

Article

Forecasting *Dendrolimus sibiricus* Outbreaks: Data Analysis and Genetic Programming-Based Predictive Modeling

Ivan Malashin ^{1,*}, Igor Masich ^{1,2,3}, Vadim Tynchenko ^{1,2,3,*}, Vladimir Nelyub ^{1,4}, Aleksei Borodulin ¹,
Andrei Gantimurov ¹, Guzel Shkaberina ^{2,3} and Natalya Rezova ^{2,3}

- ¹ Artificial Intelligence Technology Scientific and Education Center, Bauman Moscow State Technical University, 105005 Moscow, Russia; is.masich@gmail.com (I.M.)
² Laboratory “Hybrid Methods of Modeling and Optimization in Complex Systems”, Siberian Federal University, 79 Svobodny Prospekt, 660041 Krasnoyarsk, Russia; z_guzel@mail.ru (G.S.)
³ Institute of Informatics and Telecommunications, Reshetnev Siberian State University of Science and Technology, 31 Krasnoyarsky Rabochoy Prospekt, 660037 Krasnoyarsk, Russia
⁴ Scientific Department, Far Eastern Federal University, 690922 Vladivostok, Russia
* Correspondence: i.malashin@emtc.ru (I.M.); vtynchenko@emtc.ru (V.T.);
Tel.: +7-926-875-7128 (I.M.)

Abstract: This study presents an approach to forecast outbreaks of *Dendrolimus sibiricus*, a significant pest affecting taiga ecosystems. Leveraging comprehensive datasets encompassing climatic variables and forest attributes from 15,000 taiga parcels in the Krasnoyarsk Krai region, we employ genetic programming-based predictive modeling. Our methodology utilizes Random Forest algorithm to develop robust forecasting model through integrated data analysis techniques. By optimizing hyperparameters within the predictive model, we achieved heightened accuracy, reaching a maximum precision of 0.9941 in forecasting pest outbreaks up to one year in advance.

Keywords: *Dendrolimus sibiricus*; genetic programming; pest outbreaks; random forest



Citation: Malashin, I.; Masich, I.; Tynchenko, V.; Nelyub, V.; Borodulin, A.; Gantimurov, A.; Shkaberina, G.; Rezova, N. Forecasting *Dendrolimus sibiricus* Outbreaks: Data Analysis and Genetic Programming-Based Predictive Modeling. *Forests* **2024**, *15*, 800. <https://doi.org/10.3390/f15050800>

Academic Editors: Stelian Alexandru Borz and Nopparat Kaakkurivaara

Received: 21 March 2024

Revised: 19 April 2024

Accepted: 29 April 2024

Published: 30 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Dendrolimus sibiricus, commonly known as the Siberian moth (SM), is a pest insect species that periodically undergoes outbreaks, defoliating vast areas of forests. It is one of the most hazardous pests to coniferous forests [1], not only in Siberia. Historical records dating back to the late 16th century from China [2] attest to its devastating impact on Southeast Asian forests. Various subspecies exist, with the Siberian subspecies being the most prevalent and rapidly spreading due to its significantly larger food base, making it the most harmful [3]. Adult moths are large and furry, with a wingspan ranging from 53 to 104 mm, while caterpillars grow rapidly, reaching lengths of up to 8 cm [4]. The moth has a biennial life cycle, overwintering twice as larvae in forest litter, and its activity is crepuscular. Due to its threat to coniferous forests, the spread of the SM raises concerns and requires preventive measures to avoid its invasion into new territories. An imbalance in coniferous forests caused by this pest leads to serious consequences, including changes in forest structure, drying out of forest stands and changes in the habitats of game animals [5]. SM poses a potential threat to European forests due to its westward spread and susceptibility to numerous conifer species. Recent sightings near Moscow indicate its proximity to European territories [6].

Research on methods of combating SM has long been a subject of scientific inquiry, shedding light on various strategies and approaches aimed at mitigating its impact on forest ecosystems. Florov’s method [7], proposed in 1947, suggests using deviations in moisture deficit exceeding 15% from the multi-year average for 2–3 consecutive years as a signal for potential SM outbreaks. Rozhkov recommends [8] assessing outbreak conditions based on the average number of eggs and caterpillars per tree. Nikitina’s empirical-statistical

model [9] aids in forecasting based on spatiotemporal population dynamics. Improved pheromone monitoring using traps with attractants enhances pest population tracking. Identified chemical compounds formulating the sexual attractant of the SM recommend its use in pheromone traps across its range. An optimal concentration for monitoring purposes has been determined, along with a methodology for sparse population pheromone monitoring. To mitigate its introduction and spread, proactive measures such as rigorous inspections of forest products and timber are necessary. Quantitative risk assessments and modeling techniques can aid in predicting its potential spread, informing targeted prevention efforts. A mechanistic grid-based model was developed in [10] to simulate the moth's potential spread in Europe, providing valuable insights for forest managers and policymakers to prioritize surveillance and mitigation efforts.

The detrimental effects of *Dendrolimus sibiricus* outbreaks on forest ecosystems, including reduced biodiversity, altered nutrient cycling, and increased susceptibility to other disturbances, have been well-documented in numerous research studies [11–13]. Additionally, the economic impact of these outbreaks [14,15], in terms of losses to timber production, forest regeneration efforts, and ecosystem services, has led to a heightened interest in understanding and predicting the occurrence of these events [16–18].

Paper [16] identified an altitudinal belt between 400 and 800 m above sea level as indicative of outbreak development. New parameters derived from remote sensing variables can forecast forest stand susceptibility to pest attacks up to 2–3 years in advance, simplifying monitoring efforts in inaccessible taiga forests. The research [17] focuses on enhancing the monitoring of SM outbreaks in the dark-coniferous taiga and aims to identify early detection methods and predict outbreaks. Through remote sensing and field surveys conducted in 2018–2019, the study examines preferred habitats of the SM based on terrain, forest type, and inventory characteristics. Work [18] addresses the urgent issue of large-scale destruction of taiga forests by Siberian silk moth outbreaks in mid-altitude mountains. It examines the influence of landscape factors on outbreak dynamics using Landsat-8 satellite imagery and field surveys.

Consequently, there is a pressing need for effective management strategies, underpinned by accurate forecasting techniques, to mitigate the adverse effects of *Dendrolimus sibiricus* outbreaks on forest ecosystems and regional economies. The aim of our research is to utilize data on forest composition and climate history for 15,000 taiga forest plots in the Krasnoyarsk Krai region.

- Our primary goal is to identify the optimal parameters for a classification model, employing machine learning techniques rooted in genetic programming.
- Specifically, we aim to forecast Siberian silk moth outbreaks one year in advance.
- The identification of these parameters is essential for precisely distinguishing between infected and uninfected forest plots.

2. Materials and Methods

2.1. Data Collection Methods and Analysis

This scientific study delves into the characterization of forest plots, with a keen focus on crucial parameters pivotal for predicting Siberian silk moth outbreaks. Among these parameters, soil moisture and mossiness stand out as fundamental indicators of habitat suitability for the moth's proliferation [19]. Additionally, the presence of past outbreaks serves as a benchmark for anticipating future occurrences. Age, height, diameter, and density of trees within the plots offer insights into the maturity and health of the forest [20–22], influencing the susceptibility to moth infestations. Furthermore, factors such as volume and area of selected forest plots provide quantitative measures of forest coverage [23], influencing the scale of potential outbreaks. Finally, slope exposure and steepness contribute [24] valuable information about the topographic features affecting microclimates within the forest, which play a significant role in shaping the moth's habitat preferences and population dynamics.

Figure 1 shows the distribution of forestry districts in the Kracnoyarsk region. The inset provides an overview of the entire area at a reduced scale.

