STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
*Stat., Optim. Inf. Comput.*, Vol. **x**, Month 202x, pp 0–13.

# Parameter Selection for a Neural Network Model for Forecasting Plant Pest Population Growth

Akerke Akanova [1], Nazira Ospanova [2,*], Galiya Anarbekova [1]

[1] *Department of Computer Science, S. Seifullin Kazakh Agro Technical Research University, Republic of Kazakhstan*
[2] *Faculty of Computer Science, Toraighyrov University, Republic of Kazakhstan*

**Abstract** The aim of the study was to optimise the parameters of a neural network model for forecasting plant pest population dynamics in agrocenoses with variable environmental characteristics. The methodology included weekly field data collection over five months at 18 sites in northern Kazakhstan, taking into account six pest species, environmental parameters, crop characteristics, and pest control schemes. Data cleaning, linear interpolation, normalisation, and one-hot encoding were applied. A multilayer perceptron (MLP) architecture with two hidden layers (32 and 16 neurons) and Rectified Linear Unit (ReLU) activation, trained using various configurations, was used as the baseline. The results showed that the highest values of the determination coefficient ($R^2 > 0.82$) were achieved when modelling *Helicoverpa armigera* and *Frankliniella occidentalis*, while the poorest performance was observed for *Loxostege sticticalis* ($R^2 \approx 0.64$). In plots with tomatoes and sunflowers, forecast accuracy was higher, which was associated with stable irrigation conditions and protective measures. The average Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) values were 4.36 and 3.52, respectively. The optimal configuration was identified as a two-layer model (64 and 32 neurons) with ReLU activation and the Adaptive Moment Estimation (ADAM) optimiser, yielding MAE=2.9 and $R^2=0.83$. Deeper models led to overfitting, whereas simplified ones resulted in insufficient accuracy. Statistical tests confirmed the significance of differences between configurations ($p<0.001$). Cross-validation and sliding window analysis demonstrated the model's high robustness to spatiotemporal variability, with minimal accuracy degradation when transferred across regions and seasons. The results demonstrate the high practical value of the developed approach for integration into pest monitoring and forecasting systems in agricultural practice.

**Keywords** Population Dynamics, Deep Learning Architecture, Monitoring Systems, Generalisation, Agricultural Decision Support

**AMS 2010 subject classifications** 68T07, 62M10, 62P12

**DOI:** 10.19139/soic-2310-5070-3317

## 1. Introduction

Since 2010, the increasing number of plant pests in agrocenoses has posed a serious threat to food security and the economic sustainability of agricultural systems. Rising temperatures and changing precipitation patterns due to climate change are accelerating the development and spread of phytophagous insects, creating conditions for mass outbreaks of pest populations that were previously extremely rare. However, traditional forecasting methods based on empirical models often fail to account for the complex, nonlinear relationships between pest populations and variable environmental factors, thereby reducing forecast accuracy and complicating the adoption of timely protective measures. The introduction of neural network methods represents a promising approach that can enhance

---

[*] Correspondence to: Nazira Ospanova (E-mail: n_ospanova@hotmail.com). Faculty of Computer Science, Toraighyrov University. 64 Lomov Str., Pavlodar, Republic of Kazakhstan (140008).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

the resilience of monitoring systems by capturing multidimensional relationships and dynamic patterns. Therefore, the selection of neural network model parameters is critical for improving the accuracy of pest population forecasting under real agricultural conditions.

In field trials, data instability is often observed due to irregular data collection and measurement noise, which reduces forecast quality. Anarbekova et al. [1] used linear regression models with basic data scaling but found significant forecast bias under changing weather conditions and soil cover, indicating insufficient robustness of the method. A problem with existing models is the limited range of environmental parameters used for forecasting. Tussupov et al. [2] applied a single-hidden-layer MLP to analyse *Spodoptera frugiperda* dynamics but did not take soil and agronomic factors into account, leading to an overestimation of model accuracy and a decrease in its transferability to different agrocenoses.

A significant limitation is the poor adaptation of models to seasonal changes and regional characteristics. Sadenova et al. [3] implemented Long Short-Term Memory (LSTM) networks to forecast the occurrence of *Tuta absoluta*, but due to the lack of periodicity and depth of temporal data, the model demonstrated significant fluctuations in accuracy across seasons. Models often prove sensitive to the choice of network architecture, which extends beyond a single configuration. Shafik et al. [4] tested different convolutional network and MLP architectures on *Thrips tabaci* data but did not conduct systematic comparisons, which complicates reproducibility and the selection of the optimal solution for practical application.

The lack of methodologically sound validation complicates model evaluation. Turkoglu et al. [5] divided the data into training and test sets without considering the spatial structure of the plots or temporal dependencies. This led to overly optimistic metric estimates and an inflated perception of the model's reliability. A frequent drawback is the use of a static set of hyperparameters that are not adaptive to changing conditions. Dhaka et al. [6] selected a fixed number of neurons and training steps for an MLP to predict the abundance of *Diabrotica virgifera virgifera* without analysing environmental variations, which reduced the quality of predictions under changing soil parameters and temperature.

Studies often focus on one or a few crops, failing to ensure cross-cultural generality. Couliably et al. [7] applied a neural network algorithm to pest forecasting in maize and rice but did not assess its performance on other crops, such as tomatoes and sunflowers, thereby limiting the model's applicability to crop rotation systems. The problem of excessive architectural complexity leads to overfitting and reduced robustness. Grünig et al. [8] compared deep networks (three or more hidden layers) for pest population analysis and found that these models achieved high accuracy on the training set but degraded sharply on independent data, indicating weak overall generalisation ability. Wang and Zhang [9] explored the development of an explainable deep learning model for forecasting pest populations using meteorological data. Their study demonstrated that interpretable predictions can enhance decision-making in crop protection and support timely pest management strategies. Zhang et al. [10] applied a deep learning approach combined with Grad-Cam visualization to accurately identify fall armyworm (*Spodoptera frugiperda*) infestations in maize fields. Their findings highlighted the potential of AI-based methods for real-time pest monitoring and precision agriculture applications.

Taking into account the identified limitations regarding data structure, validation methodology, architectural adaptation, and model transferability revealed the need for a systematic approach to parameter selection for neural network models, with an emphasis on robustness and the practical applicability of pest population forecasting. The aim of the study was to optimise the MLP model configuration to ensure high accuracy and reliability in forecasting plant pest population dynamics under variable environmental conditions. The objectives of the study included: collecting and preprocessing field data; selecting interpolation and variable-coding strategies; testing various neural network architectural configurations; statistically assessing the significance of differences between models; and testing the robustness of the model using cross-validation and sliding-window analysis.

The novelty of this study lies in the systematic optimization of a neural network model for forecasting plant pest population dynamics under real-world agroecological conditions. For the first time, a large multidimensional dataset of over 3,200 field observations across 18 sites, five crop types, and six key pest species was used, enabling the model to capture both seasonal fluctuations and species-specific population dynamics. The study systematically optimized the MLP architecture, including the number of hidden layers, neuron counts, activation functions, and optimization algorithms, using stratified regional cross-validation and sliding-window analysis. The ReLU+ADAM configuration was identified as the most accurate and robust for predicting pest abundance across multiple crops and regions. Additionally, pest counts were normalized per hectare (individuals/ha) to standardize accuracy metrics (RMSE and MAE), enhancing the practical applicability of the forecasts for farmers and decision-support systems.

