agriculture

Review

Review of Methods and Models for Potato Yield Prediction

Magdalena Piekutowska [1] and Gniewko Niedbała [2,*]

[1] Department of Botany and Nature Protection, Institute of Biology, Pomeranian University in Słupsk, 22b Arciszewskiego St., 76-200 Słupsk, Poland; magdalena.piekutowska@upsl.edu.pl
[2] Department of Biosystems Engineering, Faculty of Environmental and Mechanical Engineering, Poznań University of Life Sciences, Wojska Polskiego 50, 60-627 Poznań, Poland
* Correspondence: gniewko.niedbala@up.poznan.pl

Abstract: This article provides a comprehensive overview of the development and application of statistical methods, process-based models, machine learning, and deep learning techniques in potato yield forecasting. It emphasizes the importance of integrating diverse data sources, including meteorological, phenotypic, and remote sensing data. Advances in computer technology have enabled the creation of more sophisticated models, such as mixed, geostatistical, and Bayesian models. Special attention is given to deep learning techniques, particularly convolutional neural networks, which significantly enhance forecast accuracy by analyzing complex data patterns. The article also discusses the effectiveness of other algorithms, such as Random Forest and Support Vector Machines, in capturing nonlinear relationships affecting yields. According to standards adopted in agricultural research, the Mean Absolute Percentage Error (MAPE) in the implementation of prediction issues should generally not exceed 15%. Contemporary research indicates that, through the use of advanced and accurate algorithms, the value of this error can reach levels of even less than 10 per cent, significantly increasing the efficiency of yield forecasting. Key challenges in the field include climatic variability and difficulties in obtaining accurate data on soil properties and agronomic practices. Despite these challenges, technological advancements present new opportunities for more accurate forecasting. Future research should focus on leveraging Internet of Things (IoT) technology for real-time data collection and analyzing the impact of biological variables on yield. An interdisciplinary approach, integrating insights from ecology and meteorology, is recommended to develop innovative predictive models. The exploration of machine learning methods has the potential to advance knowledge in potato yield forecasting and support sustainable agricultural practices.

Keywords: potato; yield prediction; statistical models; process-based models; machine learning; deep learning; neural networks; model complexity; predictive accuracy

Academic Editor: Dengpan Xiao

Received: 23 December 2024
Revised: 5 February 2025
Accepted: 8 February 2025
Published: 9 February 2025

Citation: Piekutowska, M.; Niedbała, G. Review of Methods and Models for Potato Yield Prediction. *Agriculture* **2025**, *15*, 367. https://doi.org/10.3390/agriculture15040367

1. Introduction

Potato (*Solanum tuberosum* L.) is not only one of the most important sources of food, but also a key ingredient for the economic stability of many farms around the world [1]. In the context of yield forecasting, the most important models are those that implement pre-harvest forecasts, as they provide an advance estimate of what yields can be expected with a certain probability [2]. In an era of relentless climate change, increasing resource constraints, and pressure to increase production efficiency, accurate yield forecasting is becoming indispensable. Today's technologies, such as statistical and process models and machine learning (ML), offer advanced tools to help farmers and decision-makers make informed decisions [3,4].

Since Ronald A. Fisher introduced variance analysis and the principles of experimental planning, statistical methods have played an invaluable role in analyzing the impact of various factors on agricultural production. Today, thanks to the development of computer technology and access to high-quality data, statistical predictive models have become more precise and comprehensive. The use of modern models makes it possible to consider complex interactions between data from different sources, which increases their usefulness under changing climatic conditions [5,6].

Process-based models (PBMs) introduce a new dimension to forecasting by integrating knowledge of physiological, soil, and climatic aspects of potato growth. As a result, they not only simulate plant responses to different conditions, but also help plan long-term production strategies and adapt to climate change.

Machine learning and deep learning (DL) techniques add another level of sophistication to the potato yield forecasting process. With the ability to analyze large and complex datasets, ML and DL models can create more accurate representations of crop condition, enabling better agronomic interventions and optimizing resource management. These advanced technologies are opening new horizons in precision agriculture, where the use of remote sensing data, soil analysis, and weather conditions are becoming standard [7–9].

This article offers an extended overview of statistical, process, and machine learning methods used in potato yield forecasting. It identifies forward-looking developments that combine cutting-edge technologies with sustainable management strategies to meet global food and resource challenges. The future of agriculture lies in innovations that do not only increase production efficiency, but also contribute to more sustainable resource management and environmental protection.

2. Statistical Models

2.1. Historical Background

Statistical models play a key role in interpreting complex processes that influence yields, integrating data from various sources, and supporting decision-making in agricultural practice. The use of statistical methods in agriculture dates to the early 20th century, when Ronald A. Fisher introduced the analysis of variance and formal principles of experimental design [5]. In subsequent decades, agriculture benefited from advances in mathematical statistics, incorporating multivariate regression models, nonlinear approaches, mixed (hierarchical) methods, and geostatistics [10–14]. Since the 1980s and 1990s, increased computational power has enabled the implementation of more complex estimation techniques, Bayesian approaches, and the integration of spatial, phenotypic, and genetic data (Figure 1).