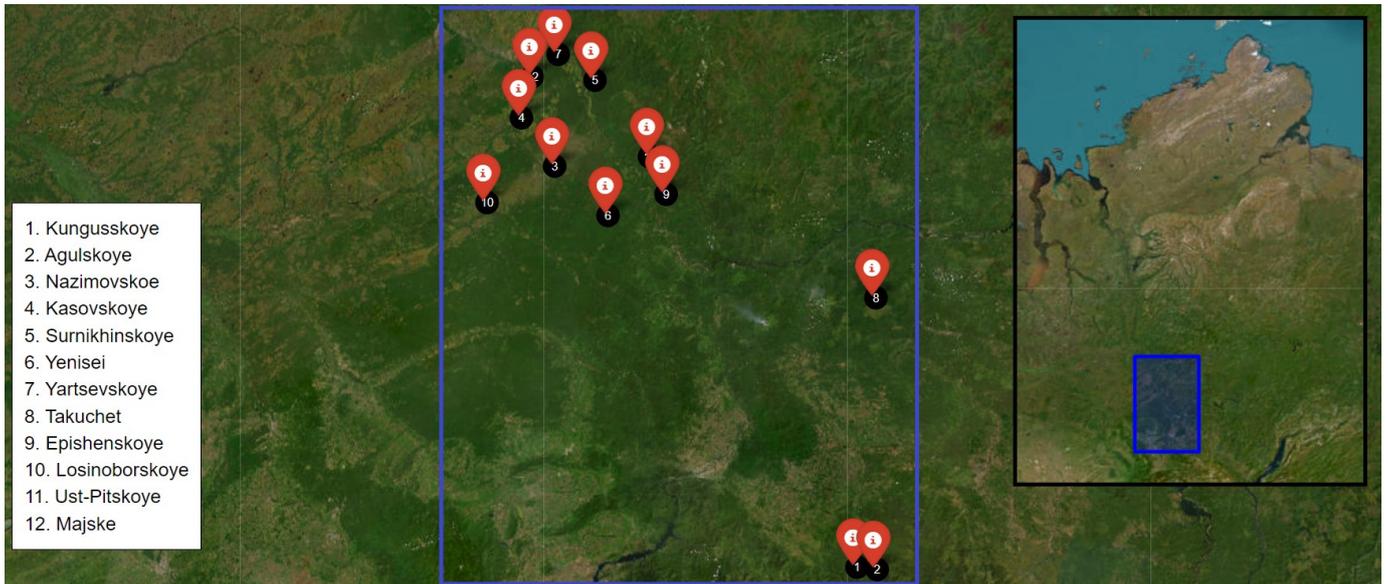


Figure 1. Forestry Districts on the Map.

Figure 2 illustrates the distribution of forest districts based on the number of moth outbreaks. Each bar represents a forestry district, and the height of the bar indicates the count of incidents within each district.

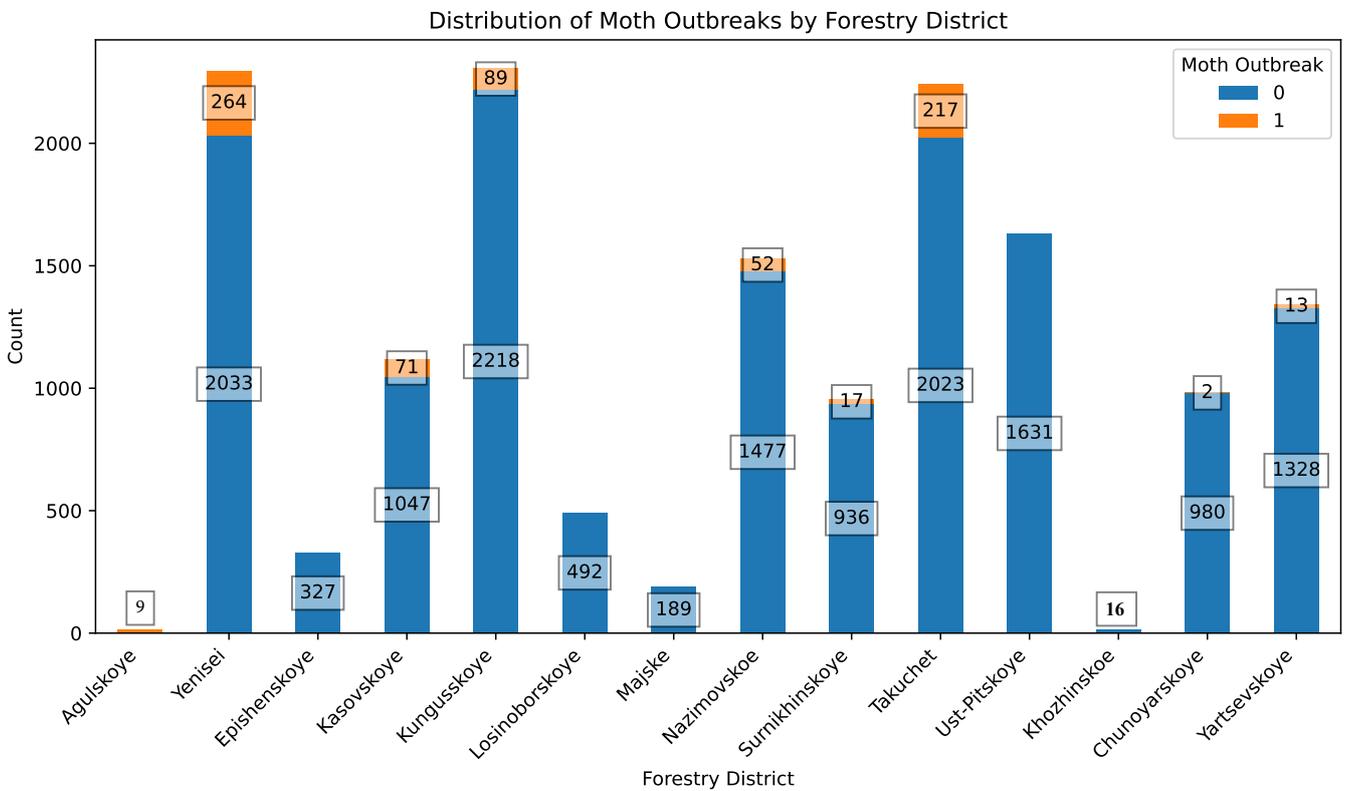


Figure 2. Distribution of Forest Districts by Number of Incidents.

Figure 3 depicts the distribution of forestry districts in the region as well as outbreaks of moth. Red dots indicate areas with outbreaks, while blue dots represent healthy forest areas.