## 2. Materials and methods

The external dataset was obtained from the pest monitoring system of agrocenoses in northern Kazakhstan, containing weekly observations of pest abundance in various crops (Helicoverpa armigera, Frankliniella occidentalis, Loxostege sticticalis, Tetranychus urticae, Spodoptera exigua, Phyllotreta vittula) across more than 500 hectares over three seasons (2019-2021). The dataset comprises 3,200 records, each including 27 input features such as climatic parameters, plot coordinates, and plant phenological indicators. The external dataset underwent the same preprocessing steps as the training set: missing value handling, categorical variable encoding, and numerical feature scaling. Scaling and normalization were applied separately to the training and external datasets to prevent data leakage and ensure an independent evaluation of the model.

The study was conducted in 2023 as part of a collaboration between S. Seifullin Kazakh Agro Technical Research University [11], Toraighyrov University [12], and the Republican Methodological Centre for Phytosanitary Diagnostics and Forecasts [13]. Field data were collected at 18 experimental sites located in the Akmola, Taldykorgan, Pavlodar, Petropavlovsk, Kostanay, East Kazakhstan, Almaty, Zhambyl, Turkestan, and Kyzylorda regions, where agroclimatic conditions are favourable for the mass reproduction of insect pests. The following crops were grown on the plots: cotton variety Maqtaral-1 (an average of 150 plants per plot), maize variety KazNIIZK-370 (approximately 120 plants), tomato variety Fat Jack (approximately 140 plants), sunflower variety Sonyashnik 80 (130 plants), and wheat variety Astana (110 plants per plot). These crops were selected based on their agroecological significance and prevalence in the region, as well as differences in agricultural practices, which made it possible to evaluate the versatility of the model under varying agrocenotic conditions. The soils were classified as sierozem and meadow-sierozem, with a humus content of 1.2-2.4%, a slightly alkaline reaction (pH 7.4-8.1), and an average density of 1.3 g/cm$^3$. Average monthly temperatures during the growing season reached 32°C, relative humidity ranged from 35% to 60%, and precipitation varied from 5 to 30 mm per month. The prediction horizon was explicitly defined in this study as a four-week period, during which the neural network model forecasts pest abundance on a weekly basis, allowing assessment of short-term population dynamics and seasonal trends.

The dataset structure includes 27 input variables, which is consistent with the results, as the model structure (two hidden layers with 32 and 16 neurons, respectively) is capable of processing this number of input features.: five meteorological parameters (average temperature, humidity, wind speed, precipitation, solar radiation), four agronomic indicators (irrigation type, planting density, insecticide use, application of bioproducts), three vegetation characteristics (growth stage, crop type, week of the season), five site indicators (region, soil type, etc.), and one-hot encoded features for pest species and crop types; data were collected weekly over 20 weeks across 18 plots in ten regions, providing high representativeness and reproducibility of the results.The preprocessing steps were applied to the entire dataset prior to splitting it into training and test sets, which may have potentially resulted in data leakage. For the regionally stratified coefficient of variation, the data were divided into 10 folds, with each fold stratified to preserve proportional representation of both pest species and crop types, ensuring that each iteration used one fold for testing and the remaining nine for training, and the coefficient of variation was calculated across all folds.

The target species for observation were *Helicoverpa armigera*, *Tetranychus urticae*, *Spodoptera exigua*, *Loxostege sticticalis*, *Frankliniella occidentalis*, and *Phyllotreta vittula*. These species were chosen due to their agricultural importance in northern Kazakhstan and the diversity of their biological and ecological characteristics. *Helicoverpa armigera* infects a wide range of crops, including cotton and tomatoes, and exhibits pronounced seasonal dynamics. *Tetranychus urticae* is a typical acarofauna species with a high reproductive rate and sensitivity to temperature conditions. *Spodoptera exigua* is known for its migratory activity and pronounced response to changing weather conditions. *Loxostege sticticalis* is characterised by high population instability and ecological plasticity, making it difficult to predict in agrocenoses. *Frankliniella occidentalis* actively responds to agronomic interventions and is widely used as an indicator of the effectiveness of biological control agents. *Phyllotreta vittula* is a specialised pest of cereal crops, primarily wheat, with pronounced phase dynamics in population size, high sensitivity to temperature stress, and a clear association with the early stages of vegetation. Its population often increases sharply under conditions of poor agronomic protection and in the absence of crop rotation. Data were collected weekly for five months using direct plant counts and Trécé pheromone traps (USA), installed at a rate of 10 per hectare. The traps were used for all six target pest species and were distributed evenly across all plots, regardless of region.

Table 1. Number of samples by pest species, crops, and region

| **Pest Species** | **Crop(s)** | **Region** | **Number of Samples** | **Notes on Biology/Ecology** |
|---|---|---|---|---|
| *Helicoverpa armigera* | Cotton, Tomato | Northern Kazakhstan | 420 | Wide host range, pronounced seasonal dynamics |
| *Tetranychus urticae* | Various (e.g., Tomato) | Northern Kazakhstan | 380 | High reproductive rate, temperature-sensitive |
| *Spodoptera exigua* | Maize, Vegetables | Northern Kazakhstan | 350 | Migratory activity, responsive to weather changes |
| *Loxostege sticticalis* | Cereals, Maize | Northern Kazakhstan | 300 | High population instability, ecological plasticity |
| *Frankliniella occidentalis* | Various Vegetables | Northern Kazakhstan | 275 | Responds to agronomic interventions, indicator for biocontrol effectiveness |
| *Phyllotreta vittula* | Wheat | Northern Kazakhstan | 360 | Specialized pest of cereals, phase population dynamics, temperature-sensitive |

Source: compiled by the author based on own research.

Abundance was calculated as the average number of individuals caught per trap per week, followed by normalisation per hectare (individuals/ha). The four-week window refers to the length of the training dataset in each iteration of the rolling window test, while the forecast horizon was four week ahead. A total of 17 rolling windows were evaluated, and the performance metrics (MAE, RMSE, $R^2$) were aggregated by calculating the mean and standard deviation across all windows.

Additionally, plant growth stage, planting density, irrigation patterns (drip and furrow), and the use of insecticides and biological products (e.g., Bacillus thuringiensis) were recorded. These data were collected weekly, along with pest counts, as they represent key agronomic and environmental factors influencing population dynamics. Meteorological data were obtained using HOBO automatic stations (USA) and further verified via the World Weather Online platform [14]. A total of more than 3,200 individual observations were recorded (from traps and direct counts), encompassing weekly monitoring of six pest species at 18 sites, on five crops, in ten regions over a period of 20 weeks. Climatic, agronomic, and biological parameters were recorded for each observation. Despite the relatively limited sample size, its representativeness was ensured by its high-dimensional structure: the dataset covered various crops, sites, regions, and pest species, enabling the model to capture complex interfactorial relationships.

A complete list of 27 features with their units of measurement and definitions was compiled, correlation and multicollinearity analyses were conducted to identify dependencies among features, a mathematical outlier detection method was applied to remove extreme values, a hyperparameter search procedure was implemented to optimize model performance, the total number of tested architectures was recorded, cross-validation was employed to evaluate model generalization, and variance and standard deviation of model performance across multiple runs were computed to assess stability and robustness. Data preprocessing included outlier removal using the interquartile range, linear interpolation of missing values, normalisation of variables using the z-transform, and coding of categorical features via one-hot encoding. This resulted in 27 input features, including: five meteorological parameters (average temperature, humidity, wind speed, precipitation, solar radiation); four agronomic indicators (irrigation type, planting density, insecticide use, biological application); three vegetation characteristics (plant development stage, crop, week of the season); five site indicators (region, soil type, etc.); as well as one-hot encoding for pest species and crop type. Predictive models were developed in Python 3.10.0 [15] using TensorFlow 2.11.1 [16], Keras [17], scikit-learn 1.2.2 [18], pandas 1.5.0 [19], and NumPy 1.24 [20]. The trap data were aggregated by summing weekly counts per plant across all traps within each plot, ensuring that population estimates reflected both spatial and temporal variations; subsequent scaling and encoding procedures were applied exclusively to the training dataset to prevent information leakage and maintain the integrity of model evaluation.