Over the past two decades, as technologies such as remote sensing, the Internet of Things (IoT), and big data analytics have advanced, statistical models have become even more versatile and powerful. Current research [15–17] focuses on integrating multi-source data, accounting for uncertainties, exploring climate change scenarios, and testing novel agronomic strategies. This process lays the foundations of so-called Agriculture 4.0, in which statistics become a key component of tools to support production management.

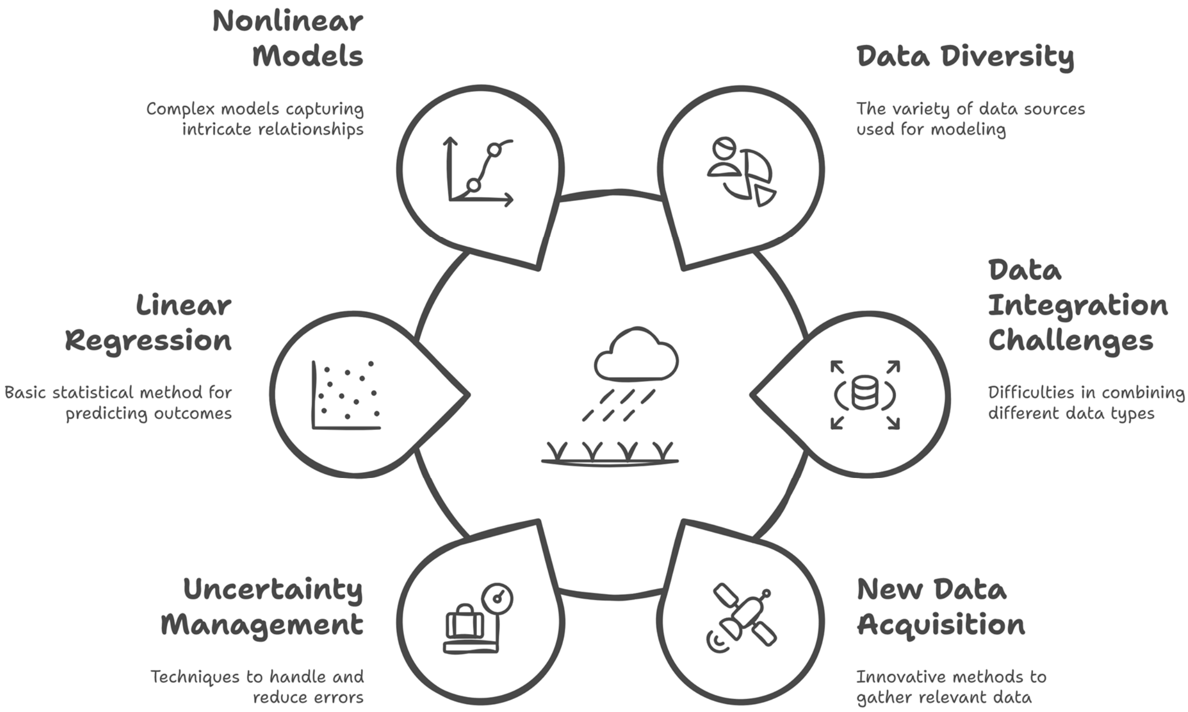

Figure 1. Overview of statistical methods and models for potato yield prediction.

2.2. Data and Model Preparation

2.2.1. Diversity and Quality of Data

Effective forecasting of potato yields relies on data from high temporal and spatial resolutions, derived from multiple sources. Several key categories can be distinguished:

1. Meteorological and Climatic Data

Traditionally, meteorological stations have provided data on precipitation, temperature, solar radiation, and humidity, but these are now complemented by climate reanalysis products, satellite-based measurements, and climate model projections. Incorporating seasonal forecasts and climate scenarios support yield predictions over both short- and long-time horizons [18,19].

2. Soil and Agronomic Data

Soil characteristics (nutrient content, pH, organic matter, water retention capacity), fertilization regimes, irrigation, plant protection measures, and the timing of planting and harvesting shape the conditions for potato growth [20,21]. These agronomic data often come from soil sampling, farm records, and monitoring systems [22].

3. Phenotypic and Genetic Data

Progress in plant phenotyping (e.g., using sensors, hyperspectral cameras, thermal imaging, or chlorophyll fluorescence analysis) and genomics enables the identification of cultivar traits and breeding lines that are sensitive or resilient to environmental stresses [23,24]. Integrating this information with ecological variables allows predictions of how specific genotypes will respond to various climate scenarios.