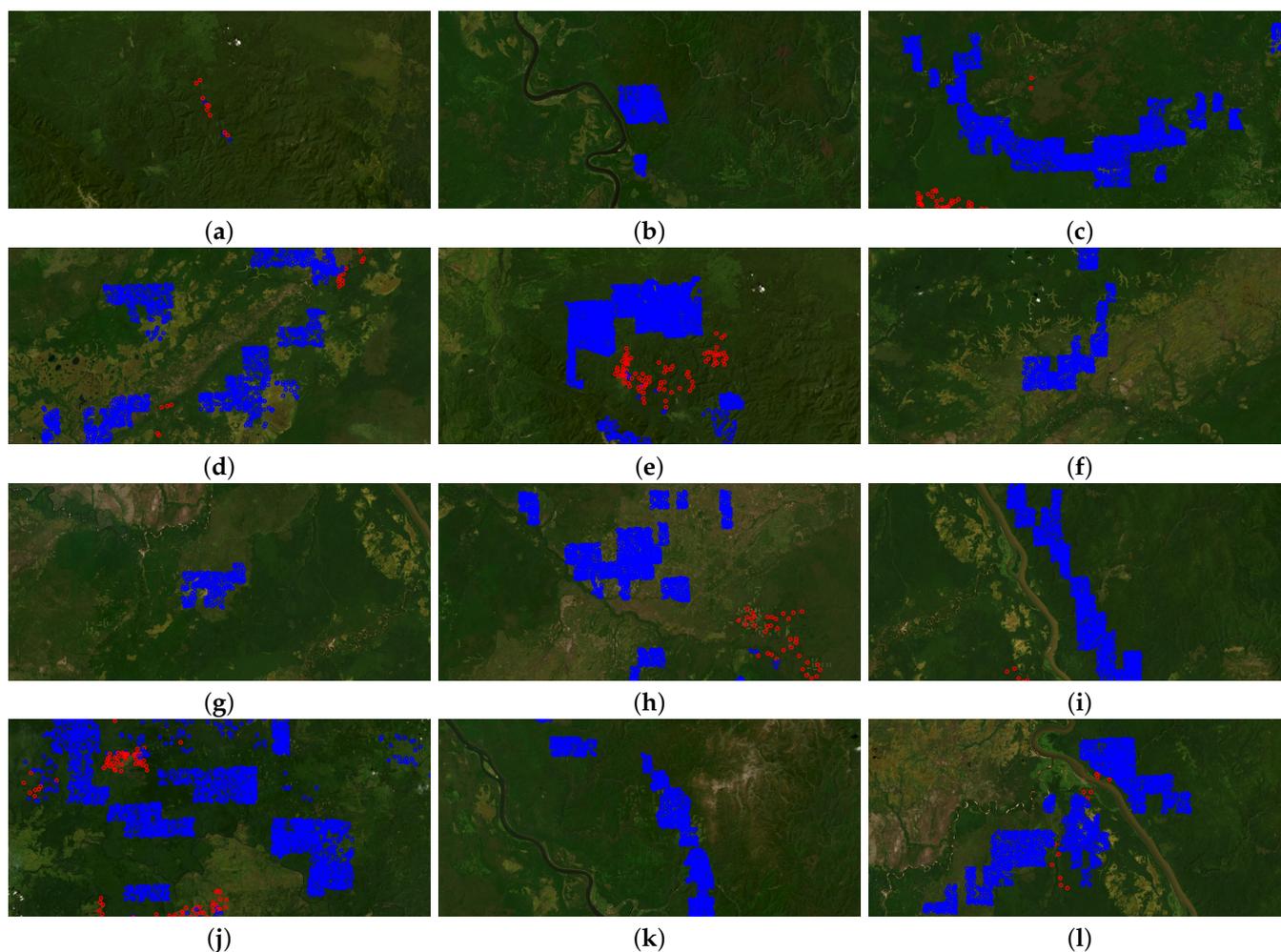


Figure 3. Distribution of *Dendrolimus Sibiricus* Outbreaks (red points) depending of forestry districts: (a) Agul'skoye (b) Epishenskoye (c) Yenisei (d) Kasovskoye (e) Kungusskoye (f) Losinoborskoye (g) Majske (h) Nazimovskoe (i) Surnikhinskoye (j) Takuchet (k) Ust-Pitskoye (l) Yartsevskoye.

The image depicts the distribution of forestry districts in the region as well as outbreaks of gypsy moth. Red dots indicate areas with gypsy moth outbreaks, while blue dots represent healthy forest areas.

Field surveys conducted by researchers entail on-site assessments aimed at obtaining data, including measurements of tree characteristics [25] such as age, height, and diameter. These measurements are typically taken using specialized tools like clinometers and diameter tapes. Soil moisture [26] and mossiness scores are also determined through direct observation and soil sampling during these surveys. Remote sensing techniques complement field surveys by providing satellite imagery [27] for large-scale data collection. These images are analyzed to assess forest parameters such as tree height, forest type, slope exposure and steepness.

Figure 4a illustrates correlation matrix of forest [28] characteristics which represents the relationships between different forest characteristics. Understanding these correlations is crucial for assessing the interdependencies among various environmental factors and their impact on forest health and ecosystem dynamics. This information aids in identifying key factors influencing forest ecosystems and informs management strategies for sustainable forest conservation and management practices.

Figure 4b illustrating the distribution of tree types across forest areas. Each whisker plot displays the distribution of tree types [29] (like Pine, Spruce, Larch, Birch, Aspen, Cedar, Fir, Willow) across different forest areas. The total count of tree types for each

area sums up to 10, reflecting the diversity and composition of tree species within the forest ecosystem. Variations in the distribution of tree species may indicate environmental conditions' peculiarities, as well as the impact of various factors such as climate, soil properties, and human activities.

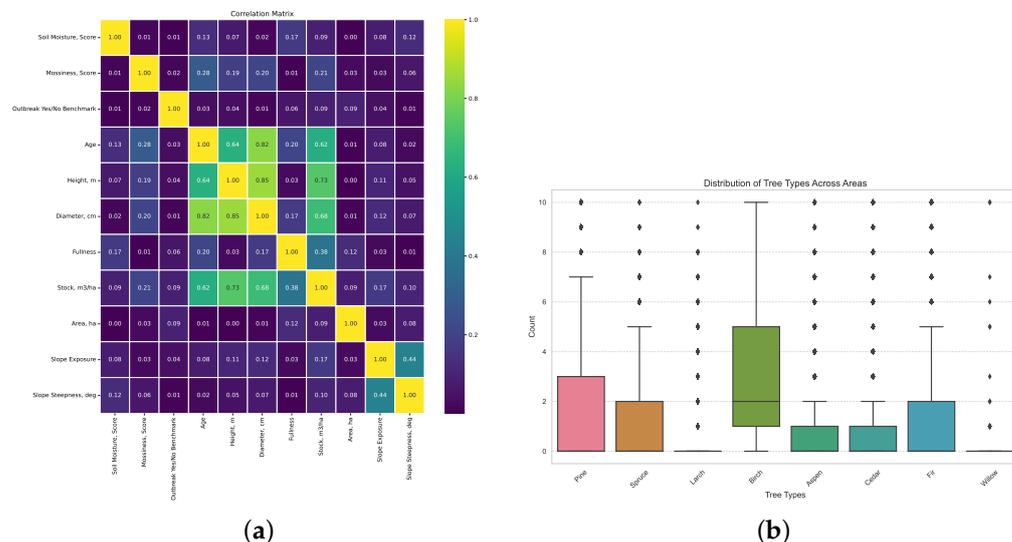


Figure 4. (a) Correlation Matrix of Forest Characteristics. (b) Distribution of Tree Types Across Forest Areas.

Climate data, obtained through the FLDAS Noah Land Surface Model [30], with specific attention directed towards the months of June and July. These months were chosen due to their significance as the summer period in Siberia, devoid of winter conditions, thus presenting optimal conditions for the proliferation of the Siberian silk moth. Key variables include soil temperature at depths of 0–10 cm below ground, maximum winter temperatures indicating thaw events, peak snow cover height, surface temperature recorded monthly from May to October, monthly evaporation rates from May to September, precipitation flux throughout the same period, and monthly soil moisture content at depths of 0–10 cm below ground.

A comparative analysis of climatic indicators for Siberian silk moth outbreaks in 1996, 2016, and 2018 reveals notable trends. The air temperature variation during the three years preceding each outbreak averaged four-six degrees Celsius lower for unaffected territories compared to those affected (Figure 5a–f). Additionally, the average evaporation level was 1.5–2 times higher for the infested area, reaching 3.5 mm/month compared to 2 mm/month for observations over four years preceding the outbreak (Figure 5g–i). Soil moisture variation over the seven years before the outbreak generally showed similar patterns for both affected and unaffected territories. However, there were localized peaks in soil moisture (Figure 5d,e) for April in 1996 and 2016 outbreaks for unaffected territories, while for affected areas, this value gradually declined to a minimum by July according to observations over seven years preceding the outbreaks. Possible explanations for these phenomena may include differences in microclimatic conditions, land use practices, and ecological factors influencing soil moisture retention and evaporation rates.

Based on the Figure 6g–l, it appears that the amount of precipitation does not seem to significantly influence the likelihood of an outbreak.