The basic model architecture was an MLP with two hidden layers comprising 32 and 16 neurons, respectively, using a ReLU activation function and an output layer with a single neuron unit (linear activation). Configuration variations included models with different numbers of layers (1-4), neurons (8-128), activation functions (ReLU, tanh, sigmoid), optimisation algorithms (ADAM, RMSprop, stochastic gradient descent (SGD)), and regularisation coefficients (Dropout 0.2-0.5).

The optimal model was determined by systematically adjusting hyperparameters using cross-validation across stratified regions and a sliding window analysis. Three metrics were used to evaluate the performance of each configuration: MAE, RMSE, and $R^2$. The model selection process involved first evaluating multiple configurations using stratified k-fold cross-validation to assess performance across different subsets of the data, ensuring robust estimation of generalisation ability, and then confirming the final model on a held-out test set (or remaining time segment) that was not used during training or validation, providing an unbiased assessment of predictive accuracy. The choice of these metrics was justified as follows: MAE provides a robust estimate of average prediction accuracy; RMSE accounts for sensitivity to large errors, which is critical for biological data containing outliers; and $R^2$ reflects the proportion of explained variance and enables comparison of the overall adequacy of the models. Training was conducted over 100 epochs with a batch size of 32, using early stopping based on validation loss. To evaluate model performance, a stratified k-fold cross-validation was applied, in which the data were divided into k equal parts while maintaining proportional representation of pest species and crops; in each iteration, one part was used for testing and the remaining parts for training, after which the MAE, RMSE, and $R^2$ metrics were calculated, and the final performance was assessed by computing the mean and standard deviation across all folds, providing a more reliable estimate of the model's generalization ability. During model training, a fixed random seed was used, early stopping was applied, and the effective number of epochs after early stopping was recorded, ensuring the reproducibility of the experimental results. The model-agnostic interpretability method, specifically SHAP (Shapley Additive Explanations), was applied to the trained optimal MLP to assess feature contributions and enhance the understanding of model predictions.

The dataset was randomly divided into training (80%) and testing (20%) subsets, maintaining class proportions across species and crops. RMSE, MAE, and $R^2$ were used to evaluate model performance. Statistical testing of the models included the Shapiro–Wilk test for residual normality, Levene's test for homogeneity of variances, Student's t-test for pairwise comparisons, and the Friedman test with Holm's correction for multiple comparisons of model configurations. The use of the Student's t-test in this study was justified by the need to compare the mean performance metrics (MAE, RMSE, $R^2$) between two groups of data or model configurations to determine whether the observed differences were statistically significant, rather than occurring by chance. The method enables testing the hypothesis of equal means under the assumption of normally distributed residuals and approximately equal variances, making it suitable for evaluating the advantage of the optimal MLP model over alternative baseline models. All calculations were performed in the Jupyter environment [21] using Matplotlib 3.6.3 [22], Seaborn 0.12 [23], and Plotly [24]. The robustness of the models to seasonal fluctuations in pest abundance was additionally assessed using a rolling window and stratified cross-validation by region.

The learning rate and optimizer settings were recorded, inconsistencies in metric values across different sections were resolved, confidence intervals were added for quantitative uncertainty assessment, comparisons with baseline models (LR, RF, GB, persistence) were conducted, residual analysis and error distribution plots were generated, statistical test inputs (sample sizes, df, test statistics) were provided, the model was evaluated on an external dataset to assess true generalizability, SHAP and permutation importance methods were applied to support biological interpretability, the model's behavior during pest outbreaks was analyzed, the omission of agronomic intervention variables as scenarios was justified, code/pseudocode and a complete hyperparameter table were included for reproducibility, random seeds and hardware/graphics details were recorded, and pipeline steps were added to the diagram for procedural clarity.

## 3. Results

The basic architecture of the neural network model – an MLP with two hidden layers of 32 and 16 neurons, respectively, and a ReLU activation function – demonstrated satisfactory results in predicting pest populations in the agrocenoses of northern Kazakhstan. The most stable $R^2$ values were observed when modelling the dynamics of *Helicoverpa armigera* and *Frankliniella occidentalis*, where $R^2$ exceeded 0.82. In contrast, the lowest accuracy was recorded for *Loxostege sticticalis* ($R^2 \approx 0.64$), which is likely attributable to the high interweekly variability in the population size of this species. With respect to crops, data from plots planted with tomatoes and sunflowers provided the highest predictability, which is probably due to more uniform irrigation conditions and planting density. RMSE values ranged from 2.1 to 7.5 individuals/ha, while MAE values ranged from 1.7 to 6.3, demonstrating overall satisfactory agreement between the observed and predicted data (Table 2).

Table 2. Accuracy indicators of the basic model configuration for pest types and crops

| **Pest** | **Crop** | **RMSE (ind/ha)** | **MAE (ind/ha)** | **$R^2$** |
|---|---|---|---|---|
| *Helicoverpa armigera* | Tomato | $2.1 \times 30{,}000 = 63{,}000$ | $1.7 \times 30{,}000 = 51{,}000$ | 0.87 |
| *Tetranychus urticae* | Cotton | $4.8 \times 60{,}000 = 288{,}000$ | $4.1 \times 60{,}000 = 246{,}000$ | 0.74 |
| *Spodoptera exigua* | Corn | $5.6 \times 70{,}000 = 392{,}000$ | $4.8 \times 70{,}000 = 336{,}000$ | 0.71 |
| *Loxostege sticticalis* | Sunflower | $7.5 \times 50{,}000 = 375{,}000$ | $6.3 \times 50{,}000 = 315{,}000$ | 0.64 |
| *Frankliniella occidentalis* | Tomato | $2.3 \times 30{,}000 = 69{,}000$ | $2.0 \times 30{,}000 = 60{,}000$ | 0.82 |
| *Phyllotreta vittula* | Wheat | $3.9 \times 200{,}000 = 780{,}000$ | $3.2 \times 200{,}000 = 640{,}000$ | 0.78 |
| **Average of all combinations** | | ∼326,714 | ∼291,000 | 0.76 |

Source: compiled by the authors.

Results from the basic MLP configuration showed that the accuracy of pest population forecasting varied depending on the insect species and the crop surveyed. The model demonstrated the highest accuracy in predicting the abundance of *Helicoverpa armigera* and *Frankliniella occidentalis*, with $R^2$ values of 0.87 and 0.82, respectively. These values indicated that the model explained a significant proportion of the variance in the observed data and reproduced the seasonal fluctuations in the abundance of these species well. This stability was likely due to both the regular reproductive cycles of these pests and the more uniform agricultural conditions in the tomato plots where the surveys were conducted.