4. Spatial and Remote Sensing Data

Satellite imagery (Sentinel, Landsat), drones, vegetation indices (NDVI, EVI), and topographical and soil maps provide information on spatial variability in crops and their condition [25–27]. These data support the creation of yield potential maps and the identification of zones requiring differentiated inputs, thus enhancing precision agriculture practices.

Satellite data are playing an increasingly central role in modern agriculture, and their importance in predicting potato growth and yield can hardly be overstated. Platforms such as Sentinel-2 and Landsat provide high-resolution imagery to monitor key environmental parameters such as plant health, soil moisture, and temperature variability. With satellite data, large-scale crop development can be tracked, allowing vegetation conditions to be analyzed even in the most remote regions, where traditional monitoring methods are difficult to apply. One of the most important tools for analyzing satellite data are vegetation indices such as the NDVI (Normalized Difference Vegetation Index) and EVI (Enhanced Vegetation Index). The NDVI, based on the ratio of red and near-infrared reflectance, provides detailed information on plant conditions. NDVI values above 0.5 indicate good crop health, especially during key potato growth phases such as tuber initiation and flowering. The EVI, which is more robust to disturbances related to soil conditions, allows precise assessment of plants at different stages of their development [4].

Other frequently used VIs in potato yield forecasting are the GNDVI (Green NDVI), NDRE (Normalized difference red-edge), SAVI (Soil-adjusted vegetation index), and RVI (Ratio vegetation index) [28].

2.2.2. Challenges of Data Integration

The greatest difficulty lies in integrating such diverse data into a single model. Data may span different years and regions, and vary in quality, precision, and sampling frequency. Harmonization, standardization, and imputation procedures are essential. Quality checks, removal of outliers (unless biologically justified), and variable transformations are necessary to meet model assumptions [29,30].

2.2.3. New Data Acquisition Approaches

Modern IoT systems allow real-time data collection (e.g., soil moisture from in-field sensors, temperature and PAR radiation from mini-weather stations). These data can be automatically processed and integrated into statistical models, enabling continuous forecast updates as new information arrives. Data fusion methods (combining satellite and ground-based measurements) and big data analytics improve the handling of large, multidimensional datasets [26,31].

2.2.4. Managing Uncertainty and Errors

Equally crucial is quantifying uncertainty. Meteorological and climate data are subject to measurement and forecasting errors, soil measurements may not represent the entire fields, and phenotypic assessments may contain noise. Introducing Bayesian approaches and cross-validation procedures facilitate the better assessment of uncertainty and model fit quality.

2.3. Overview of Statistical Models Used in Potato Yield Forecasting

Due to the diversity of data and the complexity of the phenomenon, no single universal statistical model exists. In practice, a wide range of approaches are applied, which can be combined and tailored to specific conditions. The sections below discuss selected model classes and their applications in greater detail.

2.3.1. Linear Regression and Generalized Linear Models (GLMs)

Linear regression is the simplest and most well-known statistical approach, assuming a linear relationship between yield and a set of predictors (e.g., cumulative rainfall, fertilizer application). Although simple, it offers parameter interpretability and hypothesis testing for variable significance [29]. Linear regression is a starting point for exploratory data analysis. It requires several assumptions (normality of errors, homoscedasticity) that may not hold in practice. For count data or proportions, generalized linear models (GLMs) extend linear regression to various distributions and link functions suited to the nature of the dependent variable [7,32]. For example, Poisson GLMs can model the number of tubers meeting quality standards, while binomial GLMs can model the probability of exceeding a yield threshold.

2.3.2. Nonlinear Models

Biological relationships are often nonlinear. Yield response to fertilization may increase up to an optimum and then plateau or decline. Nonlinear models [33,34] allow fitting of logistic, power, exponential, or sigmoid curves. Nonlinear methods better represent biological reality but are harder to calibrate, prone to local minima, and yield parameters that are less intuitively interpretable. Nevertheless, they capture thresholds and optima, which are crucial for modelling temperature and water availability effects on yield [35].