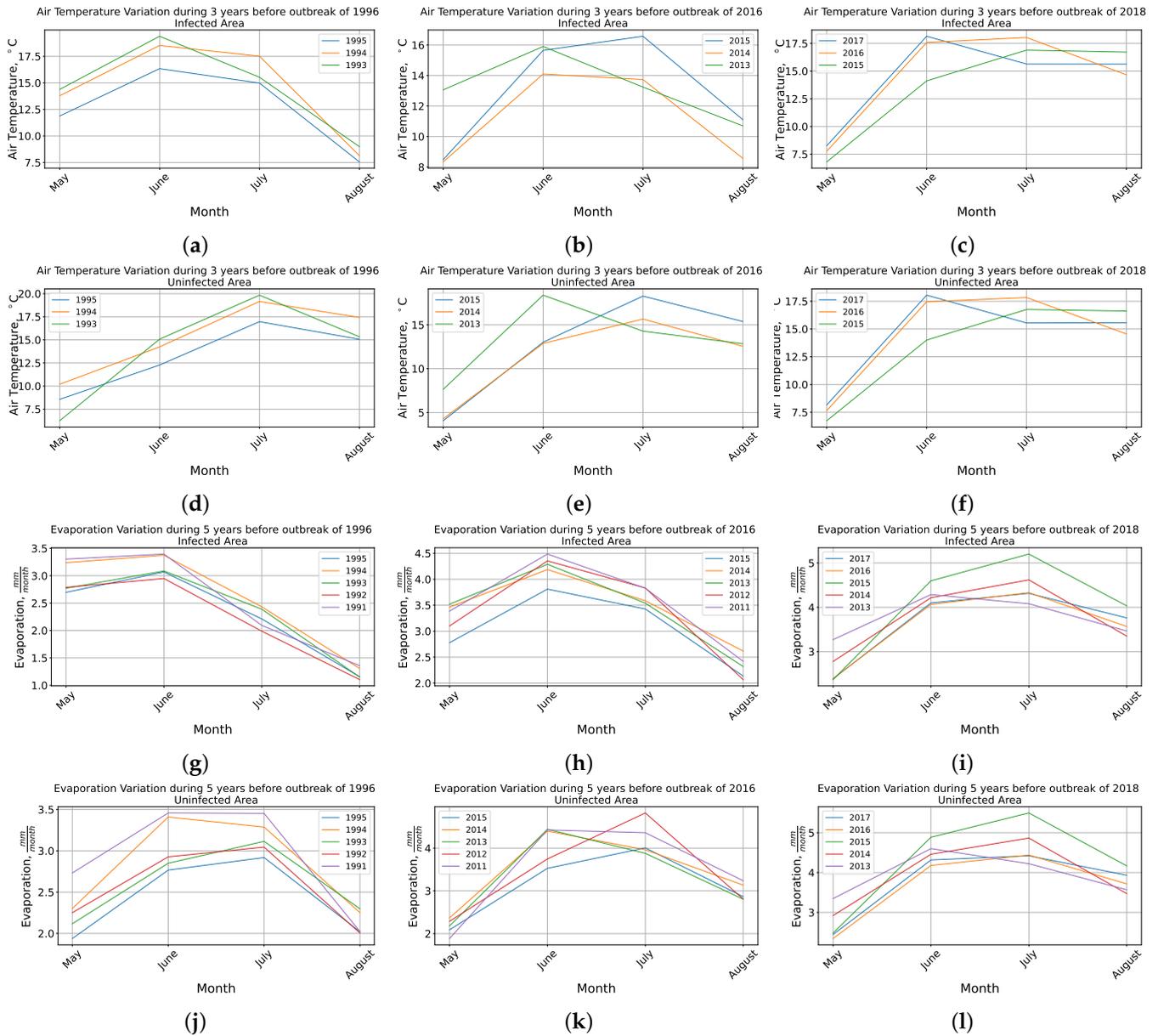


Figure 5. Temperature Variation during 3 years before outbreak of 1996, 2016, and 2018 for infected (a–c) and uninfected (d–f) areas from May to August; Evaporation Variation during 5 years before outbreak of 1996, 2016, and 2018 for infected (g–i) and uninfected (j–l) areas from May to August.

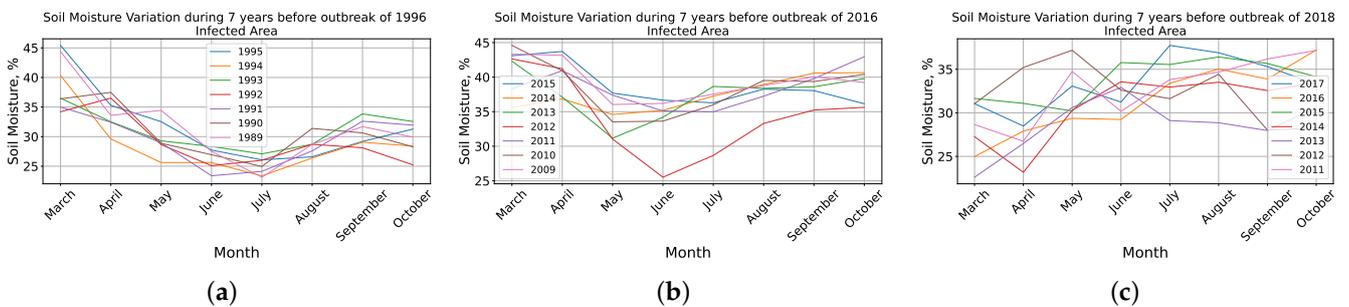


Figure 6. Cont.

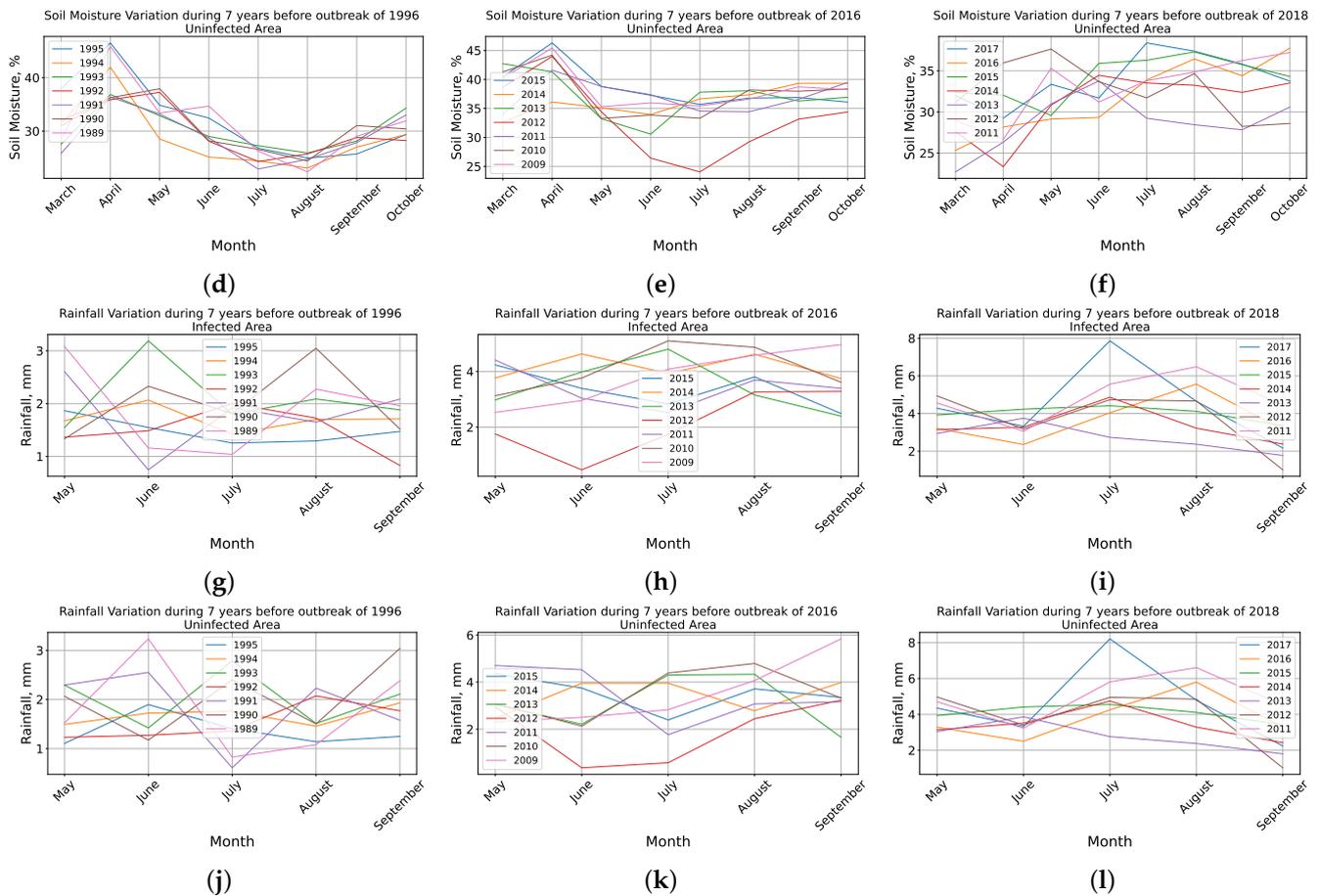


Figure 6. Soil Moisture during 7 years before outbreak of 1996, 2016, and 2018 for infected (a–c) and uninfected (d–f) areas from March to October. Rainfall Variation during 7 years before outbreak of 1996, 2016, and 2018 for infected (g–i) and uninfected (j–l) areas from May to September.

2.2. Application of Machine Learning Technique

This study explores the application of genetic programming (GP) techniques [31] to optimize the hyperparameters of a binary classifier, specifically the Random Forest [32] algorithm, for the prediction of Siberian silk moth outbreaks based on available data on forest composition and climatic indicators across 15,000 taiga forest plots in the Krasnoyarsk Krai region. The GP methodology aims to enhance the classifier’s performance by fine-tuning its parameters, thereby improving its predictive accuracy.

GP approach involves defining a fitness function [33] to evaluate the performance of each individual, which represents a set of hyperparameters for the Random Forest classifier. The algorithm’s hyperparameters, such as the number of estimators and maximum depth, are optimized iteratively using an evolutionary algorithm that mimics natural selection processes. We divided our dataset into training and testing subsets using a 80:20 ratio, with 80% of the data allocated for training and 20% for testing. This split ensured a sufficient amount of data for training while allowing robust evaluation of the model’s performance. The process continues through multiple generations, with the goal of identifying the optimal hyperparameter configuration that maximizes the classifier’s predictive accuracy.