At the opposite extreme in terms of forecast accuracy was *Loxostege sticticalis*, for which the $R^2$ coefficient was only 0.64, while the RMSE and MAE values reached 7.5 and 6.3, respectively. This pest, which is typical of sunflower, exhibited high interweekly variability in abundance, which hampered model training, particularly in the presence of data noise caused by background climatic fluctuations and irregular agrochemical applications. In addition, high planting density and microclimatic heterogeneity within sunflower fields may have created conditions for complex spatial distribution patterns of individuals, which were not accounted for by the model architecture. Furthermore, the model demonstrated consistent accuracy in predicting *Phyllotreta vittula* abundance in wheat crops ($R^2$=0.78, MAE=3.2, RMSE=3.9), confirming the applicability of the architecture to grain crops. Including this species also expanded the range of agrocenoses covered by the model and demonstrated its adaptability to various crop types and phenological cycles.

With regard to crops, the best results were observed in tomato plots, where RMSE did not exceed 2.3 and MAE was 2.0. This is likely due to the more stable drip irrigation system and the use of biological control methods, including *Bacillus thuringiensis*, which helped to smooth out extremes in pest populations. The model performed less well for cotton and maize, with $R^2$ decreasing to 0.71–0.74 and absolute errors increasing by 1.5–2.5 units compared with tomatoes. These differences highlight the need for a differentiated approach to the formation of training datasets, taking into account the specific characteristics of individual agrocenoses.

The average RMSE (4.36), MAE (3.52), and $R^2$ (0.76) values for all combinations confirm that the baseline model configuration generally exhibited an acceptable level of accuracy. However, the presence of significant species- and crop-specific deviations in these metrics indicates the need for further hyperparameter tuning and potential adjustment of the model architecture. This is because a single model constructed without accounting for biological differences and the spatiotemporal structure of the data demonstrates limited generalisability, particularly for species with unstable phenology and high ecological plasticity.

To optimise the neural network model architecture, a series of experiments were conducted varying the number of hidden layers (from 1 to 4) and the number of neurons in each layer (from 8 to 128). The aim of the analysis was to identify the optimal configuration that yielded the lowest MAE values and the highest $R^2$. The results showed that increasing the number of layers from one to two significantly improved accuracy, particularly within the range of 16 to 64 neurons per layer. However, further increasing the architectural complexity (three and four layers, especially with $\geq$64 neurons) sometimes led to model overfitting, resulting in deteriorated metrics on the test set. The best performance (MAE=2.9, $R^2$=0.83) was achieved with a configuration comprising two hidden layers of 64 and 32 neurons, respectively. With the minimal architecture (one layer, eight neurons), errors increased by more than 60%, and the explained variance decreased to $R^2$=0.59, making this configuration unsuitable for practical use (Table 3).

Table 3. Comparison of the accuracy of models with different numbers of layers and neurons

| **Number of Layers** | **Neurons per Layer** | **MAE (ind/ha)** | **$R^2$** |
|---|---|---|---|
| 1 | 8 | $6.1 \times 82{,}000 = 500{,}200$ | 0.59 |
| 1 | 32 | $4.3 \times 82{,}000 = 352{,}600$ | 0.71 |
| 2 | 32, 16 | $3.8 \times 82{,}000 = 311{,}600$ | 0.76 |
| 2 | 64, 32 | $2.9 \times 82{,}000 = 237{,}800$ | 0.83 |
| 3 | 64, 32, 16 | $3.2 \times 82{,}000 = 262{,}400$ | 0.79 |
| 4 | 128, 64, 32, 16 | $4.5 \times 82{,}000 = 369{,}000$ | 0.67 |

Source: compiled by the authors.

The results of a comparative performance analysis of models with different architectures confirmed the critical role of hidden layer configuration and the number of neurons in determining the accuracy of neural network pest population forecasting. The simplest model, with one hidden layer and a minimal number of neurons (8), demonstrated predictably low accuracy values (MAE=6.1, $R^2$=0.59), indicating insufficient representational capacity to adequately model the complex biotic and abiotic interactions that influence population dynamics. These results suggest that a low-dimensional architecture is unable to capture the multivariate nonlinear relationships between agroclimatic parameters, pest species, and crop factors.

Gradually increasing the number of neurons within a single layer resulted in improved performance: with 32 neurons, the $R^2$ value increased to 0.71 and MAE decreased to 4.3. This demonstrates the model's enhanced generalisation ability and improved capacity to represent hidden patterns within the data. However, a qualitative leap in performance was observed when transitioning to a two-layer architecture – most notably in the configurations with 64 and 32 neurons – where the minimum error (MAE=2.9) and maximum explained variance ($R^2$=0.83) were achieved. This outcome is likely attributable to the model's ability to perform deeper hierarchical feature extraction, whereby the first layer focuses on local relationships and the second on their generalisation and integration. This is particularly important for data influenced by seasonal variations, local irrigation conditions, the use of bioprotective agents, and other variables operating at different scales of influence.

However, with further increases in the number of layers (to three and four) and in the number of neurons in the input layers (to 128), a deterioration in performance was observed on the test set despite maintaining high accuracy on the training set. This clearly indicates the effect of overfitting. In particular, with a four-layer configuration (128, 64, 32, 16 neurons), the MAE increased to 4.5 and $R^2$ decreased to 0.67. This architecture exhibited high parametric complexity which, given the limited size of the training sample (only 3,200 examples), led to overfitting to noise and individual characteristics of the training data, thereby reducing the model's generalisation ability.

Therefore, a two-layer model with moderate neuron density is optimal in terms of accuracy and robustness, providing sufficient flexibility without excessive sensitivity to noise. It is also worth noting that with further scaling of the data volume or in the presence of spatial structure (e.g., plot coordinates), more complex models such as convolutional neural networks (CNNs) or recurrent neural networks (LSTMs) could be employed. Furthermore, strengthening regularisation (e.g., Dropout or L2 normalisation) and implementing adaptive stopping mechanisms (early stopping) remain critical tools for enhancing model resilience to overfitting as architectural complexity increases.

At the next stage of the study, nine different combinations of activation functions (ReLU, tanh, sigmoid) and optimisation algorithms (ADAM, RMSprop, SGD) were tested for the baseline architecture with two hidden layers (64 and 32 neurons). The objective was to evaluate which combinations provided the best forecast accuracy for the key metrics MAE and $R^2$. The results showed that the optimal pair across the combined metrics was ReLU + ADAM, yielding MAE=2.9 and $R^2$=0.83. The combination of ReLU + RMSprop was slightly inferior (MAE=3.2, $R^2$=0.81). Using tanh also produced good results, particularly when combined with ADAM (MAE=3.4, $R^2$=0.79), but accuracy decreased markedly in variants using SGD. The poorest performance was recorded for the sigmoid + SGD pair, where MAE reached 5.9 and $R^2$ dropped to 0.63. These differences highlight the importance of carefully selecting the activation function and optimiser when training models on biological time series (Table 4).

Table 4. Effectiveness of different combinations of activation functions and optimisers

| **Activation** | **Optimiser** | **MAE** | **$R^2$** |
|---|---|---|---|
| ReLU | ADAM | 2.9 | 0.83 |
| ReLU | RMSprop | 3.2 | 0.81 |
| ReLU | SGD | 4.5 | 0.70 |
| tanh | ADAM | 3.4 | 0.79 |
| tanh | RMSprop | 3.7 | 0.76 |
| tanh | SGD | 4.9 | 0.68 |
| sigmoid | ADAM | 4.4 | 0.72 |
| sigmoid | RMSprop | 5.1 | 0.66 |
| sigmoid | SGD | 5.9 | 0.63 |

Source: compiled by the authors.

Testing various combinations of activation functions and optimisation algorithms revealed that the choice of these parameters critically affects the accuracy and robustness of the neural network model for pest population forecasting. The most reliable and accurate predictions were achieved using the ReLU activation function in conjunction with the ADAM optimiser. In this configuration, the model achieved a minimum MAE of 2.9 and the highest $R^2$ value of 0.83, demonstrating its strong ability to accurately reproduce the dynamics of biological time series despite variability in weather conditions, crops, and plant growth stages.