Examples of the application of nonlinear models to potato yield forecasting include the following papers. Awad [36] uses a nonlinear model based on energy balance equations, in particular the METRIC (Mapping EvapoTranspiration at high Resolution with Internalized Calibration) equation. This model, which is less dependent on climatic measurements for calculating actual evapotranspiration, addresses the problems of limited remote data availability and climatic variability. The study showed that the METRIC model significantly improves the accuracy of yield estimation, which was confirmed by comparing the results with data from local meteorological stations. By using this model, it was possible to better deal with gaps in satellite data caused by adverse weather conditions, which in turn increased the availability of data for analysis. The use of a non-linear model in fitting and estimating biomass from remote data enabled a more comprehensive understanding of yield dynamics and contributed to improved yield forecasting. The model was tested in the context of potato cultivation, which is important for the agricultural economy in the study region. Comparison of the estimated yield values with the farmers' data showed close agreement, and the MAPE error value was less than 4%, indicating the high accuracy of the new model. Ultimately, the model developed provides tools to effectively manage agricultural practices, which is crucial in the context of changing climatic conditions and their impact on crop production. In a subsequent paper, Awad [37] uses a non-linear model based on the Trust-Region for Nonlinear Minimization (TR) algorithm, which optimizes yield data by fitting them to an exponential equation. This approach resulted in significant achievements in yield estimation, including significant improvements in estimation accuracy, as evidenced by an MAPE value of less than 4%. The model proved effective in dealing with gaps in satellite imagery data, which is crucial in the context of climate variability. In addition, yield analysis, based on data collected from local farmers, revealed close agreement between estimated and actual yields, indicating the high utility of the model in agronomic practice. An important result of this non-linear approach was also the improvement of crop biomass estimates, which contributes to a better assessment of crop yields. Although the research focused on potato cultivation, the model is versatile and has the potential to be applied to a variety of crops, opening up new possibilities for future research in sustainable agricultural production.

2.3.3. Mixed (Hierarchical) Models

Potato cultivation occurs in diverse farms and regions, creating a hierarchical data structure. Mixed models [12,38] combine fixed effects (e.g., fertilization impact on the entire population) with random effects (differences among regions, fields). These models more realistically reflect data complexity and avoid errors arising from ignoring grouping effects. They are particularly valuable in multi-season and multi-regional analyses, allowing generalization beyond specific locations and times. However, interpreting random effects is more challenging, and computational complexity is higher.

2.3.4. Geostatistical Methods

Geostatistics [13,14] model spatial variability. Kriging and cokriging are interpolation methods that consider spatial autocorrelation of yield values. Geostatistics produce yield prediction maps, pinpointing zones needing more intensive fertilization or irrigation, enabling precision agriculture. These methods require high-quality georeferenced data and sufficiently dense measurement grids. They focus on spatial aspects, so combining them with other models that include environmental or phenotypic factors is advisable.

2.3.5. Bayesian Approaches

Bayesian methods [12,24] treat model parameters as random variables with prior distributions. Updating these priors with new data results in posterior distributions that incorporate both data and prior knowledge. Advantages include natural handling of uncertainty, integrating expert knowledge (e.g., expected fertilizer effects in a region), and working with limited data. Disadvantages involve greater computational complexity (e.g., Monte Carlo methods) and choosing priors. Bayesian models are well-suited for conditions of high uncertainty, temporal and spatial variability, and multi-source data integration [22,39].

2.3.6. Dimension Reduction Methods (PCA, PLS)

With a large number of input variables (soil, meteorological, phenotypic, remote sensing data), collinearity and model complexity become problematic. PCA and PLS [30,40] extract key dimensions, reducing the number of predictors and stabilizing models. These methods are particularly useful when integrating satellite data and phenotypic indicators, where many potential indices exist. The downside is losing direct interpretation of original variables, as the analysis focuses on new, artificial components. Still, more parsimonious and stable models facilitate reliable short-term yield predictions [41].

2.3.7. Generalized Additive Models (GAMs)

GAMs [30,42] allow the modelling of nonlinear relationships between predictors and yield using smooth functions. They do not impose a specific functional form, offering flexibility. GAMs capture complex and nonlinear patterns, such as temperature effects in various crops growth stages. They enable analysis of how changes in one parameter modify relationships with others. Their drawback lies in the more complex interpretation of smoothing functions and the need for careful regularization to avoid overfitting.

2.3.8. Integration with Other Approaches

GAMs can be combined with mixed models, geostatistical approaches, or Bayesian frameworks, creating hybrid models that consider hierarchies, spatial structures, and nonlinear dependencies. Recent literature shows that GAMs can serve as modules in larger analytical systems or be integrated with mechanistic crop growth models [43].

2.4. Strengths and Limitations of Statistical Models

Statistical models facilitate data integration, interpretation of mechanisms affecting yield, hypothesis testing, and quantification of uncertainty. They allow selecting methods to match the problem's specifics, thus increasing flexibility and utility. Limitations mainly concern data quality, required model assumptions, computational complexity for advanced approaches, and difficulties in extrapolating to completely new conditions.

In summary, statistical methods play a key role in the analysis and forecasting of potato yields, allowing the integration of different datasets and the identification of key factors influencing variability. Among the most effective techniques are the following:

1. Multiple regression: this is the primary tool for investigating relationships between multiple variables affecting yield (e.g., temperature, rainfall, fertilizers applied). Although its interpretation is intuitive, these models have limitations when dealing with complex non-linear relationships.
2. Mixed (hierarchical) models: they allow regional or seasonal variability to be included in the data, allowing for more accurate forecasts. Example: analyzing yield differences between regions according to climatic conditions and cropping techniques.
3. Bayesian models: they are distinguished by their ability to incorporate uncertainty and priorities into the analysis. For example, if historical data suggest that precipitation is a major factor, Bayesian models can assign greater weight to this variable in forecasts.
4. Geostatistics: the use of methods such as kriging allows yield variability to be mapped in space, which is particularly useful in precision agriculture.