The Genetic Programming (GP) approach was implemented using Python, with the code written utilizing the DEAP library [34]. Additionally, the statistical analysis was also conducted using Python. The process involves defining the evaluation function, creating a toolbox with functions for genetic operators, initializing the population, and running the evolutionary algorithm to optimize the classifier’s hyperparameters. The best-performing individual, representing the optimal hyperparameter configuration [35], is selected, and the classifier is trained and evaluated using these parameters. Additionally, the results of

each population iteration are stored in an Excel file for further analysis and comparison, as illustrated in Figure 7.

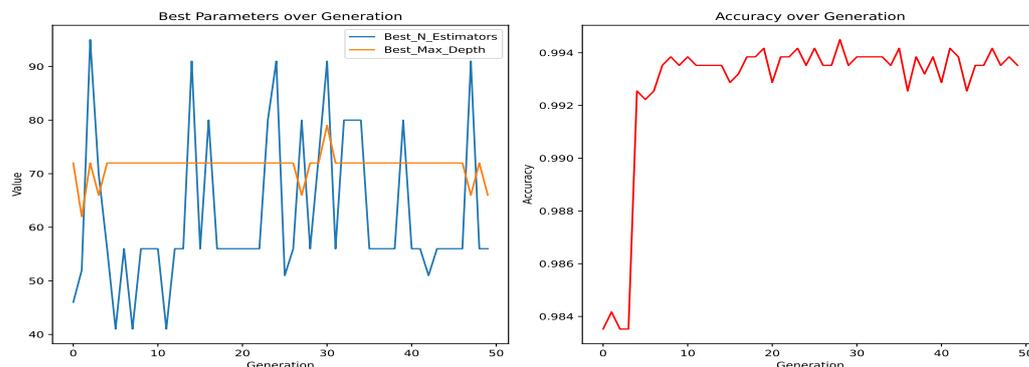


Figure 7. Evolution of Random Forest Hyperparameters and Accuracy over Generations.

Overall, this study showcases the efficacy of genetic programming techniques in optimizing the hyperparameters of a Random Forest classifier for predicting Siberian silk moth outbreaks. The optimized classifier presents potential for enhancing forest management strategies and alleviating the repercussions of pest outbreaks on taiga ecosystems. Additionally, the maximum accuracy achieved in predicting pest outbreaks one year in advance reached 0.9941 by the 8th generation.

Spatial cross-validation was conducted to evaluate the performance of the best random forest model identified through genetic algorithm optimization in predicting SM outbreaks in forests. The process involved dividing the dataset into five spatially distinct folds. The spatial cross-validation procedure was executed, resulting in accuracy scores for each fold. The obtained validation scores were as follows: 0.79, 0.84, 0.93, 0.99, and 0.96. These scores reflect the predictive capability of the model across different spatial regions, providing insights into its robustness and generalization ability.

3. Results

Due to the inherent imbalance [36] in the dataset, it was essential to monitor the precision parameter on the testing dataset to ensure the classifier's performance was accurately assessed. Precision measures the proportion of true positive predictions among all positive predictions made by the classifier, making it a crucial metric for evaluating classifier performance, especially in imbalanced datasets where the occurrence of positive cases is significantly lower than negative cases.

Through the optimization process, an optimal feature set was identified to maximize true positives (TP) while minimizing false negatives (FN). This feature set comprised a combination of variables related to forest composition and climatic indicators, carefully selected based on their predictive power and relevance to Siberian silk moth outbreak prediction. By maximizing the TP rate and minimizing FN rate, the classifier aims to improve its ability to correctly identify areas at risk of Siberian silk moth outbreaks, thereby enhancing forest management strategies and facilitating timely intervention measures. Confusion matrices illustrating various combinations of features are depicted in Figure 8.

Figure 8 dynamically illustrates how the confusion matrices evolve with varying dataset modifications. Across scenarios such as utilizing One-Hot Encoding for the 'forest type' feature, excluding this feature altogether, or incorporating temperature data without averaging, the matrices offer a visual narrative of classification accuracy and misclassifications. Notably, these insights extend to the validation dataset, providing a comprehensive exploration of model performance dynamics.

The evaluation of precision on the testing dataset allowed for the identification of the optimal feature set, which exhibited the highest precision value among all evaluated feature combinations. This optimal feature set not only maximized the classifier's ability

to correctly classify positive instances (outbreak occurrences) but also minimized the likelihood of false positive predictions, thus improving the overall reliability and accuracy of the classifier’s predictions.

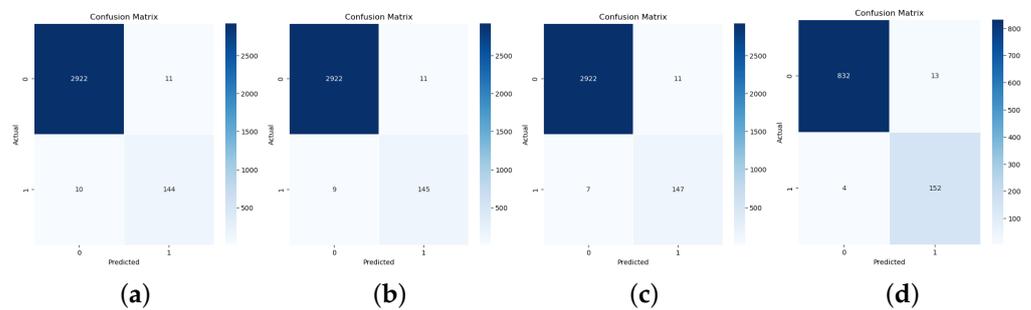


Figure 8. The confusion matrices for the test dataset of 3000 forest plots depicting the performance of the best estimator under various dataset modifications are illustrated as follows: (a) with the utilization of One-Hot Encoding (OHE) for the ‘forest type’ feature; (b) without the ‘forest type’ feature; (c) with temperature data not averaged and excluding the ‘forest type’ feature; and (d) the same as (c), but for the validation dataset consisting of 1000 forest plots.

Feature importance (Figure 9) in the context of forest dataset helps identify which attributes contribute most significantly to the model’s performance. By quantifying the influence of each feature on the model’s predictions, feature importance provides valuable insights into the underlying relationships between input variables and the target variable, such as the occurrence of forest pest outbreaks or ecosystem health. The most important features were identified as climatic variables, including rainfall for July, and soil moisture for August and September. They have the most significant impact on predicting forest conditions or potential pest outbreaks. This information is essential for prioritizing management strategies, allocating resources effectively, and informing decision-making processes aimed at preserving forest ecosystems and mitigating risks associated with environmental changes.

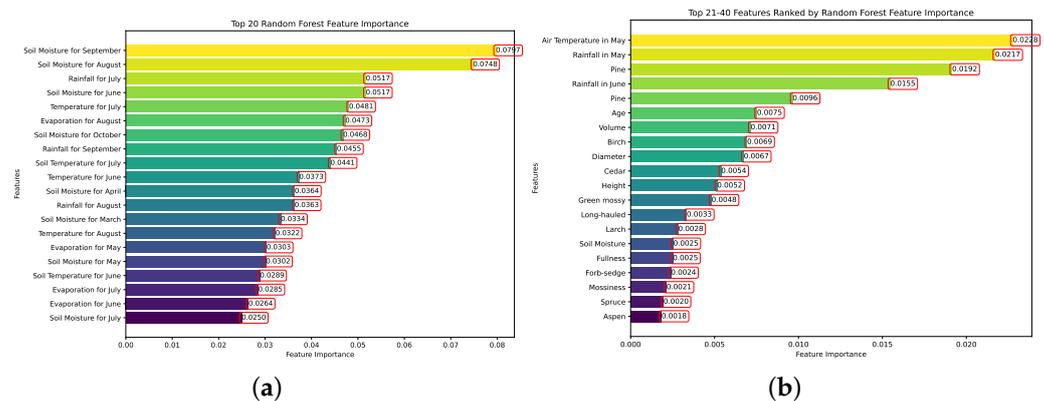


Figure 9. Feature Importance Ranked by Random Forest: (a) Top 20 Features; (b) Top 21–40 Features.

The findings highlight the significance of precision as a pivotal performance metric for classifiers dealing with imbalanced datasets, particularly within the domain of ecological modeling and the prediction of pest outbreaks. The delineation of the optimal feature set signifies a notable stride in fostering resilient and proficient predictive models for forecasting Siberian silk moth outbreaks, carrying implications for enhancing forest management strategies and safeguarding the health of taiga ecosystems.

Additionally, we systematically removed features from the dataset to investigate the impact of reduced feature sets on prediction accuracy. This process involved iteratively eliminating the least correlated features, prioritizing those with higher correlations with

the target variable. We conducted this analysis using the best random forest model identified through genetic algorithm optimization for predicting Siberian silk moth outbreaks in forests.

