.5-1.8 m tall. In open-field cultivation, it yields around 3.5-5 kg per plant, whereas in greenhouses, yields of 8-12 kg per plant have been reported. The two cultivars also differ in tolerance to stress. Moneymaker is considered hardy and adaptable to outdoor conditions, while Bull's Heart, with its larger fruit load, often benefits from the protection of a greenhouse for maximum yield. Using these two genotypes allowed us to compare how a standard commercial type (Moneymaker) and a large-fruited type (Bull's Heart) respond to seed irradiation treatments. All seeds used were of high quality and uniform size, sourced from a certified supplier. Before treatments, seeds were stored dry at room temperature. To isolate the effects of the experimental light treatments, no chemical seed coatings or primers were applied.

For germination and seedling growth, a universal bio-soil substrate (Bogatyr potting mix) was used, which is a sterile nutrient-enriched growing medium. The substrate is composed of high-moor peat, lowland peat, sand, dolomitic limestone (for pH buffering), a mineral NPK fertilizer, and beneficial soil bacteria (Bacillus subtilis). This mix provides an initial N-P-K content of approximately 300-30-400 mg/L and a pH of 5.5-6.5. The presence of Bacillus subtilis in the substrate helps suppress soil-borne pathogens and promote root growth, ensuring that any differences in germination are more likely due to seed treatment than soil disease. The substrate was lightly moistened with distilled water before sowing and was contained in plastic germination trays. After sowing, the seeds were allowed to germinate under indoor room conditions for approximately two weeks. The room temperature was maintained at approximately 22-24°C during the day and not below 18°C at night, with a relative humidity of approximately 60%.

A. Seed Irradiation Treatments

Before sowing, seeds from each cultivar were subjected to one of several light irradiation treatments or a control. A custom irradiation setup was developed, using coherent laser sources of different wavelengths. Four laser emitters (each a diode laser of specific wavelength) were mounted on a wooden plank and aligned perpendicular to a reflective base where the seeds were placed. A piece of food-grade aluminum foil served as the reflecting surface to ensure that seeds received light from all directions. The collimating optics on each laser produced a consistent circular beam spot of approximately 10 mm diameter at the seed plane. This spot size covered the seeds (placed in Petri dishes) so that all seeds in a dish were illuminated evenly. The power output of each laser was adjusted to deliver low-level intensity (on the order of a few mW/cm²), enough to stimulate but not thermally damage the seeds. Five spectral treatments corresponded to different laser wavelengths, plus an unirradiated control (Figure 1):

1. Red (630 nm): Coherent red laser light (near He-Ne laser line).

2. Green (530 nm): Green laser (frequency-doubled Nd:YAG or diode).
3. Blue (470 nm): Blue laser diode.
4. Ultraviolet (UV, 390 nm): near-UV laser diode.
5. Infrared (IR, 780 nm): a laser in the far-red/near-infrared range.
6. Control (no irradiation) seeds were kept under identical conditions but without any light exposure.

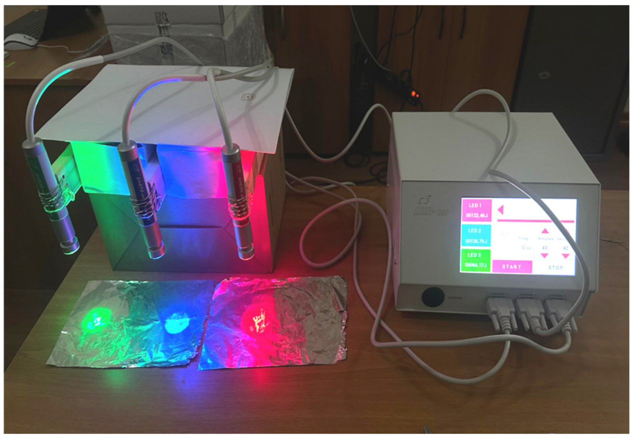

Fig. 1. Bar with laser emitters.

