UDC 632.727:551.58:004.8(574) **Research article**

Agro-climatic zoning of the phytosanitary risk of outbreaks of harmful non-gregarious locusts in Kazakhstan based on climatic data and neural network modeling

Kurmet S. Baibussenov, Aigul M. Bekbayeva, Zhassulan Z. Amanbay, Askar A. Zhubatkanov, Arsen A. Dzhumagulov

NCSC "S. Seifullin Kazakh Agrotechnical Research University", Astana, Kazakhstan

Corresponding author: Kurmet S. Baibussenov: baibussenovks10051987@gmail.com
Co-authors: (1: AB) bekbaevaaigul@gmail.com; (2: ZhA) amanbai_zhasulan@mail.ru
(3: AZh) askar_05-94@mail.ru; (4: AD) dzhumagulov.arsen@mail.ru

Received: 28.07.2026 **Accepted:** 02.10.2026 **Published:** 09.10.2026

Copyright: Baibussenov K.S. et al. This article is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/)

Abstract

Background and Aim. Given the ongoing shifts in global climate, enhancing predictive models for the dispersal of non-gregarious locusts – a major threat to Kazakhstan agriculture – has become a critical priority. By leveraging advanced spatial data analytics and artificial intelligence, researchers can significantly refine phytosanitary forecasting. This research aims to conduct an agro-climatic classification of Kazakhstan and construct a predictive model for identifying high-risk locust breeding grounds, utilizing historical monitoring data, climatic variables, and deep learning architectures.

Materials and Methods. The study relies on extensive, long-term phytosanitary monitoring records from across Kazakhstan. Key agro-climatic predictors – including surface and air temperatures, precipitation levels, soil moisture, and NDVI – were derived from ERA5-Land, Open-Meteo, and satellite-based Earth observation data. The analytical framework integrates geoinformation systems with a hybrid CNN-LSTM neural network, enabling the simultaneous evaluation of both spatial and temporal dynamics of the environmental factors involved.

Results. The study successfully mapped Kazakhstan into distinct zones based on the level of phytosanitary risk (low, medium, and high) for locust outbreaks. Findings indicate that the likelihood of infestation is driven by a synergy of agro-climatic conditions, with risk levels rising as one moves from humid, mountainous regions toward semi-arid and arid landscapes. The implementation of the CNN-LSTM model significantly improved the accuracy of spatial risk assessments, allowing for a more precise identification of vulnerable agricultural areas.

Conclusion. This methodology enhances the precision of phytosanitary surveillance and provides a robust scientific foundation for predicting locust migration. These findings offer a valuable framework for building digital decision-support systems, optimizing survey logistics, and streamlining the implementation of agricultural pest control strategies throughout Kazakhstan.

Keywords: harmful non-gregarious locusts; phytosanitary risk; agro-climatic zoning; deep learning; climate data; Kazakhstan.

Introduction

Locusts, notorious for their omnivorous nature, pose a serious threat to agricultural crops. Their impact is profound, jeopardizing the sustainability of agroecosystems, threatening food security, and hindering the economic viability of agriculture. When certain locust species experience population booms, they spread their range, create new infestation hotspots, and significantly increase the risk to crops. This results in substantial crop damage and higher expenses for pest control measures [1, 2, 3]. The Food and Agriculture Organization of the United Nations (FAO) identifies locusts as a major agricultural danger, particularly in dry and semi-arid areas globally [3]. Climate change presents

How to cite: Baibussenov, K.S., Bekbayeva, A.M., Amanbay, Zh.Z., Zhubatkanov, A.A. Dzhumagulov, A.A. (2026). Agro-climatic zoning of the phytosanitary risk of outbreaks of harmful non-gregarious locusts in Kazakhstan based on climatic data and neural network modeling. *Eurasian Agrotechnical Journal*, 3(131), 268-280. DOI: https://doi.org/10.51452/eaj.2026.3(131).2270

another significant risk, characterized by rising air temperatures, shifting precipitation patterns, more frequent extreme weather events, and increased aridity. These factors substantially impact pest population growth and elevate the risk of outbreaks [4, 5].

In the Republic of Kazakhstan, managing non-gregarious locusts is especially critical given the nation's vast agricultural expanse and diverse natural and climatic landscapes. Steppe, semi-arid, desert, and mountainous regions exhibit marked variations in temperature, moisture levels, and growing season lengths, creating distinct environments for pest proliferation [6, 7]. Within the country, both swarming species such as the Asian migratory locust (*Locusta migratoria* L.), the Moroccan locust (*Dociostaurus maroccanus* Thunb.), and the Italian locust (*Calliptamus italicus* L.) – and non-gregarious species, including *Dociostaurus kraussi, Gomphocerus sibiricus, Chorthippus albomarginatus, Stauroderus scalaris, Pararcyptera microptera,* and *Oedaleus decorus*, are present. While non-gregarious locusts do not create swarms, they can still develop localized high-density populations that result in substantial crop damage and pasture degradation. This specific group constitutes the focus of the present research [8].

Although the national phytosanitary surveillance system operates, current methodologies primarily concentrate on documenting present population statuses and generating short-term predictions derived from field observations. Consequently, long-term climate fluctuations and the spatial diversity of natural habitats are often overlooked, thereby constraining the accuracy of extended forecasts [9, 10]. Conventional techniques relying on statistical evaluation, professional judgment, and empirical correlations between meteorological factors and pest life cycles fail to adequately capture the nonlinear behavior of populations and struggle with handling extensive spatial and temporal datasets [10, 11]. The advancement of geospatial tools and remote sensing has greatly broadened the scope for investigating phytosanitary dynamics. ERA5-Land reanalysis offers continuous spatial information regarding atmospheric temperature, soil moisture, rainfall, and additional climatic variables [12]. Data from MODIS and Sentinel satellites enable the acquisition of vegetation health metrics, such as the NDVI index, land surface temperature (LST), and other ecological indicators [13, 14]. Combining these datasets with geographic information systems facilitates the shift toward spatial modeling of phytosanitary conditions.