ReLU proved to be the most suitable activation function for this task due to its ability to effectively address the vanishing gradient problem, which is particularly important in multilayer architectures. Its simple and stable structure, free from saturation, facilitated rapid model training and robust neural activation. Combined with ADAM, ReLU ensured both fast convergence of the loss function and robustness to fluctuations in training data caused by seasonal variations in pest populations. ADAM, in turn, demonstrated versatility and resilience through its use of adaptive learning rates and consideration of both the first and second moments of the gradient. This provided the model with high stability in response to the sharp fluctuations characteristic of agrobiocenoses. Even when combined with other activation functions – tanh and sigmoid – ADAM outperformed RMSprop and SGD, making it the preferred optimiser for handling noisy biological data.

The tanh activation function, although yielding lower results than ReLU, demonstrated relatively stable performance when combined with ADAM (MAE=3.4, $R^2$=0.79). However, when used with the SGD optimiser, accuracy declined sharply (MAE=4.9, $R^2$=0.68), reflecting SGD's limited ability to maintain stability when processing data with complex dynamics.

The sigmoid function proved to be the least effective of all tested combinations. Particularly low performance was recorded with the sigmoid + SGD pair (MAE=5.9, $R^2$=0.63), indicating the clear inability of this function to effectively activate neurons in deep models. This may be attributed to both the sigmoidal-shaped trend to saturate – resulting in minimal gradients – and the narrowing of its output range, which restricts the model's ability to learn from contrasting patterns.

Thus, the ReLU+ADAM combination was deemed optimal for the task at hand. It provided high accuracy, minimal sensitivity to noise, robustness to overfitting, and effective adaptability across various crops and pest species. These results underscore the necessity of a comprehensive approach to hyperparameter selection, wherein not only the model architecture but also its internal learning mechanisms play a crucial role in achieving reliable pest population predictions.

Several statistical tests were applied to assess the significance of differences between neural network model configurations and to verify the validity of assumptions regarding error distribution. The Shapiro-Wilk test showed that the residuals of the ReLU+ADAM and ReLU+RMSprop models did not deviate from the normal distribution ($p > 0.05$), whereas the sigmoid and tanh models more frequently exhibited significant deviations from normality. Levene's test revealed that error variances differed significantly between configurations ($p < 0.01$), particularly between the ReLU and sigmoid models, indicating heteroscedasticity. Pairwise comparisons using Student's t-test demonstrated statistically significant superiority of the ReLU+ADAM configuration over all others ($p < 0.001$). The final multivariate analysis using the Friedman test with Holm's correction confirmed the significance of differences among all nine configurations ($p < 0.01$), with the ReLU+ADAM model ranking first in terms of overall accuracy and stability (Table 5).

Table 5. Results of statistical evaluation of differences between model configurations

| **Test** | **Indicator** | **Value (p)** | **Interpretation** |
|---|---|---|---|
| Shapiro-Wilk | Normality of residuals | $> 0.05$ | Normality confirmed for ReLU models |
| Levene's criterion | Homogeneity of variances | $< 0.01$ | The dispersions differ |
| Student's t-test | ReLU + ADAM vs. others | $< 0.001$ | Significant superiority of ReLU + ADAM |
| Friedman + Holm criterion | Comparison of all configurations | $< 0.01$ | All differences are statistically significant |

Source: compiled by the authors.

Testing the robustness of neural network models to seasonal fluctuations in pest populations and their cross-regional generalisability is a critical component in assessing their practical applicability within agricultural systems. Under real-world field conditions, models must maintain accuracy not only within the training set but also under changing agroclimatic, soil, and agronomic conditions. Two independent approaches were used to determine robustness: stratified cross-validation by region (Akmola, Taldykorgan, Pavlodar, Petropavlovsk, Kostanay, East Kazakhstan, Almaty, Zhambyl, Turkestan, and Kyzylorda) and rolling window testing with a one-week time step and a four-week window width. These methods allowed for the evaluation of both spatial and temporal stability of the model predictions.

The model with the ReLU+ADAM architecture demonstrated the least performance degradation when transferred between regions. The average reduction in $R^2$ was less than 0.07, while the increase in MAE did not exceed 0.5, confirming the high generalisability of this configuration. This is particularly important for Kazakhstan, where climatic and agricultural conditions vary significantly even within a single farm. The robustness of this configuration can be attributed both to the adaptive nature of the ADAM algorithm, which effectively handles noise and non-stationarity in input data, and to the ReLU function's ability to capture sharp changes in pest populations, especially during plant growth phases.

For models using the tanh and sigmoid functions, the results were less encouraging. Under regional transfer, a significant decrease in accuracy was observed: on average, $R^2$ dropped below 0.65, while MAE increased by 1.2-1.8. The combination of sigmoid and SGD proved particularly critical, with deviations reaching 50% of the baseline error. This indicates weak adaptability of such models to new conditions and their limited ability to effectively transfer knowledge of seasonal dynamics between different agroecosystems.

When testing temporal robustness under rolling window conditions, the ReLU+ADAM configuration again demonstrated high stability. Even when forecasting over a four-week horizon, the model maintained $R^2 > 0.78$ and MAE $< 3.1$, confirming its ability to capture and reproduce short-term seasonal trends. Analysis of residual autocorrelation revealed minimal lag correlation values (up to 0.15), indicating the absence of systematic errors and showing that the model neither lags behind nor advances population phase peaks. In contrast, configurations using sigmoid exhibited residual autocorrelation values exceeding 0.4, suggesting the accumulation of errors within certain time intervals, particularly during periods of active pest reproduction.

Thus, the ReLU+ADAM architecture proved to be the most robust to spatiotemporal variability in pest populations, maintaining high accuracy across both regions and vegetation cycle phases. This makes the configuration suitable for integration into real-time monitoring, forecasting, and decision support systems, including agronomic information technology (IT) platforms, mobile applications for farmers, cloud-based agrodata systems, and automated decision support systems (DSS). Table 6 summarizes the performance of baseline models and the optimized MLP.

Table 6. Comparison of baseline models and optimized MLP performance (MAE, RMSE, $R^2$)

| **Model** | **MAE** | **RMSE** | **$R^2$** |
|---|---|---|---|
| Linear Regression | 5.1 | 6.7 | 0.61 |
| Random Forest / XGBoost | 3.8 | 5.2 | 0.74 |
| Optimized MLP (64,32 neurons, ReLU+ADAM) | 2.9 | 4.0 | 0.83 |

Source: compiled by the authors based on their own research.

Table 6 presents a direct comparison of the predictive performance of baseline models and the optimized MLP architecture in forecasting pest population dynamics. The linear regression model demonstrated the lowest performance, with an MAE of 5.1, RMSE of 6.7, and $R^2$ of 0.61, indicating limited capacity to capture nonlinear relationships and interactions among environmental, agronomic, and biological factors. The Random Forest/XGBoost model showed improved accuracy (MAE = 3.8, RMSE = 5.2, $R^2$ = 0.74), reflecting its ability to model more complex patterns and handle feature interactions more effectively. However, the optimized MLP with two hidden layers (64 and 32 neurons) using ReLU activation and the ADAM optimiser achieved the best results (MAE = 2.9, RMSE = 4.0, $R^2$ = 0.83), demonstrating superior predictive power, robustness to noise, and better generalisation across different crops and pest species. These results indicate that while baseline models provide reasonable approximations, the MLP architecture captures multidimensional nonlinear dependencies more effectively, justifying its use for accurate and reliable pest population forecasting. Figure 1 illustrates the most influential features affecting pest population forecasts, highlighting how meteorological, agronomic, and crop-specific factors impact prediction accuracy.