The seeds were placed approximately 10 cm below the lasers on the foil surface. The intensity was calibrated such that none of the treatments caused heating of the seed (verified by not exceeding ~25°C on the foil). For each cultivar and light wavelength, a batch of seeds was treated (n≈12 seeds per batch). To examine the effects of duration of exposure, the seeds were irradiated for varying amounts of time. Specifically, at 15-minute intervals, a subset of seeds was removed from light exposure after 15 min, 30 min, 45 min, and 60 min of irradiation. At each 15-minute interval, three seeds per cultivar were removed (so that by 60 minutes, four subsets of seeds had received 15, 30, 45, and 60 minutes of exposure, respectively) (Figure 2). This method allowed for a controlled increase in dose while using the same initial batch of seeds. All seeds removed were immediately planted in moist potting soil in labeled containers, indicating their cultivar, wavelength, and exposure time. In total, considering two cultivars times five wavelengths times 4 exposure times, plus controls, the experimental design initially comprised 2×5×4+2 (controls) = 42 treatment combinations. The seeds in each pot were sown equidistantly. After sowing, all pots were kept under identical conditions for germination, as described above. Each cultivar was treated separately but analogously under all these conditions. In addition to vegetative parameters, data were also collected on fruit yield following seed irradiation. Specifically, for the Moneymaker cultivar grown under field conditions, the number of fruits per plant was recorded at the time of the first harvest. The highest fruit yield was observed in the group subjected to Infrared (IR) laser treatment (780 nm) for 60 minutes. A moderate increase in fruit number was also noted in the IR 30-minute treatment (Figure 3).

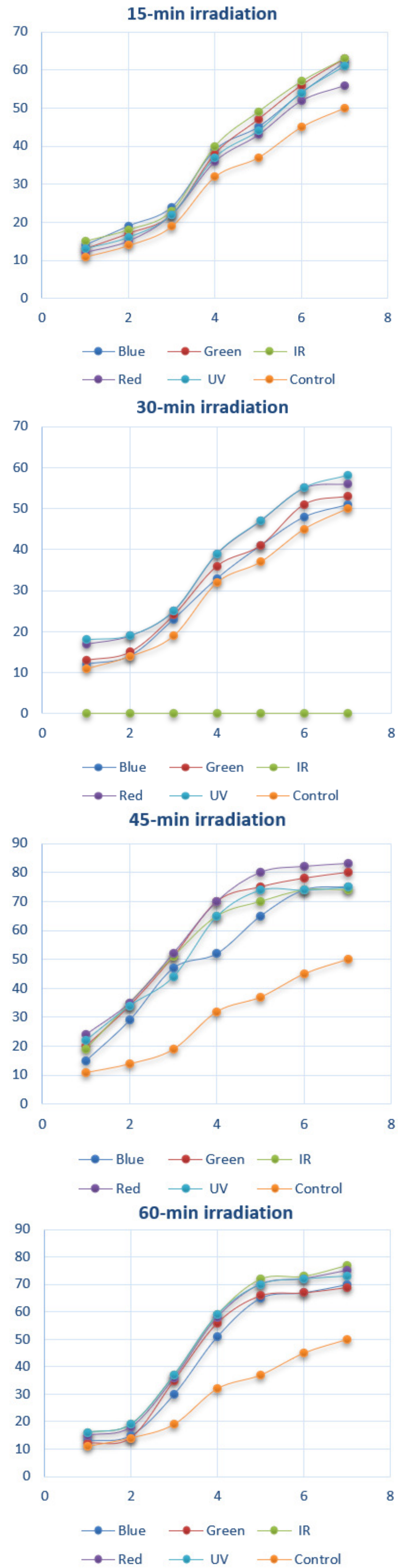

Fig. 2. Effect of seed irradiation with different wavelengths (Blue, Green, IR, Red, UV) and control (no irradiation) on plant growth dynamics.

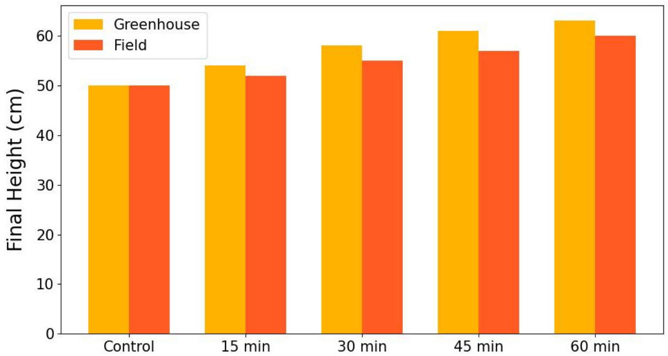

Fig. 3. Effect of seed irradiation on plant height.