Throughout the experimentation, we observed a slight decrease in accuracy when reducing the number of features, accompanied by a significant decrease in recall. This trend was particularly pronounced when nearly all features were eliminated from the dataset. From these findings, we inferred that reducing the number of features for such imbalanced datasets may not yield high accuracy specifically in identifying the target variable.

It became evident that while feature importance plots highlighted certain variables as significant for prediction within the context of the model, attributing significance to these variables without further validation would be premature. The observed reduction in accuracy and recall could be attributed to the loss of crucial information necessary for effective classification.

4. Discussion

The study's integration of climatological indicators and forest attributes to predict Siberian silk moth outbreaks across 15,000 taiga parcels in the Krasnoyarsk Krai region marks a significant methodological advancement with far-reaching implications for ecological forecasting and pest management. By synthesizing a dataset comprising climatic variables and forest characteristics, the study adopts a sophisticated approach to ecological modeling, aiming to unravel the intricate interplay between environmental factors and pest dynamics.

Climatic indicators such as soil moisture, air temperature, and evaporation play a pivotal role in the occurrence of SM outbreaks in forests. Elevated soil moisture fosters favorable conditions for the reproduction and survival of moth larvae, thereby increasing the likelihood of pest population surges [37]. Air temperature influences the developmental pace and activity of the moth within the forest ecosystem; higher temperatures can expedite moth growth and proliferation [38]. Additionally, evaporation dynamics linked to atmospheric moisture levels, which in turn impact larval viability [39]. Thus, comprehensive understanding and monitoring of these climatic parameters are imperative for forecasting and managing SM outbreaks in forest ecosystems.

Predicting pest outbreaks using ML holds paramount importance in contemporary forestry management. Forest management depends on accurately predicting insect outbreaks like the case with mountain pine beetle which was highlighted by Ramazi et al. [40], often targeting future occurrences five years ahead. Their study utilizes machine learning algorithms to forecast mountain pine beetle outbreaks across different timeframes, offering insights crucial for effective forest and pest management planning.

Modeling and simulating forest land cover changes due to epidemic insect outbreaks, like the mountain pine beetle (MPB), are crucial for effective forest management strategies. The study [41] proposes an integrative approach utilizing supervised machine learning techniques to simulate the spatiotemporal dynamics of MPB infestation over lodgepole pine forests in British Columbia, Canada. By applying generalized linear regression (GLM) and random forest (RF) algorithms to predict MPB infestation, they observed that RF algorithms outperformed GLM, with simulations for 2020 suggesting a slower rate of spread in future MPB infestations in the province.

Bark beetles, like *Dendroctonus frontalis* Zimmermann, pose significant threats to pine trees in the United States and beyond, resulting in substantial economic and ecological damages. To enhance outbreak prediction models, the study [42] integrated spatial-temporal dynamics, climate variables, terrain attributes, and vegetation indices using extreme gradient boosting. Their models accurately predicted outbreak probability and magnitude, highlighting areas at high risk for damage. This approach, incorporating climatic variables, offers valuable insights into future pest population dynamics and facilitates proactive management strategies to mitigate risks associated with bark beetle outbreaks.

Examples in scientific literature highlight the analysis of climatic trends, such as a study [43] analyzing climate trends near Lake Superior's western end from 1984 to 2013 using weather station data. The results revealed a regional warming trend, with cooler springs and warmer autumns, potentially impacting forest phenology and ecosystem dynamics in this ecologically vital region.

The predictive modeling framework employed in this study is a sophisticated synthesis of machine learning methodologies, specifically tailored to address the complex ecological dynamics inherent in pest outbreak prediction. For example, in work [44] suggests model which incorporates relevant water quality indicators, computes trophic scores using ML techniques, and implements a new classification scheme. Evaluation across diverse waterbodies in Ireland demonstrates its effectiveness compared to existing systems.

Through systematic hyperparameter optimization within the Random Forest classifier, our model demonstrates a heightened capacity to discern subtle patterns within the data and extrapolate robust predictive insights. By elucidating the relationships between climatological variables like rainfall, temperature, and soil moisture, and forest attributes such as age, density, and composition, the model encapsulates the multifaceted nature of ecosystem dynamics, thereby enhancing its predictive efficacy. For instance, the influence of spatial autocorrelation on the performance estimation of machine learning algorithms in ecological modeling was outlined in [45]. Comparing various methods, including random forest and logistic regression, authors find that spatial cross-validation yields more accurate estimates. Prioritizing spatial hyperparameter tuning ensures consistency with spatial performance estimation, mitigating the risk of overoptimistic predictions that could misguide ecological decision-making.

Leveraging GP approach embedded within the Random Forest classifier, the model undergoes iterative refinement, dynamically adjusting hyperparameters to maximize predictive performance. This iterative optimization process enables the model to adapt and evolve, iteratively fine-tuning parameters [46] such as the number of estimators and maximum depth, thereby enhancing its capacity to discern intricate patterns within the data. This theme is also highlighted in scientific literature. For example, study [47] focuses on the M3GP algorithm, a variant of GP, which facilitates feature construction by evolving hyperfeatures from original satellite image data. By applying M3GP to diverse satellite datasets from different countries, we enhance land cover classification accuracy. Integrating the evolved hyperfeatures into reference datasets notably boosts the performance of decision trees, random forests, and XGBoost algorithms in multiclass classifications.

By harnessing the power of GP-based hyperparameter optimization [48], the model transcends conventional static parameter settings, dynamically adapting to the nuances of the dataset and underlying ecological dynamics. The GA framework operates on the principle of natural selection, iteratively refining parameter configurations based on their efficacy in improving model performance metrics. Through successive generations of parameter optimization, the model converges towards an optimal configuration, characterized by heightened predictive accuracy and robustness.

Furthermore, the predictive framework's applicability extends beyond mere forecast accuracy, offering invaluable insights into the underlying mechanisms driving pest outbreaks and their spatial-temporal dynamics. By discerning the relative importance of different features within the predictive model—such as climatological variables versus forest attributes—the study sheds light on the key determinants shaping pest susceptibility and outbreak propensity across diverse ecological landscapes. This holistic understanding not only facilitates early detection and proactive management of pest outbreaks but also fosters a deeper appreciation of the intricate ecological processes underpinning forest ecosystems.

GA-based predictive modeling approach delineated in this study represents a paradigm shift in ecological forecasting, transcending traditional boundaries by integrating multi-dimensional datasets and cutting-edge machine learning techniques. By harnessing the predictive power of climatological indicators and forest attributes, this methodology holds immense promise for informing evidence-based pest management strategies, fostering

resilience in forest ecosystems, and safeguarding biodiversity in the face of mounting environmental pressures.

5. Conclusions

In conclusion, forecasting *Dendrolimus sibiricus* outbreaks is crucial for mitigating their adverse effects on forest ecosystems and regional economies. Practical recommendations are provided for improving monitoring and forecasting efforts, emphasizing the importance of collaborative research and innovative technologies. We propose several avenues for future research in this area:

1. **Integration of Additional Variables:** Explore the inclusion of supplementary environmental variables beyond those considered in the current model, such as soil properties [49] (e.g., pH, nutrient levels) and landscape characteristics [50] (e.g., topography, land use/land cover), to capture more comprehensive ecological dynamics influencing pest outbreaks.
2. **Temporal Dynamics Analysis:** Investigate the temporal dynamics of *Dendrolimus sibiricus* populations and their interaction with climatic variables over longer time scales in other regions [51,52]. Analyze historical data to identify trends and patterns in outbreak occurrences, considering factors like seasonal variability, interannual fluctuations, and long-term climate change trends.
3. **Model Refinement and Validation:** Refine the predictive model by incorporating advanced machine learning techniques or ensemble methods [53] to improve accuracy and robustness. Validate the model's performance using independent datasets or through cross-validation techniques to ensure its reliability across different spatial and temporal contexts.
4. **Spatially Explicit Modeling:** Develop spatially explicit models [54] to account for spatial autocorrelation and heterogeneity in pest distribution patterns. Utilize geospatial analysis techniques and remote sensing data to delineate spatial risk zones and identify hotspots of pest activity within the study area.
5. **Ecological Drivers Identification:** Conduct in-depth analyses to identify the key ecological drivers influencing *Dendrolimus sibiricus* outbreaks, including interactions with host plant species [55], natural enemies, and abiotic factors. Investigate how changes in forest composition, structure, and management practices may affect pest population dynamics and outbreak severity.
6. **Management Strategies Evaluation:** Evaluate the effectiveness of different pest management strategies [56], such as biological control, chemical intervention, and silvicultural practices, in mitigating *Dendrolimus sibiricus* outbreaks. Assess the ecological and socioeconomic impacts of these strategies to inform sustainable forest management decisions.
7. **Climate Change Adaptation:** Anticipate the potential effects of climate change on *Dendrolimus sibiricus* outbreaks and develop adaptive management strategies to mitigate associated risks. Investigate how projected changes in temperature [57], precipitation, and extreme weather events may alter pest phenology, distribution, and abundance in the future.
8. **Interdisciplinary Collaboration:** Foster interdisciplinary collaboration [58] between ecologists, climatologists, entomologists, remote sensing experts, and decision-makers to integrate diverse expertise and perspectives into pest management research. Promote knowledge exchange and stakeholder engagement to facilitate the translation of scientific findings into actionable management strategies.