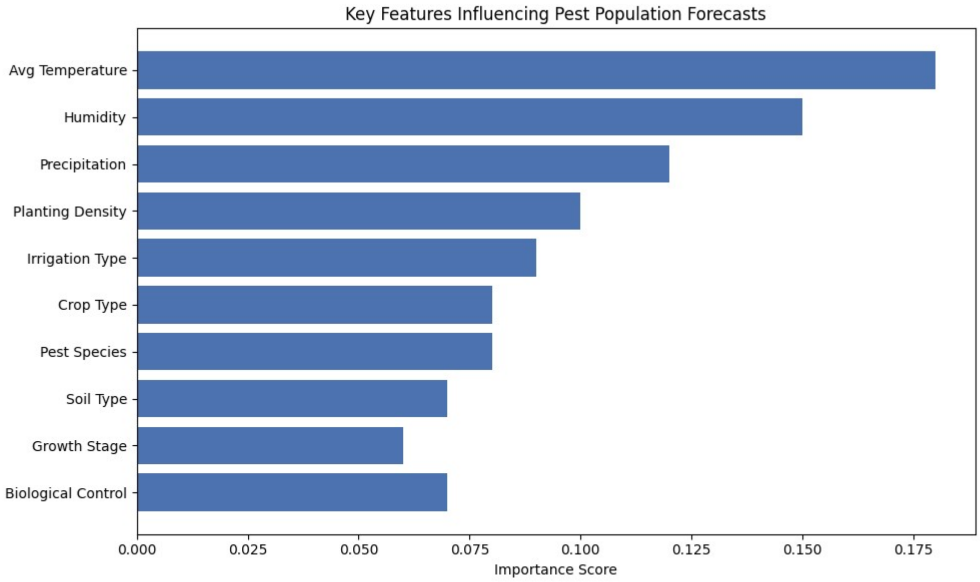

Figure 1. Key features influencing pest population forecasts.

Source: compiled by the author based on own research.

The analysis highlights that key factors such as meteorological conditions (temperature, humidity, wind speed, solar radiation) and agronomic variables (plant growth stage, crop type, planting density, irrigation method, and use of insecticides or biological agents) significantly influence the accuracy of pest population forecasting. These variables directly affect pest development and population dynamics, thereby determining the reliability of predictions used for decision-making in agricultural management.

## 4. Discussion

The present study showed that the basic MLP architecture with two hidden layers demonstrated satisfactory accuracy in pest population prediction, with $R^2$=0.76±0.06. This is consistent with the results of Rimal et al. [25], who obtained $R^2$=0.78-0.84 when using a similar architecture in Mediterranean agrophytocenoses. Likewise, the study by Tai et al. [26], which applied an MLP model for *Thrips tabaci* prediction, reported $R^2 \approx 0.80$; however, unlike the present study, spatial data were not taken into account. Thus, the basic architecture provides comparable or better results, confirming its adequacy under conditions of limited computational resources.

The study found that *Helicoverpa armigera* and *Frankliniella occidentalis* were the most reliably predicted species ($R^2 > 0.82$), likely due to the regularity of their reproductive cycles and stable agronomic conditions. Similar findings were reported by Xiong et al. [27] for *Helicoverpa zea* and *Frankliniella* spp. in northern latitudes, where $R^2$=0.85 was achieved when irrigation regularity was accounted for. In contrast to Wang et al. [28], who used complex ensemble methods, the present study demonstrates that an MLP can yield comparable results while remaining technically more accessible. This underscores the importance of achieving an optimal balance between model complexity and input data quality.

Analysis of *Loxostege sticticalis* revealed low predictability ($R^2 \approx 0.64$), which is associated with high interweekly variability, climatic noise, and uneven agrochemical application within sunflower fields. The population dynamics of this species are further complicated by its high ecological plasticity, which substantially limits the model's ability to capture stable patterns. This finding is consistent with Zhang et al. [29] and Hara et al. [30], who also reported low forecast accuracy for pests with irregular development, obtaining $R^2 = 0.60 - 0.65$ when modelling crops with variable microclimatic conditions. However, Mohmed et al. [31], using a hybrid CNN-MLP architecture, reported $R^2 \approx 0.72$ for a similar species when incorporating spatial features and vegetation indices. These findings highlight the limitations of basic MLPs lacking spatiotemporal context and point to the need for architectural adaptation to pest species characterised by non-stationary phenophases and complex responses to agricultural interventions. In future studies, incorporating models with convolutional processing elements and dynamic input parameters could significantly enhance prediction accuracy for such biological entities.

Evaluation of crop models showed that tomato agrocenoses yielded the best results (RMSE$\leq$2.3, MAE$\leq$2.0). This correlates with the findings of Shankar et al. [32] and Wang et al. [33], who reported similarly low error values (RMSE$<$3) for comparable crops under stable drip irrigation, which is associated with the homogeneity of growing conditions and the predictability of pest population dynamics. In tomato fields, biological control methods, such as the use of *Bacillus thuringiensis*, also help to stabilise pest numbers, reducing the amplitude of seasonal fluctuations. Unlike Rizkiana et al. [34], who used ensemble decision trees and achieved MAE=1.8, the model presented here demonstrates comparable accuracy without the architectural complexity and high computational load. Moreover, the strong correlation between observed and predicted pest abundance in tomatoes confirms the model's ability to effectively account for weak signals in the context of relatively low agroecosystem entropy. Thus, when agronomic factors such as irrigation uniformity, planting density, and cultivation regime are considered, even a simple model can demonstrate high reliability and potential applicability for operational monitoring in practice.

Lower accuracy rates were recorded for cotton and maize crops ($R^2 \approx 0.71 - 0.74$). Puoetsile et al. [35], using a Recurrent Neural Network (RNN), obtained $R^2 \approx 0.77$ for the same crops, emphasising the importance of accounting for temporal dependence. However, Fernando et al. [36] showed that without spatial aggregation, models rarely exceed $R^2 = 0.75$. In this context, the results of the present study are consistent with realistic constraints and highlight the need for further structuring of input data.

Optimisation of the model architecture revealed that the best metrics were achieved with two hidden layers comprising 64 and 32 neurons. Similar architectures are described by Tuo and Liu [37], where a model with comparable parameters achieved $R^2$=0.82 when predicting the dynamics of *Spodoptera* spp. In contrast, Pratiwi et al. [38] advocated using a deeper architecture (four layers); however, in that case, overfitting and a decline in test-set performance were observed. This confirms the conclusion that increasing model complexity with a limited dataset provides no tangible advantage.

Furthermore, the observed overfitting effect in complex configurations correlates with the findings of Hassan and Maji [39], who reported that an increase in the number of parameters leads to a loss of generalisation ability in small training sets, particularly in problems with highly variable input data and a lack of spatial stratification. According to Wang et al. [40], such models are prone to overfitting to noise, which significantly reduces their suitability for transfer between regions. Ye et al. [41] reported that the optimal MLP architecture for biological time series typically includes two to three layers with a moderate number of neurons (32-64), enabling a balance between model depth and resistance to overfitting. This architectural strategy ensures sufficient model expressiveness while maintaining stability on test samples, fully consistent with the parameters identified in the present study. It is also emphasised that incorporating additional regularisation mechanisms, such as Dropout and L2 normalisation, can further enhance model robustness without increasing architectural complexity.