After approximately two weeks of indoor germination, the seedlings were transplanted to their respective cultivation environments. The open-field trial was carried out in Almaty Region (Almaty), while the greenhouse trial took place in Merke, Zhambyl Region. Transplantation was carried out in early June, with the field group planted on 3 June under natural outdoor conditions and the greenhouse group relocated on 5 June to a controlled greenhouse facility. In particular, the weather in Almaty during May was cool and rainy (daytime ~17-19 °C, nighttime ~9-12°C), while early June provided warmer and more stable conditions (daytime ~24-27 °C) under protection from rainfall. As the plants matured, key growth parameters were measured: final plant height (cm) and fruit count per plant. The height of the plant was measured from the surface of the soil to the apex of the plant using a measuring tape, and the number of fruits per plant was manually counted at the time of the first harvest. All observations were recorded by hand in field notebooks and later compiled into an electronic spreadsheet for subsequent analysis. Table I provides a small sample of the dataset.

TABLE I. SAMPLE RAW DATASET ENTRIES FOR THE TOMATO GROWTH EXPERIMENT

Sample ID	Cultivar	Laser Wavelength (nm)	Exposure Time (min)	Environ-ment	Final height (cm)	Fruit count
1	Moneymaker	630	15	OF	52	8
2	Moneymaker	630	15	GH	60	10
3	Moneymaker	780	60	OF	75	14
4	Moneymaker	780	60	GH	82	12
5	Bull's Heart	470	30	OF	58	5
…	…	…	…	…	…	…
4599	Bull's Heart	470	30	GH	64	7

IV. DEEP LEARNING MODEL PERFORMANCE

Of the total dataset (~4600 samples), approximately half of the samples correspond to greenhouse-grown plants and the other half to open-field-grown plants (on the order of 2,300 samples per condition). The train-test split (80:20) was performed in a stratified manner concerning the cultivation method, ensuring that both the training and test sets maintained the same greenhouse vs. open-field sample ratio as the overall dataset. In this way, the model was evaluated fairly on each type of growing condition, without bias toward one environment.

A feedforward neural network model was trained to predict the final height of the plant according to the seed treatment parameters and conditions. The network's architecture consisted of an input layer corresponding to the experimental factors, two fully-connected hidden layers, and a single output neuron for regression. Each hidden layer was kept relatively small to match the limited dataset size (~4600 samples) and minimize overfitting. For example, the final architecture used two hidden layers with 460 and 230 neurons, respectively (after some trial-and-error tuning), each employing a Rectified Linear Unit (ReLU) activation to introduce non-linearity. The output layer produced a single continuous value (predicted plant height in cm) with linear activation, appropriate for a regression task. This structure allowed the model to learn a mapping from seed irradiation settings to the expected growth outcome. The experimental input variables were carefully encoded for the neural network. Categorical factors such as cultivar, light wavelength, and growth environment were transformed by one-hot encoding into binary input features. For instance, two binary nodes represented the cultivar (one for each cultivar, with active=1 indicating the sample's cultivar), and five features encoded the irradiation type (e.g., IR=1 for infrared treatment, while all wavelength indicators equal to 0 indicated the control without light exposure). The exposure duration (irradiation time in minutes) was a numerical input feature, which was normalized (scaled to a range of 0-1) to ensure that it was on a scale comparable to the other inputs. In total, each plant sample was described by a feature vector of eight dimensions (cultivar, environment, five wavelength indicators, and exposure time), which served as input to the neural network. This encoding ensured that the model used qualitative and quantitative factors during training.

The Adam optimization algorithm was utilized in the training, with a learning rate of 0.001, chosen for its efficiency in converging with small datasets and its ability to adaptively adjust the learning rate. A mini-batch size of 8 worked well, providing a balance between noisy gradient updates and efficient learning given the dataset size. The network was trained for a maximum of 40 epochs. However, the error in a validation split was monitored to employ early stopping if the validation loss stopped improving (which helped avoid overtraining the model). The model's complexity (number of layers/neurons) was intentionally kept low given the small dataset, as a form of regularization to reduce overfitting.

Mean Squared Error (MSE) was used as the loss function during training and as the primary performance metric for evaluation. MSE is appropriate for this regression task because it directly quantifies the average squared difference between predicted and actual plant heights (in cm^2), penalizing larger errors more heavily. In addition, the coefficient of determination (R^2) was calculated on the test results to gauge the proportion of variance in plant height explained by the model. R^2 is a standard metric for regression models, providing a more interpretable measure of goodness-of-fit (with $R^2=1$ indicating perfect prediction and $R^2=0$ indicating no better accuracy than the mean). These metrics complement each other: MSE conveys the error magnitude in physical units (useful for agronomic interpretation, since an error in cm can be related to plant growth), while R^2 indicates the model's explanatory power in relative terms. Both were chosen to give a clear sense of predictive accuracy and reliability of the model in the context of continuous growth prediction.