Author Contributions: Conceptualization, I.M. (Ivan Malashin); Data curation, I.M. (Igor Masich), V.N., A.B., N.R. and G.S.; Formal analysis, I.M. (Igor Masich), V.T., V.N. and N.R.; Funding acquisition, V.N., A.B. and A.G.; Investigation, I.M. (Igor Masich), V.T. and A.B.; Methodology, I.M. (Igor Masich); Project administration, V.T., V.N., A.B. and A.G.; Resources, A.G. and N.R.; Software, I.M. (Ivan Malashin), I.M. (Igor Masich), V.T., V.N., A.G. and G.S.; Supervision, I.M. (Igor Masich), V.N., A.B., A.G., N.R. and G.S.; Validation, I.M. (Ivan Malashin), I.M. (Igor Masich), A.B., N.R. and G.S.; Visualization, I.M. (Ivan Malashin) and V.T.; Writing—original draft, I.M. (Ivan Malashin), I.M. (Igor Masich) and V.N.; Writing—review & editing, V.T. and G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used in this study is available in the repository [Sibiricus](#) (accessed on 21 March 2024).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. EFSA Panel on Plant Health (PLH); Jeger, M.; Bragard, C.; Caffier, D.; Candresse, T.; Chatzivassiliou, E.; Dehnen-Schmutz, K.; Gilioli, G.; Jaques Miret, J.A.; MacLeod, A.; et al. Pest categorisation of *Dendrolimus sibiricus*. *EFSA J.* **2018**, *16*, e05301.
2. Sliwa, E. Occurrence of *Dendrolimus pini* and its control in the forests of Poland. *Sylvan* **1966**, *110*, 51–57.
3. Skrzecz, I.; Ślusarski, S.; Tkaczyk, M. Integration of science and practice for *Dendrolimus pini* (L.) management—A review with special reference to Central Europe. *For. Ecol. Manag.* **2020**, *455*, 117697. [[CrossRef](#)]
4. Vinokurov, N.N.; Isaev, A.P. The Siberian moth in Yakutia. *Sci. Technol. Yakutia* **2002**, *2*, 53–56.
5. Koltunov, E.; Erdakov, L. Cyclicity features of the multi-year dynamics of outbreaks of mass reproduction of different geographical populations of the Siberian moth (*Dendrolimus superans sibiricus tschetv*) in Siberia. In *Modern Problems of Science and Education*; Moscow State University of Psychology and Education (MSUPE): Moscow, Russia, 2013; p. 700.
6. Gninenko, Y.I.; Orlinkii, A. *Dendrolimus sibiricus* in the coniferous forests of European Russia at the beginning of the twenty-first century. *EPPO Bull.* **2002**, *32*, 481–483. [[CrossRef](#)]
7. Florov, D. *Forest Insect Pests*; OGIZ, Irkutsk Regional Publishing House: Irkutsk, Russia, 1948. (In Russian)
8. Rozhkov, A.S. *Siberian Moth: Systematic Position, Phylogeny, Distribution, Economic Significance, Structure, and Way of Life*; AS USSR Press: Moscow, Russia, 1963. (In Russian)
9. Nikitina, Y. Development of a point model of the Siberian moth population. *Interexpo-Geo-Sib.* **2006**, *3*, 156–161.
10. Flø, D.; Rafoss, T.; Wendell, M.; Sundheim, L. The Siberian moth (*Dendrolimus sibiricus*), a pest risk assessment for Norway. *For. Ecosyst.* **2020**, *7*, 48. [[CrossRef](#)]
11. Pavlov, I.; Litovka, Y.A.; Golubev, D.; Astapenko, S.; Chromogin, P. New outbreak of *Dendrolimus sibiricus tschetv.* in Siberia (2012–2017): Monitoring, modeling and biological control. *Contemp. Probl. Ecol.* **2018**, *11*, 406–419. [[CrossRef](#)]
12. Kirichenko, N.; Flament, J.; Baranchikov, Y.; Grégoire, J.C. Native and exotic coniferous species in Europe—possible host plants for the potentially invasive Siberian moth, *Dendrolimus sibiricus* 1 Tschv. (Lepidoptera, Lasiocampidae). *EPPO Bull.* **2008**, *38*, 259–263. [[CrossRef](#)]
13. Sul'tson, S.; Mikhaylov, P.; Kulakov, S.; Goroshko, A. Opportunities for assessing the risk of an outbreak of Siberian silkworm (*Dendrolimus superans sibiricus* Tschetv.) in taiga forests. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *548*, 052051. [[CrossRef](#)]
14. Buck, J.H. Effects of Natural Disturbances Caused by the Siberian Moth, *Dendrolimus Superans Sibiricus* (Tschetverikov), and Fire on the Dynamics of Boreal Forests in Krasnoyarsk Krai, Russia. Ph.D. Thesis, School for Environment and Sustainability, Ann Arbor, MI, USA, 2008.
15. Demidko, D.A.; Trefilova, O.V.; Kulakov, S.S.; Mikhaylov, P.V. Pine Looper *Bupalus piniaria* (L.) Outbreaks Reconstruction: A Case Study for Southern Siberia. *Insects* **2021**, *12*, 90. [[CrossRef](#)]
16. Soukhovolsky, V.; Kovalev, A.; Goroshko, A.A.; Ivanova, Y.; Tarasova, O. Monitoring and Prediction of Siberian Silk Moth *Dendrolimus sibiricus* Tschetv. (Lepidoptera: Lasiocampidae) Outbreaks Using Remote Sensing Techniques. *Insects* **2023**, *14*, 955. [[CrossRef](#)]
17. Sultson, S.M.; Goroshko, A.A.; Mikhaylov, P.V.; Demidko, D.A.; Ponomarev, E.; Verkhovets, S.V. Improving the Monitoring System Towards Early Detection and Prediction of the Siberian Moth Outbreaks in Eastern Siberia. In Proceedings of the 1st International Electronic Conference on Entomology, Online, 1–15 July 2021; pp. 1–15.
18. Sultson, S.M.; Goroshko, A.A.; Verkhovets, S.V.; Mikhaylov, P.V.; Ivanov, V.A.; Demidko, D.A.; Kulakov, S.S. Orographic factors as a predictor of the spread of the Siberian silk moth outbreak in the mountainous southern taiga Forests of Siberia. *Land* **2021**, *10*, 115. [[CrossRef](#)]
19. Bruijnzeel, L.; Kappelle, M.; Mulligan, M.; Scatena, F.N. *Tropical Montane Cloud Forests: State of Knowledge and Sustainability Perspectives in a Changing World*; Cambridge University Press: Cambridge, UK, 2010; pp. 691–740.