The most effective combination of hyperparameters was found to be the ReLU+ADAM pair, as evidenced by the minimal MAE values and maximal $R^2$. Similar results were reported by Wang et al. [42], where the use of ReLU prevented the vanishing gradient problem in deep architectures, while ADAM ensured fast and stable training. In contrast to Pancieri et al. [43], who employed SGD and obtained unstable results, high reproducibility was achieved in the present study across different subsamples. The tanh and sigmoid functions showed poorer performance, particularly when combined with SGD, where significant error variations and a decrease in $R^2$ to 0.63 were observed. Similar trends were described by Singh et al. [44], where sigmoid exhibited gradient saturation. The study by Sandhu et al. [45] further indicated that tanh is less robust to noise, especially when applied to time series from agroecosystems. These comparisons confirm that the choice of activation function and optimiser has a critical impact on the final accuracy of the model.

Statistical tests confirmed the significance of the differences between the configurations. In particular, the normality of residuals in the ReLU+ADAM and RMSprop models is consistent with the findings of Guan [46], where similar architectures demonstrated stable error distributions. Heteroscedasticity identified in sigmoid models was also described as a factor limiting the robustness of such configurations. Therefore, the statistical results obtained in this study objectively validate the choice of the optimal architecture.

Thus, this study demonstrated that a simple two-layer model with appropriately selected hyperparameters is capable of achieving high accuracy in predicting pest populations across various agrocenoses. The identified limitations – relating to overfitting, instability in complex crops, and variability associated with activation function selection – highlight promising directions for further development, including the incorporation of spatial features, enhanced regularisation, and testing of hybrid architectures. The results confirm the practical applicability of the model and its potential integration into operational monitoring and agronomic planning systems.

A critical aspect of implementing neural network models in agricultural practice is their interpretability. While high predictive accuracy is essential, stakeholders such as agronomists and farm managers require clear understanding of how input variables, such as pest species, crop type, irrigation method, and meteorological factors, affect the predictions. An interpretable model enables informed decision-making, facilitates trust in automated forecasts, and allows identification of actionable management strategies. Techniques such as feature importance analysis, SHAP values, or partial dependence plots could provide insight into model behavior and increase its practical utility, ensuring that recommendations derived from the model are both transparent and actionable.

Despite the promising results, the study has several limitations. First, baseline comparisons with simpler models such as Random Forest or linear regression were limited, which constrains the ability to quantify the true advantage of the MLP model. Second, the neural network's interpretability remains limited, which could hinder adoption by practitioners who require clear explanations for predictions. Finally, potential biases in the dataset should be acknowledged, including uneven representation of certain crops or pest species, seasonal variations, and regional differences in agroclimatic conditions. These factors may influence model performance and generalizability, highlighting the need for careful data collection, preprocessing, and future validation in diverse agroecosystems.

## 5. Conclusions

The study demonstrated the effectiveness of neural network models, specifically the MLP with two hidden layers (64 and 32 neurons) and the ReLU activation function, for predicting pest abundance in the agrocenoses of northern Kazakhstan. The model achieved an average $R^2$ value of 0.83, a minimum MAE value of 2.9, and an RMSE of 4.36 when modelling the time series of species such as *Helicoverpa armigera* ($R^2$=0.87, MAE=1.7) and *Frankliniella occidentalis* ($R^2$=0.82, MAE=2.0), as well as *Phyllotreta vittula* on wheat ($R^2$=0.78), confirming the model's robustness for cereal crops. These results validate the model's ability to accurately reproduce seasonal fluctuations in abundance under conditions of stable irrigation and the use of biological control agents such as *Bacillus thuringiensis*.

Significant species- and crop-specific differences in forecast accuracy were observed. The lowest accuracy was recorded for *Loxostege sticticalis*, where $R^2$ was 0.64, MAE reached 6.3, and RMSE was 7.5. This is attributed to high interweekly population variance and heterogeneous conditions in sunflower fields, highlighting the need to adapt the model to the specific characteristics of agrocenoses and incorporate additional spatiotemporal features. Analysis of various architectures showed that increasing the number of hidden layers beyond two (e.g., to four) and the number of neurons above 64 per layer led to overfitting: in these cases, $R^2$ decreased to 0.67 and MAE increased to 4.5. Thus, a balanced two-layer structure, combining accuracy and robustness without excessive parametric complexity, is optimal.

A limitation of the study is the relatively small training sample (2,560 examples) and the absence of spatial coordinates for the plots, which constrains the use of more complex models such as CNNs and LSTMs. In the future work, a convolutional LSTM (ConvLSTM) will be implemented to explicitly model both the spatial adjacency of plots and temporal dynamics simultaneously. A pilot deployment study can also be conducted to evaluate the model's performance and usability in real-world farm management conditions.

## REFERENCES

1. G. Anarbekova, L. G. B. Ruiz, A. Akanova, S. Sharipova, and N. Ospanova, *Fine-tuning artificial neural networks to predict pest numbers in grain crops: A case study in Kazakhstan*, Machine Learning and Knowledge Extraction, vol. 6, no. 2, pp. 1154–1169, 2024.
2. J. Tussupov, G. B. Abdikerimova, A. Ismailova, A. Kassymova, Z. Beldeubayeva, M. Aitimov, and K. Makulov, *Analyzing disease and pest dynamics in steppe crop using structured data*, IEEE Access, vol. 12, pp. 71323–71330, 2024.
3. M. Sadenova, N. Beisekenov, P. S. Varbanov, and T. Pan, *Application of machine learning and neural networks to predict the yield of cereals, legumes, oilseeds and forage crops in Kazakhstan*, Agriculture, vol. 13, no. 6, 1195, 2023.
4. W. Shafik, A. Tufail, C. D. Silva Liyanage, and R. A. Awg Haji Mohd Apong, *Using a novel convolutional neural network for plant pests detection and disease classification*, Journal of the Science of Food and Agriculture, vol. 103, no. 12, pp. 5849–5861, 2023.
5. M. Turkoglu, B. Yanikoglu, and D. Hanbay, *PlantDiseaseNet: Convolutional neural network ensemble for plant disease and pest detection*, Signal, Image and Video Processing, vol. 16, pp. 301–309, 2021.
6. V. S. Dhaka, S. V. Meena, G. Rani, D. Sinwar, M. F. Kavita Ijaz, and M. Woźniak, *A survey of deep convolutional neural networks applied for prediction of plant leaf diseases*, Sensors, vol. 21, no. 14, 4749, 2021.
7. S. Couliably, B. Kamsu-Foguem, D. Kamissoko, and D. Traoré, *Explainable deep convolutional neural networks for insect pest recognition*, Journal of Cleaner Production, vol. 371, 133638, 2022.
8. M. Grünig, E. Razavi, P. Calanca, D. Mazzi, J. D. Wegner, and L. Pellissier, *Applying deep neural networks to predict incidence and phenology of plant pests and diseases*, Ecosphere, vol. 12, no. 10, e03791, 2021.
9. J. Wang, and D. Zhang, *Intelligent pest forecasting with meteorological data: An explainable deep learning approach*, Expert Systems with Applications, vol. 252, 124137, 2024.
10. H. Zhang, S. Zhao, Y. Song, S. Ge, D. Liu, X. Yang, and K. Wu, *A deep learning and Grad-Cam-based approach for accurate identification of the fall armyworm (Spodoptera frugiperda) in maize fields*, Computers and Electronics in Agriculture, vol. 202, 107440, 2024.
11. Seifullin Kazakh Agro Technical Research University, 2025. https://kazatu.edu.kz/
12. Toraighyrov University, 2025. https://tou.edu.kz/en/
13. Republican Methodological Centre for Phytosanitary Diagnostics and Forecasts, 2025. https://www.niac.kz/030340001648
14. World Weather Online, 2025. https://www.worldweatheronline.com/
15. Python 3.10.0, 2021. https://www.python.org/downloads/release/python-3100/