After training, the neural network demonstrated good predictive accuracy on the hold-out test data. Figure 4 shows the predicted vs. actual plant heights for the test set (averaged results for each set), where nearly all points lie close to the 1:1 diagonal line, indicating that the model's predictions closely matched the true observed heights. Quantitatively, the model achieved a test MSE of approximately 9.5 cm². This corresponds to a Root Mean Squared error (RMSE) of about 3.1 cm, meaning that, on average, the model's predictions were within about three centimeters of the actual plant height. Given that tomato plants ranged from roughly 50 to 80 cm high, an error of ~3 cm is relatively small (on the order of 5% of the plant's height). R^2 was about 0.92 on the test set, indicating that ~92% of the variance in final height was explained by the model's learned relationship. In practice, an R^2 of 0.92 suggests that the network captured almost all the systematic variation in height outcomes due to the seed treatments and other inputs, with only a small portion of variation left unexplained (likely due to biological randomness or factors not included in the inputs). For example, the model predicted a final height of ~60 cm for one Moneymaker plant that actually reached 56 cm (under a red-light treatment), and ~78 cm for a Bull's Heart plant that grew to 75 cm (with an IR treatment), illustrating how the predictions were very close to the measured values. The maximum prediction error observed on any test sample was about 5-6 cm, and most predictions were accurate to within a few centimeters. This high accuracy on the independent test set confirms that the trained network successfully learned the underlying mapping from irradiation parameters and cultivar to plant growth outcome.

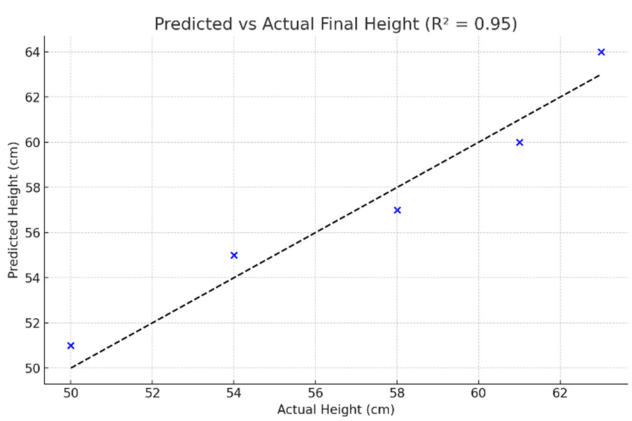

Fig. 4. Effect of seed irradiation on plant height.

Beyond final heights, the model's performance in predicting intermediate growth stages was examined to further validate its behavior. When training and testing the network on plant heights at earlier time points (e.g., mid-season height measurements), the accuracy was slightly lower, with R^2 around 0.80 for those intermediate predictions. This reduced R^2 is understandable, as early-stage growth is more variable and can be influenced by transient environmental factors (for instance, an unexpected cold spell in open-field trials caused temporary stunting in some seedlings). Such short-term effects were not fully captured by the input features, which mainly describe treatment and initial conditions. However, even in these intermediate cases, the model's predictions reflected the correct relative trends: plants from more stimulative treatments (such as longer IR exposure) were predicted to be taller than those from control or less effective treatments, albeit with less absolute precision than for final height. This suggests that the neural network was indeed learning the intended cause-and-effect signals from the data, but its predictive confidence improves as the cumulative effects manifest in the final growth outcomes.

The weights of the features and contributions to the network were analyzed to understand the learned relationships of the model. Although the neural network functions as a black box to some extent, examining the learned weights of the first layer provides insight into which inputs were the most influential. The network assigned the highest positive weight to the IR wavelength feature, aligning with the experimental result that infrared irradiation had the strongest effect on promoting height growth. The cultivar input was also a critical factor: the model learned to differentiate between the two tomato cultivars, recognizing that Bull's Heart plants generally grow taller than Moneymaker plants when all other conditions are equal. This is evidenced by the network's bias for the Bull's Heart indicator being positive. Specifically, if the input cultivar is Bull's Heart, the predicted height is higher by a certain amount, reflecting the cultivar's inherent growth potential.