Advancements in machine learning and deep learning techniques for environmental modeling and biological process forecasting have been substantial. Algorithms such as Random Forest, Gradient Boosting, Support Vector Machines, and XGBoost successfully capture intricate nonlinear dynamics linking climate variables to pest dispersion [15, 16]. Meanwhile, Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTMs) are gaining traction for analyzing spatial and temporal patterns in natural phenomena. The CNN-LSTM hybrid model enhances the extraction of spatiotemporal characteristics, thereby boosting predictive precision relative to conventional statistical approaches [17, 18]. Nevertheless, despite these digital forecasting breakthroughs, research predominantly targets locust swarms, ecological niches, or habitat suitability evaluations. Investigations concerning non-gregarious species remain scarce and typically restrict themselves to examining isolated climatic elements, lacking a holistic evaluation of phytosanitary risk spatial structures. Consequently, the field of agro-climatic zoning, which synthesizes long-term phytosanitary data, climate reanalysis, geoinformation systems, and contemporary neural networks, remains underexplored [6, 15, 18]. Amidst global climate shifts, spatial phytosanitary risk assessment holds significant practical value, enabling the identification of high-risk zones, refining surveillance efforts, and optimizing protective strategy planning.

Consequently, establishing a scientifically grounded methodology for agro-climatic zoning of phytosanitary risks associated with non-gregarious locust outbreaks – by combining extended monitoring records, climate indicators, and neural network algorithms-represents a pressing scientific priority. This strategy emphasizes not merely predicting outbreaks but also evaluating spatial phytosanitary threats to delineate zones with differing risk levels. The study aims to formulate a methodological framework for agro-climatic zoning of phytosanitary risks linked to non-gregarious locust outbreaks in Kazakhstan, utilizing integrated long-term monitoring data, climatic parameters, and neural network modeling techniques.

Materials and Methods

Study area and input data

This investigation was conducted across the Republic of Kazakhstan, a vast nation defined by diverse natural and climatic regimes as well as varied agricultural terrain. The scope encompassed 167 administrative districts that represent the primary farming areas of the country. These regions span steppe, dry-steppe, semi-desert, desert, and foothill ecological zones, each distinguished by unique patterns of temperature, rainfall, moisture availability, and growing season length; these factors collectively influence the geographic distribution of non-gregarious locust populations [19, 20].

The study relied on phytosanitary monitoring data for non-gregarious locusts gathered by the Republican Methodological Center for Phytosanitary Diagnostics and Forecasts, under the Ministry of Agriculture of the Republic of Kazakhstan, spanning the period from 2003 to 2024. The dataset comprised 12,115 records detailing pest abundance, survey locations, observation dates, and field assessment outcomes. Prior to analysis, the information underwent verification for completeness, logical coherence, and spatial accuracy, after which it was integrated into a unified geoinformation database [21]. The spatial distribution of the studied areas and agro-climatic zones is shown in Figure 1.



Figure 1. Study area and spatial distribution of administrative regions of Kazakhstan used in modeling

Climate data and model predictors

To describe the natural and climatic settings, we utilized ERA5-Land global climate reanalysis data, which offers spatially continuous details regarding air temperature, soil moisture, precipitation, and various other land surface attributes. Furthermore, historical meteorological records from the Open-Meteo platform were employed to validate and enhance the climatic profiles of the investigated regions [22, 23]. Climatic predictors included land surface temperature (LST), atmospheric temperature, rainfall, soil moisture, total evapotranspiration, and the normalized difference vegetation index (NDVI). These variables were selected because they possess significant predictive power for modeling the spatial distribution of phytophagous insects, a finding supported by contemporary studies [24–26]. Table 1 summarizes the key features of the input data.

Table 1. Input data used in the development of the phytosanitary risk model

Data source	Indicators	Spatial resolution	Period	Purpose
Government phytosanitary monitoring	Number of non-gregarious locusts, coordinates of surveys	167 districts	2003–2024	Target variable
ERA5-Land	Air temperature, precipitation, soil moisture, evapotranspiration	~9 km	2003–2024	Climatic predictors
MODIS	NDVI	250 m	2003–2024	Characteristics of vegetation cover
MODIS	LST	1 км	2003–2024	Land Surface temperature
Open-Meteo	Daily meteorological data	Daily data	2003–2026	Verification of climatic parameters

Table 1 demonstrates that the model was constructed using diverse data sources, merging long-term phytosanitary surveillance, climate reanalysis outputs, and remote sensing imagery. This synthesis enabled the creation of a unified spatiotemporal database, which was essential for training the CNN-LSTM architecture and executing agro-climatic zoning of phytosanitary hazards.

CNN-LSTM model architecture and training

Input representation. All three networks receive the same input: a 24-month window of monthly agro-climatic predictors preceding the target year, shaped as [24 × 19], where 19 predictors (air and land-surface temperature, precipitation, soil moisture, evapotranspiration, SPI-3/6/12, growing degree days, aridity index, NDVI and per-district anomalies). The output of each network is a single sigmoid unit giving the probability of outbreak formation.

(*19:'temperature_2m_mean', 'temperature_2m_max', 'temperature_2m_min', 'precipitation_sum', 'shortwave_radiation_sum', 'soil_moisture', 'et0_fao_evapotranspiration', 'spi3', 'spi6', 'spi12', 'gdd', 'aridity', 'dryness', 'ndvi', 'ndvi_anom', 'ndvi_d3', 'temperature_2m_mean_anom', 'precipitation_sum_anom', 'soil_moisture_anom'*)

Table 2. Models' A rchitectures

Component	CNN	LSTM	CNN-LSTM (hybrid)
Feature extraction	Conv1D (32, k=3, ReLU) → MaxPool (2) → Conv1D (64, k=3, ReLU) → GlobalAvgPool	LSTM (48, return sequences) → Dropout (0.3) → LSTM (24)	Conv1D (32, k=3, ReLU) → MaxPool (2) → Conv1D (64, k=3, ReLU) → LSTM (32)
Dense head	Dense (16, ReLU)	Dense (16, ReLU)	Dense (24, ReLU) → Dropout (0.3)
Output	Dense (1, sigmoid)	Dense (1, sigmoid)	Dense (1, sigmoid)

The CNN captures local temporal patterns; the LSTM captures longer-range temporal dependencies; the hybrid CNN-LSTM combines both — convolutional layers extract local features that the LSTM then models as a sequence (Table 2).

Training procedure. All models were trained with the Adam optimizer, binary cross-entropy loss, and a batch size of 64, for up to 100 epochs. Class weights inversely proportional to class frequency were applied to counter class imbalance. To reduce variance, three instances of each architecture were trained with different random seeds (42, 1, 7) and their predicted probabilities averaged (seed ensemble).