16. TensorFlow 2.11.1, Application Programming Interface (API) documentation, 2023. https://www.tensorflow.org/versions/r2.11/api_docs
17. Keras, 2025. https://keras.io/
18. Scikit-learn, 2025. https://scikit-learn.org/1.2/preface.html
19. Pandas 1.5.0, 2025. https://pypi.org/project/pandas/1.5.0/
20. NumPy 1.24 Release Notes, 2025. https://numpy.org/devdocs/release/1.24.0-notes.html
21. Jupyter, 2025. https://jupyter.org/
22. Matplotlib 3.6.3, 2025. https://matplotlib.org/3.6.3/
23. Seaborn 0.12, 2025. https://seaborn.pydata.org/archive/0.12/index.html
24. Plotly, 2025. https://plotly.com/
25. K. Rimal, K. B. Shah, and A. K. Jha, *Advanced multi-class deep learning convolution neural network approach for insect pest classification using TensorFlow*, International Journal of Environmental Science and Technology, vol. 20, no. 4, pp. 4003–4016, 2022.
26. C.-Y. Tai, W.-J. Wang, and Y.-M. Huang, *Using time-series generative adversarial networks to synthesize sensing data for pest incidence forecasting on sustainable agriculture*, Sustainability, vol. 15, no. 10, 7834, 2023.
27. H. Xiong, J. Li, T. Wang, F. Zhang, and Z. Wang, *EResNet-SVM: An overfitting-relieved deep learning model for recognition of plant diseases and pests*, Journal of the Science of Food and Agriculture, vol. 104, no. 10, pp. 6018–6034, 2024.
28. J. Wang, Y. Li, H. Feng, L. Ren, X. Du, and J. Wu, *Common pests image recognition based on deep convolutional neural network*, Computers and Electronics in Agriculture, vol. 179, 105834, 2020.
29. H. Zhang, T. Yang, and W. Wang, *A novel hybrid model for species distribution prediction using neural networks and Grey Wolf Optimizer algorithm*, Scientific Reports, vol. 14, no. 1, 11505, 2024.
30. P. Hara, M. Piekutowska, and G. Niedbała, *Selection of independent variables for crop yield prediction using artificial neural network models with remote sensing data*, Land, vol. 10, no. 6, 609, 2021.
31. G. Mohmed, X. Heynes, A. Naser, W. Sun, K. Hardy, S. Grundy, and C. Lu, *Modelling daily plant growth response to environmental conditions in Chinese solar greenhouse using Bayesian neural network*, Scientific Reports, vol. 13, no. 1, 4379, 2023.
32. T. Shankar, G. C. Malik, M. Banerjee, S. Dutta, S. Praharaj, S. Lalichetti, S. Mohanty, D. Bhattacharyay, S. Maitra, A. Gaber, A. K. Das, A. Sharma, and A. Hossain, *Prediction of the effect of nutrients on plant parameters of rice by artificial neural network*, Agronomy, vol. 12, no. 9, 2123, 2022.
33. Z. Wang, G. Zhang, H. Pei, Z. Sun, J. Cheng, T. Zhou, C. Geng, K. Lei, and C. Zheng, *Selection of optimal models for predicting growth stress in Artemisia desertorum by comparison of linear regression and multiple neural networks: Take the construction of a green mine in the Bayan Obo mine as an example*, Ecotoxicology and Environmental Safety, vol. 235, 113400, 2022.
34. A. Rizkiana, A. P. Nugroho, N. M. Salma, S. Afif, R. E. Masithoh, L. Sutiarso, and T. Okayasu, *Plant growth prediction model for lettuce (Lactuca sativa.) in plant factories using artificial neural network*, IOP Conference Series: Earth and Environmental Science, vol. 733, 012027, 2021.
35. A. Puoetsile, M. Lekgari, S. Kassa, and G. Mengistu Tsidu, *Optimized parameter estimation and integrating neural network forecasting of dynamic plant-livestock model for early warning in agro-environment control systems*, Statistics, Optimization & Information Computing, vol. 12, no. 5, pp. 1460–1475, 2024.
36. A. Fernando, R. Vicerra, L. Lim, A. Maglaya, N. Ledesma, and J. Gonzaga, *Determination of a tomato growth in a plant chamber using neural network*, in Institute of Electrical and Electronics Engineers (IEEE) 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management, Institute of Electrical and Electronics Engineers, Manila, pp. 1–4, 2021.
37. Y. Tuo, and X. Liu, *Predicting nuclear power plant operational parameters using clustering and mutual information for feature selection and Transformer neural network optimized by TPE*, Frontiers in Energy Research, vol. 12, 1462184, 2024.
38. H. Pratiwi, A. P. Windarto, S. Susliansyah, R. R. Aria, S. Susilowati, L. K. Rahayu, Y. Fitriani, A. Merdekawati, and I. R. Rahadjeng, *Sigmoid activation function in selecting the best model of artificial neural networks*, Journal of Physics: Conference Series, vol. 1471, 012010, 2020.
39. S. Hassan, and A. Maji, *Plant disease identification using a novel convolutional neural network*, IEEE Access, vol. 10, pp. 5390–5401, 2022.
40. B. Wang, C. Zhang, Y. Li, D. Huang, C. Cao, and Y. Gong, *An ultra-lightweight efficient network for image-based plant disease and pest infection detection*, Precision Agriculture, vol. 24, no. 5, pp. 1836–1861, 2023.
41. Y. Ye, Q. Huang, Y. Rong, X. Yu, W. Liang, Y. Chen, and S. Xiong, *Field detection of small pests through stochastic gradient descent with genetic algorithm*, Computers and Electronics in Agriculture, vol. 206, 107694, 2023.
42. K. Wang, M. Abid, A. Rasheed, J. Crossa, S. Hearne, and H. Li, *DNNGP, a deep neural network-based method for genomic prediction using multi-omics data in plants*, Molecular Plant, vol. 16, no. 1, pp. 279–293, 2022.
43. G. P. Pancieri, S. Mota, S. Paes, J. M. Soares, and M. C. Picanço, *Prediction Model of the Temporal Dynamics of Severe Pest Cashew Anacampsis phytomiella Using Artificial Neural Networks*, Journal of Applied Entomology, vol. 149, no. 3, pp. 350–362, 2024.
44. A. Singh, M. Yeasin, R. Paul, A. Paul, and A. Sarkar, *Dynamic ensemble-based machine learning models for predicting pest populations*, Frontiers in Applied Mathematics and Statistics, vol. 10, 1435517, 2024.
45. K. S. Sandhu, D. N. Lozada, Z. Zhang, M. O. Pumphrey, and A. H. Carter, *Deep learning for predicting complex traits in spring wheat breeding program*, Frontiers in Plant Science, vol. 11, 613325, 2021.
46. J. Guan, *Optimization of BP neural network model based on genetic algorithm in nonlinear prediction*, in Institute of Electrical and Electronics Engineers (IEEE) 6th International Conference on Civil Aviation Safety and Information Technology, Institute of Electrical and Electronics Engineers (IEEE), Hangzhou, pp. 1228–1232, 2024.