Additionally, by including the environment as one of the inputs, the model captured the performance difference between the greenhouse and field conditions. The network tended to predict slightly higher heights for greenhouse-grown plants compared to field-grown plants (given the same seed treatment and cultivar), which is consistent with the empirical data showing that the plants in the greenhouse had more favorable growth. These internal weight patterns demonstrate that the neural network not only fits the data well but also learned meaningful relationships that make biological sense (e.g., IR light has a positive effect, Bull's Heart is taller, and greenhouse conditions yield taller plants). In essence, the model's behavior echoed the key findings of the experiment, providing confidence that it was generalizing the treatment effects rather than just memorizing individual data points.

Throughout model development, careful measures were taken to monitor and prevent overfitting. During training, the losses on the training set and a hold-out validation set (in this case, the 20% test split used as a proxy for validation during development) were monitored at each epoch. The training loss decreased steadily and eventually plateaued, while the validation loss remained closely aligned with the training loss (differing only slightly at convergence). This parallel behavior indicated that the model was not overfitting significantly; if it had been, the training loss would have continued decreasing while the validation loss would have started to increase (diverging curves). In addition, early stopping was employed, meaning that if the validation MSE did not improve for a number of epochs, training was halted to avoid over-training.

Furthermore, as previously mentioned, the network's architecture was intentionally kept simple, and the total number of training epochs was limited, which inherently constrains the model's capacity to memorize the training data. These precautions helped ensure that the model's performance on the training and unseen test data was comparable. Indeed, the final training MSE was on the same order as the test MSE (within about 10-15% of it), and the training R^2 reached about 0.95, which is only modestly higher than the test R^2 of 0.92. This suggests the model generalized well to new samples and did not purely overfit the training examples.

This study did not perform a k-fold cross-validation. Although k-fold cross-validation (repeatedly training the model on subsets of the data and validating on the remaining portion) can provide a more robust estimate of model performance on small datasets, this study opted for a simpler hold-out validation strategy. The focus was on demonstrating the feasibility of using DL to capture irradiation effects, rather than exhaustive model benchmarking. Using an 80:20 split, a reasonable number of samples was preserved for training so that the model could learn the underlying patterns, and an independent test set was maintained to evaluate generalization. This approach is a common compromise in proof-of-concept studies.

Finally, limited hyperparameter tuning was carried out to refine the model's performance. Instead of an automated grid search or Bayesian optimization, which would be difficult to justify with so few data points, this study relied on manual trial-and-error exploration. Using significantly larger layers led to overfitting (the model would perfectly fit training data but then exhibit higher error on validation data), whereas much smaller layers underfit (failed to capture some of the nonlinear relationships, yielding higher error on both training and validation). The learning rate was adjusted in small increments (evaluating 0.0005, 0.001, 0.005), observing that 0.001 provided stable and fast convergence without causing oscillation in loss. The batch size (8) and the number of epochs (up to 40 with early stopping) were reached similarly by trying a few alternatives and choosing the settings that gave the lowest validation loss. This study did not tune every parameter exhaustively, but this practical approach allowed us to obtain a well-performing model given the constraints. All hyperparameters chosen are documented (optimizer = Adam, learning rate = 0.001, batch size = 8, two hidden layers with 460 and 230 neurons, ReLU activations, etc.) to ensure reproducibility.

In future work with a larger dataset, a more systematic hyperparameter optimization and possibly the use of cross-validation would be advisable to further validate and potentially improve the model. For now, the chosen configuration proved to be adequate for capturing the key effects in the data without overfitting. In summary, the DL model performed well in predicting tomato plant height outcomes based on seed irradiation treatments, achieving high accuracy (explaining over 90% of the variance in height) on the test data and keeping prediction errors to within a few centimeters. The model architecture and training regime were intentionally kept straightforward to suit the small dataset, and the results indicate that the network effectively learned the important treatment-response patterns (e.g., the growth benefits of infrared and green light, and the inherent differences between cultivars and environments). These findings illustrate the promise of applying even relatively simple neural networks in agronomic studies, with careful design and validation, to derive predictive insights from experimental data.