20. Roberts, A.J.; Crowley, L.M.; Sadler, J.P.; Nguyen, T.T.; Gardner, A.M.; Hayward, S.A.; Metcalfe, D.B. Effects of elevated atmospheric CO₂ concentration on insect herbivory and nutrient fluxes in a mature temperate Forest. *Forests* **2022**, *13*, 998. [CrossRef]
21. Giupponi, L.; Leoni, V.; Pedrali, D.; Giorgi, A. Restoration of Vegetation Greenness and Possible Changes in Mature Forest Communities in Two Forests Damaged by the Vaia Storm in Northern Italy. *Plants* **2023**, *12*, 1369. [CrossRef]
22. Harris, R.C.; Kennedy, L.M.; Pingel, T.J.; Thomas, V.A. Assessment of canopy health with drone-based orthoimagery in a Southern Appalachian red spruce forest. *Remote Sens.* **2022**, *14*, 1341. [CrossRef]
23. Ganz, S.; Adler, P.; Kändler, G. Forest cover mapping based on a combination of aerial images and Sentinel-2 satellite data compared to National Forest Inventory data. *Forests* **2020**, *11*, 1322. [CrossRef]
24. Jourgholami, M.; Karami, S.; Tavankar, F.; Lo Monaco, A.; Picchio, R. Effects of slope gradient on runoff and sediment yield on machine-induced compacted soil in temperate forests. *Forests* **2020**, *12*, 49. [CrossRef]
25. Balenović, I.; Jazbec, A.; Marjanović, H.; Paladinić, E.; Vuletić, D. Modeling tree characteristics of individual black pine (*Pinus nigra* Arn.) trees for use in remote sensing-based inventory. *Forests* **2015**, *6*, 492–509. [CrossRef]
26. Pastor, J.; Post, W. Influence of climate, soil moisture, and succession on forest carbon and nitrogen cycles. *Biogeochemistry* **1986**, *2*, 3–27. [CrossRef]
27. USGS EarthExplorer. Available online: <https://earthexplorer.usgs.gov/> (accessed on 20 March 2024).
28. Guan, B.T. Effects of correlation among parameters on prediction quality of a process-based forest growth model. *For. Sci.* **2000**, *46*, 269–276. [CrossRef]
29. Chen, W. Tree size distribution functions of four boreal forest types for biomass mapping. *For. Sci.* **2004**, *50*, 436–449. [CrossRef]
30. Jacob, J.; Slinksi, K. *FLDAS Noah Land Surface Model L4 Global Monthly 0.1 × 0.1 Degree (GDAS and CHIRPS-PRELIM)*; Goddard Earth Sciences Data and Information Services Center (GES DISC): Greenbelt, MD, USA, 2021. Available online: https://disc.gsfc.nasa.gov/datasets/FLDAS_NOAH01_CP_GL_M_001/summary (accessed on 1 March 2024).
31. Ahvanooy, M.T.; Li, Q.; Wu, M.; Wang, S. A survey of genetic programming and its applications. *KSII Trans. Internet Inf. Syst. (TIIS)* **2019**, *13*, 1765–1794.
32. Rigatti, S.J. Random forest. *J. Insur. Med.* **2017**, *47*, 31–39. [CrossRef]
33. Baresel, A.; Sthamer, H.; Schmidt, M. Fitness function design to improve evolutionary structural testing. In Proceedings of the 4th Annual Conference on Genetic and Evolutionary Computation, New York, NY, USA, 9–13 July 2002; pp. 1329–1336.
34. Kim, J.; Yoo, S. Software review: Deap (distributed evolutionary algorithm in python) library. *Genet. Program. Evolvable Mach.* **2019**, *20*, 139–142. [CrossRef]
35. Yang, L.; Shami, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* **2020**, *415*, 295–316. [CrossRef]
36. Ramyachitra, D.; Manikandan, P. Imbalanced dataset classification and solutions: A review. *Int. J. Comput. Bus. Res. (IJCBR)* **2014**, *5*, 1–29.
37. Saravesi, K.; Aikio, S.; Wäli, P.R.; Ruotsalainen, A.L.; Kaukonen, M.; Huusko, K.; Suokas, M.; Brown, S.P.; Jumpponen, A.; Tuomi, J.; et al. Moth outbreaks alter root-associated fungal communities in subarctic mountain birch forests. *Microb. Ecol.* **2015**, *69*, 788–797. [CrossRef]
38. Casey, T.M. Flight energetics and heat exchange of gypsy moths in relation to air temperature. *J. Exp. Biol.* **1980**, *88*, 133–146. [CrossRef]
39. Judd, G.J.; Gardiner, M.G.; Thomson, D. Control of codling moth in organically-managed apple orchards by combining pheromone-mediated mating disruption, post-harvest fruit removal and tree banding. *Entomol. Exp. Appl.* **1997**, *83*, 137–146. [CrossRef]
40. Ramazi, P.; Kunegel-Lion, M.; Greiner, R.; Lewis, M.A. Predicting insect outbreaks using machine learning: A mountain pine beetle case study. *Ecol. Evol.* **2021**, *11*, 13014–13028. [CrossRef] [PubMed]
41. Harati, S.; Perez, L.; Molowny-Horas, R. Integrating neighborhood effect and supervised machine learning techniques to model and simulate forest insect outbreaks in British Columbia, Canada. *Forests* **2020**, *11*, 1215. [CrossRef]
42. Munro, H.L.; Montes, C.R.; Gandhi, K.J. A new approach to evaluate the risk of bark beetle outbreaks using multi-step machine learning methods. *For. Ecol. Manag.* **2022**, *520*, 120347. [CrossRef]
43. Garcia, M.; Townsend, P.A. Recent climatological trends and potential influences on forest phenology around western Lake Superior, USA. *J. Geophys. Res. Atmos.* **2016**, *121*, 13–364. [CrossRef]
44. Uddin, M.G.; Nash, S.; Rahman, A.; Dabrowski, T.; Olbert, A.I. Data-driven modelling for assessing trophic status in marine ecosystems using machine learning approaches. *Environ. Res.* **2024**, *242*, 117755. [CrossRef]
45. Schratz, P.; Muenchow, J.; Iturrutxa, E.; Richter, J.; Brenning, A. Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecol. Model.* **2019**, *406*, 109–120. [CrossRef]
46. Fu, Z.; Yang, H.; So, A.M.C.; Lam, W.; Bing, L.; Collier, N. On the effectiveness of parameter-efficient fine-tuning. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 12799–12807.
47. Batista, J.E.; Cabral, A.I.; Vasconcelos, M.J.; Vanneschi, L.; Silva, S. Improving land cover classification using genetic programming for feature construction. *Remote Sens.* **2021**, *13*, 1623. [CrossRef]
48. Bergstra, J.; Bardenet, R.; Bengio, Y.; Kégl, B. Algorithms for hyper-parameter optimization. In Proceedings of the NIPS'11: 24th International Conference on Neural Information Processing Systems, Granada, Spain, 12–15 December 2011.

49. Turczański, K.; Rutkowski, P.; Dyderski, M.K.; Wrońska-Pilarek, D.; Nowiński, M. Soil pH and organic matter content affects European ash (*Fraxinus excelsior* L.) crown defoliation and its impact on understory vegetation. *Forests* **2019**, *11*, 22. [CrossRef]
50. Kefalas, G.; Lorilla, R.S.; Xofis, P.; Poirazidis, K.; Eliades, N.G.H. Landscape Characteristics in Relation to Ecosystem Services Supply: The Case of a Mediterranean Forest on the Island of Cyprus. *Forests* **2023**, *14*, 1286. [CrossRef]
51. Mikkola, K.; Ståhls, G. Morphological and molecular taxonomy of *Dendrolimus sibiricus* Chetverikov stat. rev. and allied lappet moths (Lepidoptera: Lasiocampidae), with description of a new species. *Entomol. Fenn.* **2008**, *19*, 65–85. [CrossRef]
52. Lukin, A. New Data on the Distribution and Abundance of *Dendrolimus sibiricus* (Tshetverikov, 1908) (Lepidoptera: Lasiocampidae) in the Komi Republic. 2021. Available online: <https://assets.researchsquare.com/files/rs-900432/v1/85264be7-71fb-4004-9b0d-d4f968aec7b1.pdf?c=1637245744> (accessed on 20 March 2024).
53. Dietterich, T.G. Ensemble methods in machine learning. In Proceedings of the International Workshop on Multiple Classifier Systems, Cagliari, Italy, 21–23 June 2000; Springer: Berlin/Heidelberg, Germany, 2000; pp. 1–15.
54. DeAngelis, D.L.; Yurek, S. Spatially explicit modeling in ecology: A review. *Ecosystems* **2017**, *20*, 284–300. [CrossRef]
55. Canelles, Q.; Aquilué, N.; James, P.M.; Lawler, J.; Brotons, L. Global review on interactions between insect pests and other forest disturbances. *Landsc. Ecol.* **2021**, *36*, 945–972. [CrossRef]
56. Moricca, S.; Panzavolta, T. Recent Advances in the Monitoring, Assessment and Management of Forest Pathogens and Pests. *Forests* **2021**, *12*, 1623. [CrossRef]
57. Nunes, L.J.; Meireles, C.I.; Gomes, C.J.P.; Ribeiro, N.M.A. The impact of climate change on forest development: A sustainable approach to management models applied to Mediterranean-type climate regions. *Plants* **2021**, *11*, 69. [CrossRef] [PubMed]
58. Brandstädter, S.; Sonntag, K. Interdisciplinary collaboration: How to foster the dialogue across disciplinary borders? In *Advances in Ergonomic Design of Systems, Products and Processes, Proceedings of the Annual Meeting of GfA 2015*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 395–409.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.