3. Process-Based Models

Various approaches are currently used to model yields, including process-based models (PBMs)—also known as mechanistic or biophysical crop growth models. They stand out as tools that integrate knowledge of plant physiology, soil, and climate, enabling a better understanding of the mechanisms shaping potato yields and simulating their responses to changing environmental and management conditions.

3.1. Historical Background

The initial attempts at process-based modelling in agriculture emerged in the 1960s and 1970s, when researchers began formalizing knowledge about photosynthesis, transpiration, carbon assimilation, and phenology into mathematical equations. The works of Monteith [44] and Penning de Vries et al. [45] established mechanistic approaches to describing plant growth, accounting for energy balances and the efficiency of converting solar radiation into biomass.

In subsequent decades, general crop growth models such as SUCROS, WOFOST, and LINTUL formed the foundation for developing crop-specific models, including those for potatoes [46,47]. Dedicated potato models, such as SUBSTOR-Potato [48,49], SIMPOTATO [50], and LINTUL-POTATO [51,52], integrated knowledge about tuber phenology, biomass allocation, gas exchange, and environmental factors.

During the 1990s and 2000s, process-based models began to be implemented in decision support systems (DSSAT, APSIM, STICS, CropSyst), enabling their broader use in agricultural practice [53,54]. With advances in computing power, availability of high-resolution meteorological data, GPS, GIS, and remote sensing technologies, PBMs could be increasingly calibrated, validated, and applied to diverse environmental and management scenarios.

In recent years, researchers have shown growing interest in integrating PBMs with phenotypic, genomic, and machine learning methods [55]. Studies have emerged combining PBMs with other model classes (e.g., statistical or ML-based) in hybrid approaches [56]. Advances in remote sensing (drones, Sentinel, Landsat satellites), IoT data (soil and climate sensors), and multi-scale analyses (field to global assessments) have made PBMs indispensable tools for forecasting potato yields.

The AgMIP (Agricultural Model Intercomparison and Improvement Project) initiative further promotes PBM development by comparing different crop models and climate scenarios, improving methods, and understanding uncertainties [55,57].

3.2. Strengths and Limitations of Process-Based Models

3.2.1. Strengths

1. Biological Interpretability:

PBMs describe plant growth, development, and yield formation based on known physiological processes: photosynthesis, respiration, transpiration, water and nutrient uptake, and organ development [58,59]. Results can be interpreted through biological mechanisms.

2. Potential for Extrapolation to New Conditions:

Unlike statistical models, PBMs do not rely solely on historical correlations. Their process-based nature allows the simulation of yields under novel climatic conditions, unknown management practices, genetic changes, or extreme environments [60].

3. Integration of Multiple Factors:

PBMs can incorporate a wide range of environmental factors (temperature, precipitation, radiation, CO_2), soil conditions (moisture, nutrient availability), management practices (fertilization, irrigation, crop protection), and genetic parameters (cultivar traits). They are flexible and easily extendable [59,61].

4. Support for Long-Term Scenario Analyses:

PBMs can conduct long-term simulations to assess the impacts of climate change, agricultural policies, breeding strategies, or new technologies on future potato yields [62].

3.2.2. Limitations

1. High Data and Parameter Requirements:

PBMs need extensive input data: physiological parameters, soil properties, genetic coefficients, and high-resolution meteorological data. Collecting and processing these data is costly and time-consuming, limiting usage in data-poor environments [63,64].

2. Calibration and Validation Requirements:

Models must be calibrated against empirical data (field experiments, phenotypic measurements) and validated before use for forecasting [46,56].

3. Simplifications and Omissions:

Despite aiming for biological realism, PBMs include simplifications—assuming uniform soil conditions, omitting certain diseases or pests, and applying simplified plant structures [59].

4. Sensitivity to Input Data Quality:

Errors in meteorological, soil, or phenotyping data can propagate and cause forecast uncertainty [63].

Despite these constraints, PBMs remain integral to modern agricultural analytics. Technological progress (IoT, remote sensing, genomic selection) and improved data integration methods are reducing barriers related to data availability and accuracy.

3.3. Overview of Process-Based Models

Below are ten process-based models applied to potato yield forecasting. The selection includes both well-established models and newer projects. Recent studies integrate these models with innovative data sources and computational methods (Figure 2).