V. LIMITATIONS AND FUTURE DIRECTIONS

This study combined an agronomic experiment on seed irradiation with advanced modeling to dissect the results on tomato growth. Several important findings emerge. First, pre-sowing laser irradiation of seeds can enhance subsequent plant growth, but the effects depend on the light wavelength, exposure duration, and cultivar. The results showed that longer exposures (45-60 min) to red, green, blue, and IR light improved germination success and seedling vigor compared to shorter exposures. In particular, 1 hour of infrared (780 nm) treatment stood out as a potent stimulant, as IR-irradiated seeds offered the tallest plants and highest yields under favorable conditions. This aligns with previous studies suggesting that infrared or red light can accelerate germination and growth by influencing phytochrome-mediated pathways. Phytochrome, which absorbs red/far-red, likely perceives IR treatment and triggers growth-promoting signals similar to those in shade-avoidance responses (where plants elongate and grow faster under far-red light). In addition, IR radiation might have gently warmed the seeds, helping to germinate (infrared often has a slight heating effect, although the low intensity used in this study was intended not to overheat). The net result was an early acceleration of growth. Green laser exposure also consistently benefited both cultivars, corroborating reports that green light can penetrate seeds and tissues effectively and enhance seedling performance.

Blue and UV light, which have higher energy, did not show dramatic improvements but still produced plants somewhat taller than controls. Blue light receptors (cryptochromes) might induce protective or compact growth. Interestingly, blue-treated plants were sturdy and dark green, suggesting high chlorophyll, even if height gains were moderate. UV at 390 nm is close to the UV-A range. Low UV can induce stress-protective phenolics, but too much can stunt growth. The UV 60 min treatment did not harm the seedlings, and these plants grew on par with or just above the control, indicating a slight benefit or at least no negative effect from the UV dose used. In this case, 60 minutes was roughly optimal for most wavelengths (which is a much larger dose than in [6], possibly due to the lasers in this study being lower power or the seeds needing longer to absorb enough photons). The observed cultivar-specific responses are particularly interesting. Bull's Heart, the larger variety, benefited greatly from red laser treatment in the greenhouse. Those plants were the tallest of all, even more than the IR-treated Bull's Heart. Moneymaker, on the contrary, saw more modest gains from red (only ~10% taller than control) but thrived under IR. This could be related to genetic differences in photoreceptor sensitivity or seed coat properties.

It is important to acknowledge the limitations of this study. Although the observed differences were large enough to be significant even at that sample size, experimenting with more seeds and replicates would bolster confidence. Additionally, since the experiment was conducted in a single season at one location, environmental conditions unique to that season may have influenced the results. Repeating the field vs. greenhouse comparison in a different year or location would be informative to see if the same patterns hold. The DL model, although effective, was trained on a very limited dataset. Thus, its predictive use beyond this dataset should be viewed as a proof of concept. The model was kept simple to reduce the risk of overlearning. In a future study, collecting a larger dataset (perhaps including image-based features and multi-time-point data) would allow training of more sophisticated networks (such as LSTMs as in other studies) to predict growth trajectories in a truly generalizable way.

In terms of broader implications, the findings support the idea that physical seed priming (such as laser irradiation) can be a sustainable agricultural practice to improve crop performance without chemical inputs. Especially in stress-prone regions or organic farming, this technique could improve seedling establishment and yield. The DL model illustrates how modern AI can help in decision support, helping tailor these interventions to specific contexts (cultivar/environment). This synergy between traditional agronomy and AI is an emerging paradigm for precision agriculture.

VI. CONCLUSION

This study confirms that laser-based seed irradiation is an effective strategy to improve tomato growth and productivity. Seeds exposed to certain laser wavelengths, notably infrared (780 nm) and green (530 nm), grew into more vigorous plants than the untreated controls. These irradiated seedlings developed a higher final height and produced higher fruit yields, demonstrating a measurable increase in productivity. Importantly, similar improvements were observed in both examined tomato cultivars, suggesting that laser priming can robustly enhance early growth across different genetic backgrounds when applied under suitable conditions. Moreover, these benefits were achieved without any chemical input or detrimental effects on plant health, underscoring the method's promise as a sustainable and safe means of boosting crop performance. The DL model proved to be a reliable tool for optimizing seed irradiation practices. The trained neural network achieved high predictive accuracy in relating seed treatment parameters to final plant height, indicating that it successfully captured the complex relationships between laser wavelength, duration of exposure, and growth outcomes. Its strong performance suggests that such a model can be trusted to identify the most effective irradiation settings, effectively pinpointing optimal seed treatment conditions to maximize tomato plant growth. In practice, the model's predictions could serve as a decision-support resource, guiding growers in selecting the irradiation parameters that are most likely to benefit a given tomato cultivar and cultivation environment.