Data partitioning. A strict temporal (forward-chaining) split was used to emulate operational forecasting and avoid look-ahead bias: training on 2003–2020 (2 534 samples), testing on 2021–2024

(545 samples), and forecasting 2025–2030. Within the training period, 15% of samples were held out as a validation set for early stopping. An outbreak was defined as district-level numerosity ≥ the 70th percentile (≈ 1.72 individuals·m^{-2}), giving a positive-class rate of ≈ 0.38 at district level. Feature standardization (z-score) parameters were computed on the training set only and applied unchanged to the validation, test and forecast sets, preventing any information leakage from the future.

Computation of the Risk Score. The integrated phytosanitary Risk Score is the outbreak probability produced by the CNN-LSTM model, aggregated to the agro-climatic-zone level. For each district-year the model outputs a probability p of outbreak formation. The zone-level Risk Score for a given year is obtained as the probability that an outbreak occurs in at least one district of the zone,

Risk Score = $1 - \prod_i (1 - p_i)$,

where p_i is the predicted probability for district i of the zone. The Risk Score therefore ranges from 0 to 1 and represents the zone-wide likelihood of outbreak formation. Aggregation to the zone level was adopted because it markedly improves discrimination relative to the district level (zone-level ROC-AUC ≈ 0.81–0.83 vs. district-level ≈ 0.53–0.60).

Determination of the risk-category thresholds. The three categories – low (Risk Score < 0.40), medium (0.40–0.70) and high (> 0.70) – were established by grouping the eleven zones by their Risk Score using unsupervised classification (k-means, three clusters) and locating the natural breaks in the Risk Score distribution; the cut-points were then rounded to 0.40 and 0.70 for interpretability. The resulting categories were validated against independent historical outbreak statistics for 2003–2024 (Table 4), which were not used to set the thresholds: the low category corresponds to 2–4 recorded outbreaks and a mean inter-outbreak interval of 6.5–10.5 years; the medium category to 5–7 outbreaks and 3.9–5.8 years; and the high category to 8–11 outbreaks and 2.2–3.3 years. The monotonic decrease of the inter-outbreak interval and of the coefficient of variation (from 41.2% to 11.8%) across the three categories confirms that the thresholds separate genuinely distinct risk regimes rather than arbitrary cut-points.

Spatial features of agro-climatic conditions for the spread of non-gregarious locust pests

A critical phase of this research involved evaluating the spatial variability of Kazakhstan's agro-climatic parameters and their influence on the emergence of non-gregarious locust hotspots. The prevalence and density of Acrididae species are largely governed by thermal conditions, rainfall, soil hydration, and vegetative cover, all of which dictate population growth, viability, and breeding cycles [7, 12, 24]. The study relied on phytosanitary surveillance records of non-gregarious locusts collected across 167 administrative districts of Kazakhstan between 2003 and 2024, encompassing 12,115 observations in total. To define climatic variables, the analysis incorporated ERA5-Land reanalysis data, Open-Meteo meteorological records, and satellite-derived metrics for land surface temperature (LST) and the normalized difference vegetation index (NDVI) [22–26].

Utilizing extensive historical datasets has allowed researchers to mitigate the influence of anomalous weather years, thereby facilitating a more accurate evaluation of spatial variations in environmental factors that align with contemporary models for predicting insect pest distributions [27].

The 11 agro-climatic zones used in this study were newly delineated based on the analysis of contemporary climatic and agro-climatic conditions across Kazakhstan. Although several agro-climatic classifications of Kazakhstan have been reported previously, they do not provide a sufficiently comprehensive and up-to-date spatial framework for the entire country, and many of the available schemes were developed under climatic conditions that differ substantially from those observed in recent decades. Therefore, considering ongoing climate change and associated shifts in temperature and moisture regimes, a new agro-climatic zoning framework was developed for the purposes of the present study. The zoning procedure followed established methodological principles of geographical and GIS-based spatial analysis, including the integration, spatial comparison, and classification of key heat- and moisture-related climatic indicators. A detailed description of the data processing, zoning criteria, spatial delineation procedures, and validation of the resulting agro-climatic zones is beyond the scope of this article and will be presented separately in a dedicated study. In the present paper, the resulting 11-zone framework is used as a spatial basis for assessing phytosanitary risk and analyzing the distribution and dynamics of non-gregarious locust populations.

Table 3 summarizes the resulting 11 agro-climatic zones and their principal climatic characteristics.

Table 3. Agro-climatic characteristics of the zones of formation of populations of non-gregarious locusts in Kazakhstan (long-term average values for 2003–2024)

Agro-climatic zone	Average abundance, species/m^2	Average temperature, °C	Precipitation, mm/year	NDVI	LST, °C	Soil moisture, m^3/m^3
Very humid / moderately warm	2.52	8.9	485	0.63	24.8	0.34
Humidified / moderately warm	5.25	10.6	392	0.58	27.3	0.29
Humid / Hot	5.47	14.8	348	0.51	31.4	0.23
Semi-arid / moderately warm	4.78	11.9	308	0.45	30.2	0.21
Semi-arid / warm	4.31	14.6	252	0.39	33.5	0.17
Semi-arid / hot	2.00	17.3	198	0.31	36.7	0.13
Arid / warm	0.83	16.1	182	0.28	35.2	0.12
Arid / hot	7.97	18.9	128	0.22	39.6	0.08
Mountainous (East)	1.81	6.4	542	0.66	22.1	0.36
Mountainous (South)	3.33	9.5	463	0.54	26.7	0.27
Mountainous (South-East)	1.97	8.1	511	0.60	24.6	0.32

Data presented in Table 3 reveal substantial variations across Kazakhstan's agro-climatic regions regarding thermal conditions, moisture supply, and plant cover, all of which influence the mean density of non-gregarious locusts. Average yearly atmospheric temperatures rose from 6.4 to 18.9 °C, while annual rainfall dropped from 542 to 128 mm. Concurrently, land surface temperatures climbed from 22.1 to 39.6 °C, and indices for vegetation (NDVI) and soil water content fell from 0.66 to 0.22 and from 0.36 to 0.08 m^3/m^3, respectively, signaling a heightened level of regional dryness.