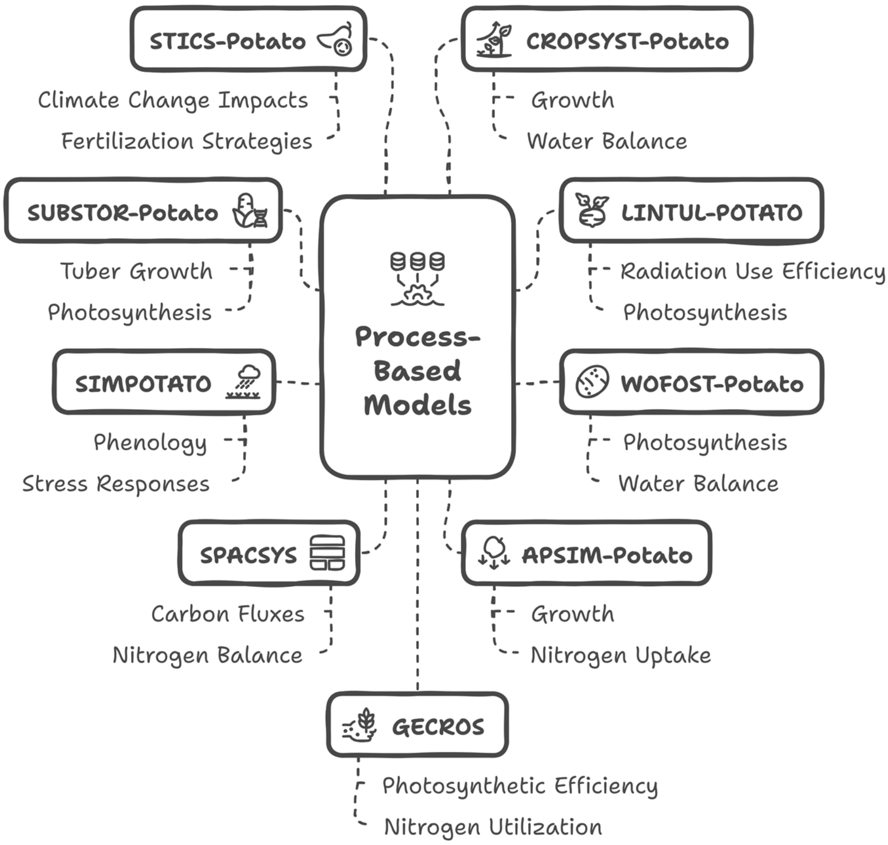

Figure 2. Overview of process-based models for potato yield prediction.

3.3.1. SUBSTOR-Potato (DSSAT Module)

SUBSTOR-Potato, part of DSSAT, simulates tuber growth based on photosynthesis, respiration, phenology, and water and nutrient uptake [48]. It requires cultivar parameters and detailed weather and soil data. Daccache et al. [65] and Wolf and Van Oijen [66] applied DSSAT to analyze climate and CO_2 effects on potato yields in various scenarios.

Application: widely used for evaluating fertilization strategies, irrigation scheduling, climate change impacts, and testing new cultivars and management strategies [56].

3.3.2. LINTUL-POTATO

LINTUL-POTATO is a simplified model focusing on radiation use efficiency and photosynthesis [51]. It considers temperature and soil moisture influences on biomass accumulation and tuber growth. Franke et al. [67] employed simplified models like LINTUL to assess temperature stress sensitivity during tuber formation.

Application: its simplicity makes it suitable for scenario analyses of environmental factors' impacts on yields, initial assessments of climate change implications, or testing new conditions.

3.3.3. WOFOST-Potato

WOFOST is a general crop growth model with potato-specific implementations addressing photosynthesis, water balance, nutrient uptake, and phenology [47,68]. De Wit et al. [69] utilized WOFOST with remote sensing data to evaluate potato growth conditions at the continental scale.

Application: used for regional analyses, integration with remote sensing data, and spatio-temporal yield variability assessments.

3.3.4. SIMPOTATO

SIMPOTATO is specifically designed for potatoes, detailing phenology, biomass allocation, and responses to water and temperature stress [70]. Maiorano et al. [71] enhanced understanding of extreme weather impacts on tuber phenology and yields using models like SIMPOTATO.

Application: applied in breeding research, variety testing, analyzing drought or cold periods, and assessing management strategies under changing climate conditions.

3.3.5. SPACSYS

SPACSYS is an ecosystem model that simulates C, N, and water fluxes. Although not exclusively for potatoes, it can be parameterized for the crop [72]. Wu et al. [73] used ecosystem models to evaluate nitrogen use efficiency under changing climate conditions, applicable to potato scenarios.

Application: useful in analyzing nitrogen balance, greenhouse gas emissions, and the environmental impact of cropping systems, including potato cultivation.

3.3.6. APSIM-Potato (APSIM Module)

APSIM is a platform for farming system simulation. Its potato module describes growth, nitrogen and water uptake, phenology, and management/environment interactions [53]. Jägermeyr et al. [56] used APSIM in AgMIP comparisons, evaluating climate scenarios and multiple crops, including potatoes.