Unlike previous studies that are mainly based on environmental sensor data, UAV imagery, or image-based yield estimation for tomato growth prediction or disease detection [17-23], this study introduces a novel integration of physical seed-level interventions and DL. Although prior work has successfully applied CNNs, RNNs, and Transformer-based models to analyze environmental conditions or fruit images, they have not incorporated the biological effects of pre-sowing treatments as predictive variables. This work fills this gap by demonstrating that the parameters of seed irradiation, such as wavelength and exposure duration, can be effectively modeled to predict the final height of the plant. This not only broadens the scope of AI applications in agriculture but also introduces a new data dimension that has been largely overlooked in earlier research.

Beyond the specific case of tomatoes, the approach of combining physical seed treatments with AI-based predictive modeling has promising implications for agriculture in general. The successful results obtained here indicate that similar DL frameworks could be applied to other crop species to predict and improve their growth trajectories. Using neural networks to analyze how various seed priming techniques affect different plants, researchers and practitioners may be able to achieve comparable gains in germination rates and yields in a range of crops. In essence, the AI-driven methodology of this study offers a blueprint to improve crop productivity beyond tomatoes, highlighting the potential of data-driven models to inform and improve agricultural practices. Looking ahead, a key direction for future research is to develop a more universal predictive model that transcends the specific cultivars and conditions studied here. By incorporating data from a wider range of tomato varieties and even other crop species, a neural network could potentially capture broader patterns in how seeds respond to irradiation, resulting in a model capable of providing reliable growth predictions across diverse genetic backgrounds and environmental contexts. Ultimately, such a generalized AI tool could guide seed irradiation protocols and other agronomic decisions on a much wider scale.

ACKNOWLEDGMENT

This research has been funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP 19677201).

REFERENCES

[1] "FAOSTAT: Crops and livestock products, Tomatoes," *Food and Agriculture Organization of the United Nations (FAO).* https://www.fao.org/faostat/en/#data/QCL.

[2] S. Murad *et al.*, "Strategies to overcome drought stress for improving plant growth under sustainable agriculture," *Soil and Environment*, vol. 43, no. 2, pp. 148–159, Dec. 2024, https://doi.org/10.25252/SE/2024/243179.

[3] S. P. Mohammed *et al.*, "Heat stress in tomato plants: current challenges and future directions for sustainable agriculture," *New Zealand Journal of Crop and Horticultural Science*, pp. 1–25, Dec. 2024, https://doi.org/10.1080/01140671.2024.2432624.

[4] A. Aitkazina, O. Mamyrbayev, K. Alimhan, D. Oralbekova, and N. Zhumazhan, "Infrared Laser Processing in Seed Treatment: A Biotechnical Approach," *Journal of Electrical Systems*, vol. 20, no. 10s, pp. 6241–6247, Jul. 2024.

[5] Samiya, S. Aftab, and A. Younus, "Effect of low power laser irradiation on bio-physical properties of wheat seeds," *Information Processing in Agriculture*, vol. 7, no. 3, pp. 456–465, Sep. 2020, https://doi.org/10.1016/j.inpa.2019.12.003.

[6] C. F. Rivera-Talamantes, A. Michtchenko, G. González-López, A. V. Budagovsky, D. C. Coyac, and J. Acosta, "Influence of pre-sowing red laser irradiation of tomato seeds on the initial plant development, salinity

stress tolerance, and harvest yield," *Emirates Journal of Food and Agriculture*, Oct. 2022, https://doi.org/10.9755/ejfa.2023.v35.i1.2934.

[7] Y. P. Chen, M. Yue, and X. L. Wang, "Influence of He–Ne laser irradiation on seeds thermodynamic parameters and seedlings growth of Isatis indogotica," *Plant Science*, vol. 168, no. 3, pp. 601–606, Mar. 2005, https://doi.org/10.1016/j.plantsci.2004.09.005.

[8] M. Hasan, M. M. Hanafiah, I. H. H. Alhilfy, and Z. Aeyad Taha, "Comparison of the effects of two laser photobiomodulation techniques on bio-physical properties of *Zea mays* L. seeds," *PeerJ*, vol. 9, Jan. 2021, Art. no. e10614, https://doi.org/10.7717/peerj.10614.

[9] P. S. Swathy, K. R. Kiran, M. B. Joshi, K. K. Mahato, and A. Muthusamy, "He–Ne laser accelerates seed germination by modulating growth hormones and reprogramming metabolism in brinjal," *Scientific Reports*, vol. 11, no. 1, Apr. 2021, Art. no. 7948, https://doi.org/10.1038/s41598-021-86984-8.