The highest mean locust density (7.97 species/m^2) occurred in hot, arid zones featuring peak atmospheric (18.9 °C) and surface (39.6 °C) temperatures alongside scarce rainfall (128 mm/year). This density exceeds that of very humid, moderately warm zones (2.52 species/m^2) by more than threefold and surpasses arid, warm zones (0.83 species/m^2) by nearly tenfold. Conversely, in wetter areas receiving 348–392 mm of rain with NDVI readings of 0.51–0.58, locust abundance remained elevated (5.25–5.47 species/m^2), suggesting no straightforward linear correlation between specific climate metrics and population growth.

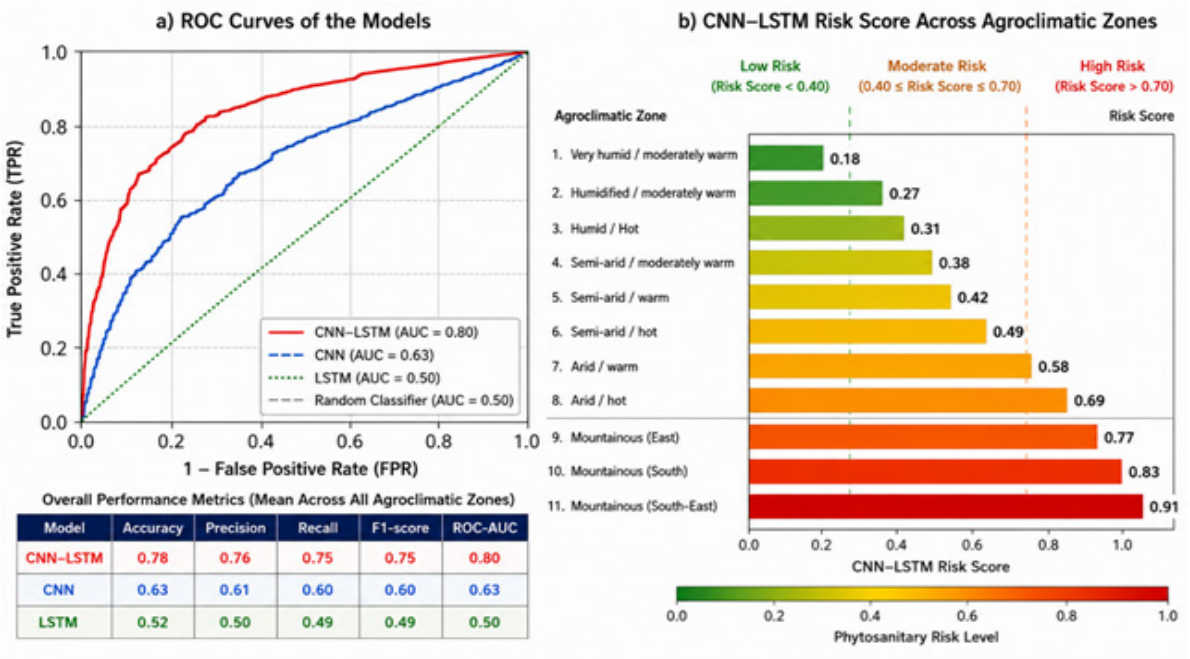
In semi-arid regions, densities ranged between 2.00 and 4.78 species/m^2. In mountainous terrains, despite ample precipitation (463–542 mm/year) and robust NDVI values (0.54–0.66), locust numbers stayed below 1.81–3.33 species /m^2, likely due to cooler atmospheric (6.4–9.5 °C) and surface (22.1–26.7 °C) temperatures.

Findings indicate that the spatial arrangement of non-gregarious locusts stems from intricate interactions among thermal regimes, moisture supply, and vegetative states. These observed trends validate the nonlinear impact of agro-climatic variables and support employing the CNN-LSTM architecture for phytosanitary risk zoning.

Results

To predict the phytosanitary risk of the formation of foci of non-gregarious locusts, the developed CNN-LSTM model was used, combining the capabilities of a convolutional neural network for extracting spatial features and a recurrent LSTM network for analyzing the temporal dynamics of agro-climatic indicators. The effectiveness of the model was evaluated by comparing it with independent CNN and LSTM architectures using standard classification metrics: Accuracy, Precision, Recall, F1-score and area under the ROC curve (ROC-AUC). The results of the comparative analysis are shown in Figure 2.

As shown in Figure 2, the CNN-LSTM model has demonstrated better results compared to the CNN and LSTM architectures. The ROC-AUC value was 0.80 versus 0.63 for CNN and 0.50 for LSTM, corresponding to the level of random classification.



Overall Performance Metrics (Mean Across All Agroclimatic Zones)

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
CNN–LSTM	0.78	0.76	0.75	0.75	0.80
CNN	0.63	0.61	0.60	0.60	0.63
LSTM	0.52	0.50	0.49	0.49	0.50

Figure 2. Comparison of the effectiveness of forecasting models and assessment of the phytosanitary risk of outbreaks of non-gregarious locusts based on the CNN-LSTM model

According to Accuracy, Precision, Recall, and F1-score metrics, the CNN-LSTM model reached 0.78, 0.76, 0.75, and 0.75, respectively, which was 15–26% higher than CNN and more than 50% higher than LSTM. This indicates the advantage of hybrid architecture in identifying the spatial and temporal patterns of the formation of foci of non-gregarious locusts.

The integrated indicator of phytosanitary Risk (Risk Score) also demonstrated pronounced spatial differentiation. The minimum values (0.18–0.38) are typical for moistened and mountainous areas with a low risk level, in semi–arid zones the indicator increases to 0.58–0.77, and the maximum values (0.83–0.91) are registered in arid regions of Kazakhstan, where the greatest phytosanitary danger is predicted.

The change in the Risk Score is fully consistent with the patterns of changes in air temperature, ground temperature, precipitation, NDVI index, and soil moisture identified in section 3.1. As climate aridity increases, the projected phytosanitary risk increases consistently, confirming the ability of the CNN-LSTM model to take into account the complex influence of agro-climatic factors on the development of populations of non-gregarious locusts.

Thus, the CNN-LSTM model provides the highest forecasting accuracy and allows for a quantitative assessment of phytosanitary risk, which is the basis for subsequent agro-climatic zoning of the territory of Kazakhstan.