Application: employed for strategic production analysis, long-term yield forecasting, and optimizing fertilization and irrigation under climate variability.

3.3.7. STICS-Potato

STICS is a multi-criteria model covering plant physiology, soil, weather, and management [54]. Potato parameterizations simulate tuber phenology, N, and water dynamics. Jégo et al. [74] evaluated the impact of various agricultural practices on nitrate leaching under the root zone of potato and sugar beet using the STICS soil–crop model.

Application: used to assess climate change impacts, test fertilization strategies, optimize water use, and integrate phenotypic and remote sensing data.

3.3.8. CROPSYST-Potato

CropSyst models growth, water and nitrogen balances, soil erosion, greenhouse gas emissions, and management effects. Its potato parameterization allows simulations under various systems and conditions [75]. Alva et al. [76] carried out a study to integrate and validate the CSPotato model with CropSystVB to predict yield, soil processes, and nitrogen transport in irrigated potato rotational systems, demonstrating its suitability as a decision-making tool.

Application: long-term scenario analysis, evaluating crop rotations, fertilization and irrigation strategies, and environmental impacts.

3.3.9. GECROS (Generic Crop Growth Simulation Model) for Potatoes

GECROS was developed mainly for cereals but can be adapted for potatoes. It emphasizes photosynthesis, respiration, nitrogen distribution, and canopy-level photosynthetic efficiency [77]. Khan et al. [78] conducted a study to apply the GECROS model to analyze differences in potato tuber yield between five varieties of different maturity and 100 diploid F1 genotypes, to identify key nitrogen-related traits (Nmax and tuber N concentration) affecting yield, and optimize these traits in the context of specific growing environments.

Application: studying photosynthetic efficiency, nitrogen utilization, responses to elevated CO_2 and temperatures, and strategic breeding approaches.

3.3.10. Local Models (e.g., SOLANUM) and Hybrid Projects

Under names like SOLANUM, locally developed potato models emerge in research projects targeting specific regions. These models integrate detailed phenological data, local cultivar parameters, and soil information [51].

Application: local condition analysis, testing new varieties, supporting breeding, and adapting to agro-climatic conditions. Often serves as a base for integration with modern techniques (machine learning).

3.4. Prospects for the Development and Integration of Process-Based Models

Advances in Potato PBMs Are Tied to Several Key Trends

1. Integration with Phenotypic and Genetic Data:

Merging PBMs with genetic and phenotypic profiles enables prediction of genotype-specific responses to environmental conditions [79].

2. Hybrid Models with Machine Learning:

Combining process knowledge with ML methods (neural networks, Bayesian approaches, Random Forest) improves prediction accuracy, stability, and handling of data gaps and uncertainties [80].

3. Integration with Remote Sensing and IoT:

Drone and satellite imagery (Sentinel, Landsat), along with ground-based sensors, allow real-time updates to model parameters, verifying canopy status, soil moisture, and nutrient availability [26].

4. Inclusion of Biotic Stresses and Tuber Quality:

Efforts to integrate disease, pest, and tuber quality modules into PBMs are underway. This would enable more complete field-condition simulations [59].

5. Multi-Scale and Policy Analyses:

Linking PBMs with economic, social, and climatic models allows evaluation of production stability at regional and global scales, informing food security policy, agricultural strategies, and trade considerations [56].

4. Machine and Deep Learning Models

The development of machine learning (ML) technologies is bringing significant benefits to agriculture, especially in the context of potato yield forecasting. ML models use advanced algorithms to analyze complex and nonlinear relationships in large datasets, making them extremely useful tools with high predictive and adaptive potential. Their ability to analyze data in detail allows for predictions both at the level of entire fields and individual plants, enabling precise agronomic interventions.

By integrating a variety of data, such as satellite imagery, information on soil physical parameters, or nutrient availability, for example, machine learning models can create a comprehensive picture of crop health and yield post-tenure. This, in turn, supports farmers in improving irrigation practices, fertilizer application accuracy, and pest control, which in the long run contributes to increased productivity and production efficiency [81,82].

The use of advanced predictive models is important not only at the level of individual farms, but also for the entire agricultural supply chain. Accurate yield forecasts allow for proper planning of storage, transportation, and sales, which stabilizes the market and minimizes the risk of food shortages. Although the use of ML comes with challenges, such as the need for high-quality data or end-user understanding of the results, its potential in transforming agriculture and supporting sustainable development strategies is enormous [83,84]. From the point of view of agricultural practice, the most valuable predictive models are those that report the amount of expected pre-harvest yields for the current growing season.