[10] A. C. Hernández, C. A. Carballo, A. Artola, and A. Michtchenko, "Laser irradiation effects on maize seed field performance," *Seed Science and Technology*, vol. 34, no. 1, pp. 193–197, Apr. 2006, https://doi.org/10.15258/sst.2006.34.1.21.

[11] M. L. Foschi, M. Juan, B. Pascual, and N. Pascual-Seva, "Influence of Lighting and Laser Irradiation on the Germination of Caper Seeds," *Agriculture*, vol. 12, no. 10, Oct. 2022, Art. no. 1612, https://doi.org/10.3390/agriculture12101612.

[12] E. Amri, Y. Gulzar, A. Yeafi, S. Jendoubi, F. Dhawi, and M. S. Mir, "Advancing automatic plant classification system in Saudi Arabia: introducing a novel dataset and ensemble deep learning approach," *Modeling Earth Systems and Environment*, vol. 10, no. 2, pp. 2693–2709, Apr. 2024, https://doi.org/10.1007/s40808-023-01918-9.

[13] Y. Gulzar and Z. Ünal, "Time-Sensitive Bruise Detection in Plums Using PlmNet with Transfer Learning," *Procedia Computer Science*, vol. 257, pp. 127–132, 2025, https://doi.org/10.1016/j.procs.2025.03.019.

[14] Y. Gulzar and Z. Ünal, "Optimizing Pear Leaf Disease Detection Through PL-DenseNet," *Applied Fruit Science*, vol. 67, no. 1, Feb. 2025, Art. no. 40, https://doi.org/10.1007/s10341-025-01265-2.

[15] Y. Gulzar, Z. Ünal, T. Kızıldeniz, and U. M. Umar, "Deep learning-based classification of alfalfa varieties: A comparative study using a custom leaf image dataset," *MethodsX*, vol. 13, Dec. 2024, Art. no. 103051, https://doi.org/10.1016/j.mex.2024.103051.

[16] O. Poplavskyi *et al.*, "High-performance information technology for processing large datasets and biomedical images to improve the accuracy of computer-aided decision support systems," in *Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2024*, Lublin, Poland, Dec. 2024, Art. no. 25, https://doi.org/10.1117/12.3057444.

[17] Y. Wang *et al.*, "Progress in Research on Deep Learning-Based Crop Yield Prediction," *Agronomy*, vol. 14, no. 10, Oct. 2024, Art. no. 2264, https://doi.org/10.3390/agronomy14102264.

[18] Y. Ge *et al.*, "Tracking and Counting of Tomato at Different Growth Period Using an Improving YOLO-Deepsort Network for Inspection Robot," *Machines*, vol. 10, no. 6, Jun. 2022, Art. no. 489, https://doi.org/10.3390/machines10060489.

[19] Y. Mu, T. S. Chen, S. Ninomiya, and W. Guo, "Intact Detection of Highly Occluded Immature Tomatoes on Plants Using Deep Learning Techniques," *Sensors*, vol. 20, no. 10, May 2020, Art. no. 2984, https://doi.org/10.3390/s20102984.

[20] R. Rajamohanan and B. C. Latha, "An Optimized YOLO v5 Model for Tomato Leaf Disease Classification with Field Dataset," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12033–12038, Dec. 2023, https://doi.org/10.48084/etasr.6377.

[21] K. Johansen *et al.*, "Predicting Biomass and Yield in a Tomato Phenotyping Experiment Using UAV Imagery and Random Forest," *Frontiers in Artificial Intelligence*, vol. 3, May 2020, Art. no. 28, https://doi.org/10.3389/frai.2020.00028.

[22] K. Grunberg, R. Fernández-Muñoz, and J. Cuartero, "Growth, Flowering, and Quality and Quantity of Pollen of Tomato Plants Grown Under Saline Conditions," *Acta Horticulturae*, no. 412, pp. 484–489, Nov. 1995, https://doi.org/10.17660/ActaHortic.1995.412.58.

[23] H. M. Zayani *et al.*, "Deep Learning for Tomato Disease Detection with YOLOv8," *Engineering, Technology & Applied Science Research*, vol. 14, no. 2, pp. 13584–13591, Apr. 2024, https://doi.org/10.48084/etasr.7064.