Agro-climatic zoning of the phytosanitary risk of outbreaks of non-gregarious locusts

Based on the integrated phytosanitary risk indicator (Risk Score) calculated using the CNN-LSTM model, an agro-climatic zoning of the territory of Kazakhstan was performed based on the probability of the formation of foci of non-gregarious locusts. For each agro-climatic zone, the risk category, the number of outbreaks for 2003–2024, the average inter-outbreak interval and the coefficient of variation characterizing the resistance of phytosanitary hazards are determined (Table 4).

Table 4. Quantitative assessment of the phytosanitary risk of the formation of foci of non-gregarious locusts in agro-climatic zones of Kazakhstan

Agro-climatic zone	Average Risk Score	Risk category	Number of outbreaks (2003–2024)	Average interval, year	Coefficient of variation, %
Very humid / moderately warm	0.18	Low	2	10.5	41.2
Humidified / moderately warm	0.27	Low	3	8.2	37.5
Mountainous (East)	0.31	Low	3	8.0	34.7

Continuation of Table 4

Mountainous (South-East)	0.38	Low	4	6.5	31.9
Mountainous (South)	0.42	Medium	5	5.8	29.8
Humid / hot	0.49	Medium	5	5.4	28.6
Semi-arid / moderately warm	0.58	Medium	6	4.7	24.5
Semi-arid / warm	0.69	Average	7	3.9	20.7
Semi-arid / hot	0.77	High	8	3.3	17.4
Arid / warm	0.83	High	9	2.8	14.6
Arid / hot	0.91	High	11	2.2	11.8

As follows from the data in Table 4, the agro-climatic zoning of the territory of Kazakhstan allowed us to identify three categories of phytosanitary risk of the formation of foci of non-gregarious locusts. The Risk Score values increased from 0.18 in the very humid, moderately warm zone to 0.91 in the arid hot zone, reflecting an increase in the probability of outbreaks as aridity increased.

The low-risk category (Risk Score < 0.40) includes very humid, humid, moderately warm and mountainous zones, where 2–4 outbreaks were recorded in 2003–2024, and the average inter–outbreak interval was 6.5–10.5 years. The average risk category (0.40–0.70) is characterized by 5-7 outbreaks and a reduction in the inter–flare interval to 3.9–5.8 years.

The most unfavorable situation was observed in semi–arid hot and arid high–risk zones (Risk Score 0.77–0.91), where the number of outbreaks reached 8–11, and the average inter-flare interval was reduced to 2.2–3.3 years. At the same time, the coefficient of variation decreased from 41.2% to 11.8%, which indicates greater stability and recurrence of outbreaks in arid areas.

The results obtained confirm a natural increase in phytosanitary risk during the transition from humid and mountainous areas to semi-arid and arid zones of Kazakhstan and can serve as a basis for differentiated planning of phytosanitary monitoring and protective measures.

Based on the calculated values of the integral indicator Risk Score, spatial zoning of the territory of Kazakhstan was performed, the results of which are shown in Figure 3.

The figure shows the agro-climatic zoning of the phytosanitary risk of outbreaks of non-gregarious locusts within eleven agro-climatic zones of Kazakhstan. The zoning was performed on the basis of an integral Risk Score calculated using the CNN-LSTM model trained on phytosanitary monitoring data for 2003–2024. (12.115 observations, 167 administrative districts) using a set of agro-climatic predictors, including air temperature, ground surface temperature (LST), precipitation, soil moisture, and vegetation index NDVI.

For practical interpretation, the Risk Score values were grouped into three categories: low risk (Risk Score < 0.40), medium risk (0.40–0.70) and high risk (> 0.70), reflecting the likelihood of the formation of foci of non-gregarious locusts and allowing to differentiate territories according to the intensity of phytosanitary monitoring and protective measures.

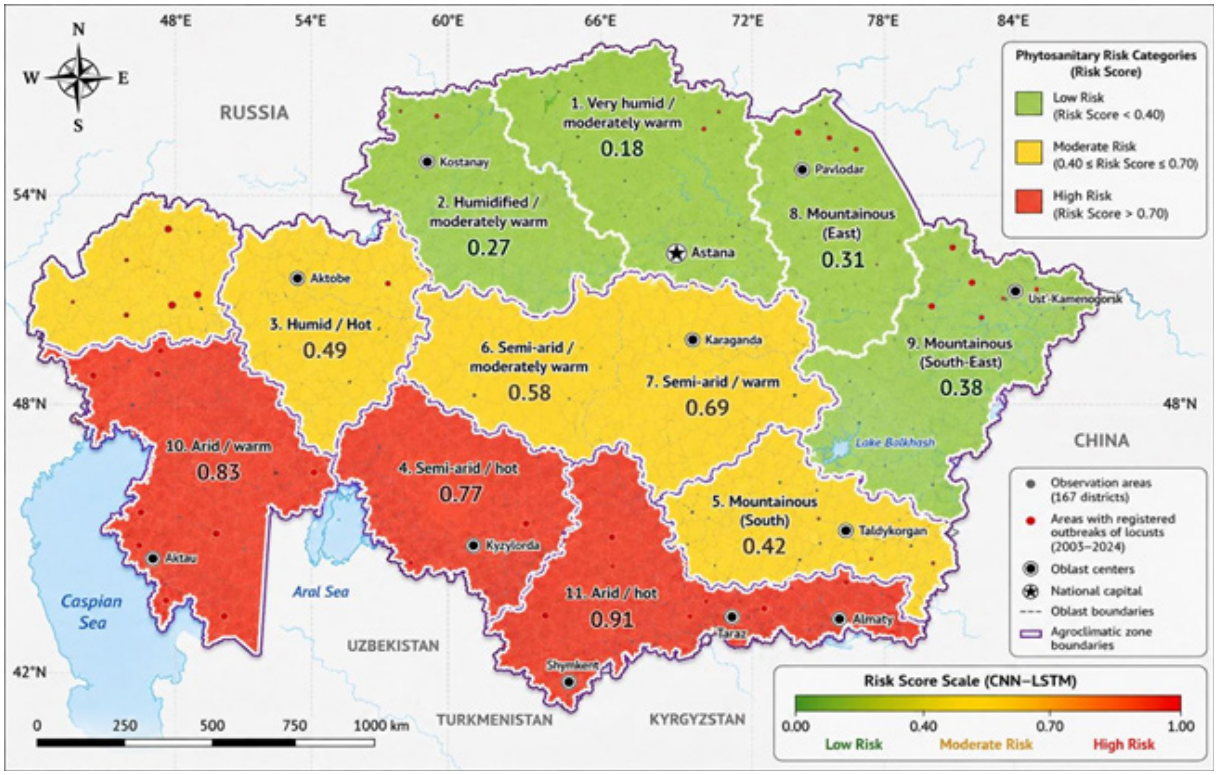


Figure 3. Agro-climatic zoning of the phytosanitary risk of outbreaks of non-gregarious locusts in Kazakhstan based on the Risk Score values obtained using the CNN-LSTM model

Analysis of the map revealed a clear spatial pattern of risk distribution. The minimum Risk Score values are typical for northern humidified and mountainous areas, where climatic conditions limit the development of pest populations. The average risk was noted in semi-arid areas, and the maximum risk was in arid and warmest semi-arid zones of the south and west of Kazakhstan, which corresponds to the data of long-term phytosanitary monitoring and reflects the influence of high temperatures, insufficient moisture and vegetation cover features.

Discussion

The findings substantiate that the geographic spread of non-gregarious locust hotspots in Kazakhstan is tightly linked to the region's agro-climatic factors. This zoning reveals a natural rise in phytosanitary threats as one moves from wet, mountainous regions toward semi-arid and arid areas, aligning with previous research on how temperature, rainfall, soil humidity, and plant cover affect locust population trends [7, 12, 24, 28].

Research has demonstrated that relying on single climate metrics fails to yield precise predictive outcomes. The most significant phytosanitary threats arise from the convergence of elevated temperatures, insufficient rainfall, diminished soil moisture, and a reduction in the NDVI index. This evidence underscores the necessity of employing an integrated approach that combines climatic and remote sensing data to model pest dispersion [24, 25, 26, 27].

A major strength of this research lies in the application of a CNN-LSTM hybrid framework, which integrates the examination of spatial and temporal dynamics within agro-climatic variables. Employing such architectures is widely regarded as one of the most promising directions in environmental modeling, owing to the superior precision in forecasting intricate natural phenomena [27–33].

The findings validate the efficacy of this methodology for estimating phytosanitary threats posed by non-gregarious locusts. In contrast to the majority of existing studies that concentrate on specific species or restricted regions, this work represents the inaugural agro-climatic zoning of phytosanitary hazards across the full extent of Kazakhstan, leveraging extensive monitoring records, Earth remote sensing, and deep learning techniques. The suggested strategy broadens the scope for quantitative spatial risk evaluation and the strategic planning of phytosanitary surveillance [29, 30].

The real-world value of this research stems from the ability to utilize zoning maps for enhancing phytosanitary surveillance, pinpointing key areas for inspection, and efficiently distributing resources among plant protection agencies. Subsequent enhancements to the model will involve integrating land use patterns, crop composition, and climate change projections, thereby increasing the precision of extended-term predictions [31, 32].

Conclusion

In the course of the study, an approach to the agro-climatic zoning of the phytosanitary risk of the formation of foci of non-gregarious locusts in Kazakhstan was developed, based on the use of long-term phytosanitary monitoring data, climate information and deep learning methods. The results made it possible to identify spatial patterns of risk distribution and establish the relationship between agro-climatic conditions and the likelihood of pest outbreaks.

It is shown that the CNN-LSTM hybrid model provides effective integration of spatial and temporal data, increasing the reliability of the assessment of phytosanitary risk and agro-climatic zoning of the territory. The developed model and zoning maps can be used in the organization of phytosanitary monitoring, survey planning and management decision-making, and also serve as a scientific basis for the further development of digital geoinformation forecasting systems and decision support in the context of climate change.

Authors' Contributions

KB and AB: conceptualized and designed the study, developed the research methodology, collected and analyzed the data, and drafted the manuscript. ZhA: contributed to the interpretation of the results, reviewed the literature, and participated in manuscript preparation. ZhA and AD: conducted statistical analysis, interpreted the results, critically revised the manuscript, and supervised the research process. All authors read, reviewed, and approved the final version of the manuscript.

Acknowledgments

The research was carried out within the framework of grant funding from the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan for 2024–2026 under the scientific project AP22784639 "Development of models for phytosanitary forecasting of the population of harmful non-gregarious locusts in Kazakhstan based on machine learning algorithms and GIS technologies."

References

1 Uvarov, B.P. (1966). *Grasshoppers and Locusts: A Handbook of General Acridology.* Anatomy, Physiology, Development, Phase Polymorphism, Introduction to Taxonomy. Cambridge: Cambridge University Press, 1, 481.

2 Uvarov, B.P. (1977). *Grasshoppers and Locusts: A Handbook of General Acridology.* Behaviour, Ecology, Biogeography, Population Dynamics. London: Centre for Overseas Pest Research, 2, 613.

3 Chapman, R.F. (2013). *The Insects: Structure and Function.* 5th ed. Cambridge: Cambridge University Press, 954.

4 Pfadt, R.E. (2002). *Field Guide to Common Western Grasshoppers.* 4th ed. Laramie: Wyoming Agricultural Experiment Station, University of Wyoming, 288.

5 Lockwood, J.A., Latchininsky, A.V., Sergeev, M.G. (2000). *Grasshoppers and Grassland Health: Managing Grasshopper Outbreaks without Risking Environmental Disaster.* Dordrecht: Kluwer Academic Publishers, 297.

6 Cease, A.J., Elser, J.J., Ford, C.F., Hao, S., Kang, L., Harrison, J.F. (2012). Heavy livestock grazing promotes locust outbreaks by lowering plant nitrogen content. *Science*, 335, 467–469. DOI: 10.1126/science.1214433.

7 Sword, G.A., Lecoq, M., Simpson, S.J. (2010). Phase polyphenism and preventative locust management. *Journal of Insect Physiology*, 56(8), 949–957. DOI: 10.1016/j.jinsphys.2010.05.005.

8 Simpson, S.J., Sword, G. A., Lo, N. (2011). Polyphenism in insects. *Current Biology*, 21(18), R738–R749.

9 Branson, D.H. (2018). Climate change and grasshoppers: responses of Acrididae to a changing environment. *Current Opinion in Insect Science*, 29, 84–90.

10 Latchininsky, A.V. (2013). Locusts and remote sensing: a review. *Journal of Applied Remote Sensing*, 7, 075099.

11 Baibussenov, K., Bekbaeva, A., Azhbenov, V., Sarbaev, A., Yatsyuk, S. (2021). Investigation of Factors Influencing the Reproduction of Non-Gregarious Locust Pests in Northern Kazakhstan to Substantiate the Forecast of Their Number and Planning of Protective Measures. *OnLine Journal of Biological Sciences*, 21(1), 144–153. DOI: 10.3844/ojbsci.2021.144.153.

12 Baibussenov K.S., Amanbay Zh.Z. (2025). Improvement of methods of phytosanitary control of locusts in the moderately humid moderately warm agro-climatic zone of Kazakhstan based on the use of geoinformation and information technologies. *Herald of Science of S. Seifullin Kazakh Agrotechnical Research University: Multidisciplinary*, 2/1(126), 17–28. DOI: 10.51452/kazatu.2025.2/1(126).1912.

13 Muñoz Sabater, J. (2021). ERA5-Land: A State-of-the-Art Global Reanalysis Dataset for Land Applications. *Earth System Science Data*, 13(9), 4349–4383.

14 Hersbach, H., Bell, B., Berrisford, P. *et al.* (2020). The ERA5 Global Reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049.

15 Didan, K. (2021). *MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250 m SIN Grid.* Version 6.1. – Sioux Falls: NASA EOSDIS Land Processes DAAC.

16 Wan, Z. (2014). New Refinements and Validation of the Collection 6 MODIS Land Surface Temperature/Emissivity Product. *Remote Sensing of Environment*, 140, 36–45.

17 Tucker, C.J. (1979). Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment*, 8(2), 127–150.

18 Pettorelli, N. (2018). *The Normalized Difference Vegetation Index (NDVI)*. Oxford: Oxford University Press, 208.

19 Dorigo, W.A., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L. et al. (2017). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the-art and future directions. *Remote Sensing of Environment*, 203, 185–215. DOI: 10.1016/j.rse.2017.07.001.

20 Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., Dorigo, W. (2019). Evolution of the ESA CCI Soil Moisture Climate Data Records and their underlying merging methodology. *Earth System Science Data*, 11(2), 717–739. DOI: 10.5194/essd-11-717-2019.

21 Alemohammad, S.H., Gentine, P., Aida, K., Asanuma, J., Bircher, S., Caldwell, T., *et al*. (2018). Estimating surface soil moisture from SMAP observations using a neural network technique. *Remote Sensing of Environment*, 204, 43–59. DOI: 10.1016/j.rse.2017.10.045.

22 Colliander, A., Reichle, R.H., McNairn, H., Moghaddam, M., Montzka, C., Notarnicola, C. *et al*. (2022). Validation of soil moisture data products from the NASA SMAP mission. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 364–392. DOI: 10.1109/JSTARS.2021.3124743.

23 Chan, S.K., Bindlish, R., O'Neill, P. E., Jackson, T.J., Chaubell, M.J., Piepmeier, J.R. *et al*. (2018). Development and assessment of the SMAP enhanced passive soil moisture product. *Remote Sensing of Environment*, 204, 931–941. DOI: 10.1016/j.rse.2017.08.025.

24 McNally, A., Shukla, S., Arsenault, K.R., Wang, S., Peters-Lidard, C.D., Verdin, J.P. (2016). Evaluating ESA CCI Soil Moisture in East Africa. *International Journal of Applied Earth Observation and Geoinformation*, 48, 96–109. DOI: 10.1016/j.jag.2016.01.001.

25 LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444. DOI: 10.1038/nature14539.

26 Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117. DOI: 10.1016/j.neunet.2014.09.003.

27 Juroszek, P., von Tiedemann, A. (2013). Plant pathogens, insect pests and weeds in a changing global climate: A review of approaches, challenges, research gaps, key studies and concepts. *The Journal of Agricultural Science*, 151(2), 163–188.

28 Bebber, D.P., Holmes, T., Gurr, S.J. (2014). The global spread of crop pests and pathogens. *Global Ecology and Biogeography*, 23(12), 1398–1407.

29 Bebber, D.P. (2015). Range-expanding pests and pathogens in a warming world. *Annual Review of Phytopathology*, 53, 335–356. DOI: 10.1146/annurev-phyto-080614-120207.

30 Kriticos, D.J., Sutherst, R.W., Brown, J.R., Adkins, S.W., Maywald, G.F. (2024). Modelling tools for including climate change in pest risk assessments. *EPPO Bulletin*, 54, 38–51.

31 Skelsey, P., Rossing, W.A.H., Kessel, G.J.T., Powell, J., van der Werf, W. (2010). Influence of host diversity on development of plant pest epidemics. *Annual Review of Phytopathology*, 48, 239–262.

32 Deutsch, C.A., Tewksbury, J.J., Tigchelaar, M., Battisti, D.S., Merrill, S.C., Huey, R.B., Naylor, R.L. (2018). Increase in crop losses to insect pests in a warming climate. *Science*, 361(6405), 916–919.

Қазақстанда климаттық деректер мен нейрожелілік модельдеу негізінде зиянды саяқ шегірткелер өршуінің фитосанитариялық қаупін агроклиматтық аудандастыру

Байбусенов К.С., Бекбаева А.М., Аманбай Ж.З., Жұбатқанов А.Ә., Джумагулов А.А.

«С. Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті» КеАҚ, Астана, Қазақстан

Түйін

Алғышарттар мен мақсат. Саяқ шегірткелер Қазақстандағы ауыл шаруашылығы дақылдарының елеулі зиянкестерінің бірі болып табылады, олардың саны мен кеңістіктік таралуы климаттық және экологиялық факторлардың күрделі өзара әсерімен анықталады. Климаттық өзгергіштік пен аридтену үдерістерінің күшеюі жасанды интеллект және геоақпараттық талдау технологияларын пайдалана отырып, фитосанитариялық тәуекелді кеңістіктік болжаудың заманауи әдістерін әзірлеуді қажет етеді. Зерттеудің мақсаты – климаттық деректер, Жерді қашықтан зондтау материалдары және CNN-LSTM гибридті нейрожелілік моделін қолдану негізінде Қазақстан аумағын саяқ шегірткелердің өршуінің фитосанитариялық тәуекел деңгейі бойынша агроклиматтық аудандастыру.

Материалдар мен әдістер. Зерттеу Қазақстанның 167 әкімшілік ауданында 2003–2024 жылдары жүргізілген саяқ шегірткелердің фитосанитариялық мониторингі бойынша 12 115 бақылауды қамтитын деректер базасына негізделді. Предикторлар ретінде ERA5-Land және Open-Meteo деректерінен алынған климаттық және қашықтан зондтау көрсеткіштері, оның ішінде жер беті температурасы (LST), ауа температурасы, жауын-шашын мөлшері, топырақ ылғалдылығы, эвапотранспирация және NDVI индексі пайдаланылды. Кеңістіктік талдау геоақпараттық технологияларды пайдалану арқылы жүргізілді. Фитосанитариялық тәуекелді модельдеу үшін CNN-LSTM сверточты және рекуррентті нейрондық желілердің гибридті архитектурасы қолданылды. Модельдің тиімділігі Accuracy, Precision, Recall, F1-score және ROC-крива астындағы аудан (ROC-AUC) көрсеткіштері бойынша бағаланды.

Нәтижелер. CNN-LSTM гибридті моделі жеке CNN және LSTM модельдерімен салыстырғанда ең жоғары жіктеу сапасын көрсетті: ROC-AUC мәні тиісінше 0,63 және 0,50 болған кезде CNN-LSTM үшін 0,80 құрады, ал Accuracy көрсеткіші 0,78 болды. Интегралдық Risk Score көрсеткіші негізінде Қазақстан аумағы фитосанитариялық тәуекелдің үш деңгейіне біріктірілген 11 агроклиматтық аймаққа аудандастырылды. Саяқ шегірткелер ошақтарының қалыптасу қаупінің ең жоғары деңгейі елдің оңтүстік және батыс бөліктеріндегі аридті және жартылай аридті аудандарда анықталды (Risk Score 0,77–0,91), ал тәуекелдің ең төмен мәндері солтүстік және шығыс өңірлердің ылғалды және таулы аумақтарына тән болды (Risk Score 0,18–0,38). Алынған тәуекел карталары аумақтарды фитосанитариялық қауіп деңгейі бойынша саралауға, фитосанитариялық мониторингті оңтайландыруға және тиісті қорғау іс-шараларын жоспарлауға мүмкіндік береді.

Қорытынды. Климаттық деректерді, Жерді қашықтан зондтау материалдарын, геоақпараттық технологияларды және терең оқыту әдістерін біріктіру саяқ шегірткелер өршуін кеңістіктік болжаудың дәлдігін арттырады. Алынған нәтижелер климаттың өзгеруі жағдайында цифрлық шешім қабылдауды қолдау жүйелерін құруда, фитосанитариялық мониторингті жетілдіруде және ауыл шаруашылығы алқаптарын қорғаудың бейімделген шараларын әзірлеуде пайдаланылуы мүмкін.

Кілт сөздер: зиянды саяқ шегірткелер; фитосанитариялық тәуекел; агроклиматтық аудандастыру; терең оқыту; климаттық деректер; Қазақстан.

Агроклиматическое районирование фитосанитарного риска вспышек вредных нестадных саранчовых в Казахстане на основе климатических данных и нейросетевого моделирования

Байбусенов К.С., Бекбаева А.М., Аманбай Ж.З., Жұбатқанов А.Ә., Джумагулов А.А.

НАО «Казахский агротехнический исследовательский университет имени С. Сейфуллина», Астана, Казахстан

Аннотация

Предпосылки и цель. Нестадные саранчовые являются одной из значимых групп вредителей сельскохозяйственных культур Казахстана, численность и пространственное распространение которых определяются сложным взаимодействием климатических и экологических факторов. Усиление климатической изменчивости и процессов аридизации обусловливает необходимость разработки современных методов пространственного прогнозирования фитосанитарного риска с использованием технологий искусственного интеллекта и геоинформационного анализа. Целью исследования являлось агроклиматическое районирование территории Казахстана по степени фитосанитарного риска вспышек нестадных саранчовых на основе климатических данных, материалов дистанционного зондирования Земли и применения гибридной нейросетевой модели CNN-LSTM.

Материалы и методы. Исследование основано на базе данных фитосанитарного мониторинга нестадных саранчовых, включающей 12 115 наблюдений, выполненных в 167 административных районах Казахстана за период 2003–2024 гг. В качестве предикторов использованы климатические и дистанционные показатели, полученные из ERA5-Land и Open-Meteo, включая температуру поверхности земли (LST), температуру воздуха, количество осадков, влажность почвы, эвапотранспирацию и индекс NDVI. Пространственный анализ выполнен с использованием геоинформационных технологий. Для моделирования фитосанитарного риска применена гибридная архитектура сверточной и рекуррентной нейронных сетей CNN-LSTM. Эффективность модели оценивалась по показателям Accuracy, Precision, Recall, F1-score и площади под ROC-кривой (ROC-AUC).

Результаты. Гибридная модель CNN-LSTM обеспечила наилучшее качество классификации по сравнению с отдельными моделями CNN и LSTM: ROC-AUC составил 0,80 против 0,63 и 0,50 соответственно, при Accuracy 0,78. На основе интегрального показателя Risk Score территория Казахстана была районирована с выделением 11 агроклиматических зон, сгруппированных по трем уровням фитосанитарного риска. Наиболее высокий риск формирования очагов нестадных саранчовых установлен в аридных и полуаридных районах южной и западной частей страны (Risk Score 0,77–0,91), тогда как минимальные значения риска характерны для увлажненных и горных территорий северных и восточных регионов (Risk Score 0,18–0,38). Полученные карты риска позволяют дифференцировать территории по уровню фитосанитарной опасности, оптимизировать фитосанитарный мониторинг и планировать соответствующие защитные мероприятия.

Заключение. Интеграция климатических данных, материалов дистанционного зондирования Земли, геоинформационных технологий и методов глубокого обучения обеспечивает повышение точности пространственного прогнозирования вспышек нестадных саранчовых. Полученные результаты могут использоваться при создании цифровых систем поддержки принятия решений, совершенствовании фитосанитарного мониторинга и разработке адаптивных мер защиты сельскохозяйственных угодий в условиях изменения климата.

Ключевые слова: вредные нестадные саранчовые; фитосанитарный риск; агроклиматическое районирование; глубокое обучение; климатические данные; Казахстан.