The use of machine learning in potato yield prediction offers many advantages, but it also comes with significant limitations that are worth considering in detail before tackling such analyses. One of the biggest advantages of cranking out is the ability to automate processes related to crop monitoring and management. With accurate yield forecasts, farmers can automatize decisions on irrigation, fertilization, and crop protection, which not only saves time, but also increases operational efficiency. This automation allows for more sustainable resource management, which is crucial in the face of growing food security and environmental challenges [85]. In addition, the use of machine learning can lead to the increased ecological efficiency of crops. Accurate yield predictions and optimization of fertilization and irrigation allow farmers to reduce the number of chemicals applied, thus benefiting soil health and ecosystem protection. In the long term, this approach can contribute to sustainable agriculture, which is becoming increasingly important in the face of climate change and environmental degradation [84].

However, despite its numerous advantages, the use of ML in the prediction of potato yields faces significant limitations. First, the implementation of machine learning models involves the need for expertise in data analysis and programming. Farmers who lack experience in these areas may face difficulties in using these technologies effectively. As agriculture becomes more technical, it becomes necessary to introduce education and training programs to help farmers understand and take advantage of the opportunities that ML brings.

Another challenge is the sensitivity of ML models to environmental changes. These models can fall short when faced with unforeseen climatic events, such as extreme rainfall or drought, which were not accounted for during model training. In the face of sudden changes, this can lead to significant errors in forecasts, which in turn affect decision-making in crop management [86,87]. Therefore, it is important to keep learning yields updated with non-typical data as well. While the application of machine learning in potato yield prediction holds great potential for automation, ecological efficiency, and process optimization, it is at the same time essential to address the challenges associated with the need for specialized knowledge and environmental sensitivity. As technologies evolve, it will become crucial to establish a solid educational and infrastructural foundation that will enable farmers to realize the full potential of ML in the sustainable management of their fields.

Deep learning (DL) is gaining prominence in potato yield forecasting, offering advanced modelling methods that can significantly improve prediction accuracy. In a world where agriculture faces increasing challenges related to food security, climate change, and limited natural resources, the use of innovative technologies such as deep learning is becoming crucial. Deep independent models allow for more sophisticated analysis, making it possible to capture subtle patterns in the data that can have a significant impact on crop performance [88].

In potato yield forecasting, deep learning is applied to the analysis of a variety of data, including meteorological, agronomic, and spectral data. For example, convolutional neural networks (CNNs) are particularly effective in analyzing remote sensing data. With the ability to process images, CNNs can effectively assess crop performance by identifying plant health from images taken from UAVs (unmanned aerial vehicles). This enables ongoing monitoring of health parameters such as leaf density and vegetation cover. On the other hand, Long Short-Term Memory (LSTM) networks are ideal for modelling temporal data. They can capture temporal relationships and better predict how climate changes, such as temperature and precipitation, affect potato growth. LSTMs can analyze long-term trends, as well as short changes in atmospheric conditions, allowing for more accurate predictions. Deep learning also allows the creation of hybrid models that integrate various ML algorithms, such as Random Forest, with neural networks. This type of approach combines the power of both methods, leading to even better forecasting results.

Despite its many advantages, there are challenges to implementing deep learning. The required amount of data to effectively train the models and the need for adequate data quality remain key barriers. Additionally, the interpretability of deep learning models is an important issue, as users need to understand how various factors affect predictions. In the future, it will be important to develop more transparent models that enable better interpretation of results [80,89–92].

4.1. Example Techniques

4.1.1. Random Forest

Random Forest (RF) is one of the most popular machine learning techniques that is gaining importance in potato yield forecasting. It is an ensemble method that relies on creating multiple decision trees and then combining their results to produce a more accurate and stable prediction. With its ability to efficiently handle large datasets and analyze complex, non-linear relationships between different variables, Random Forest is particularly useful in the agricultural context [10].

For potato yield forecasting, the technique can consider several factors, such as soil conditions, meteorological data, cultivation techniques, and plant variety characteristics. Random Forest is also resistant to over-fitting, making it an effective tool in situations where the available data are complex or contain outliers. As a result, Random Forest-based models can provide reliable yield forecasts, allowing farmers to make informed decisions about crop management and optimizing production processes.

The algorithm starts by randomly selecting samples from the entire dataset (Figure 3). Each decision tree is trained on a different random subset of the data, allowing for a variety of models. When building a tree for each sample, the algorithm makes split decisions based on feature values. At each decision node, the algorithm makes a split based on the best criterion (e.g., Gini coefficient or entropy) to maximize the difference between categories in the data. Another key element is random feature selection. At each branching of the tree, the algorithm decides from a limited number of features chosen at random which features to use for partitioning. This makes the trees less correlated, resulting in increased diversity in the model. After building multiple trees, Random Forest makes predictions, considering the results from all the trees. For classification problems, it decides the class based on a majority vote (voting method), while in a regression context, it averages the results of all the trees to get the final prediction value.

Random Forest is also equipped with model evaluation techniques, such as cross-validation, which allows for parameter optimization and model efficiency. Due to its design, the technique can adapt to complex and non-linear relationships in the data, making it the preferred method for many applications, including potato yield forecasting [8,9].

The RF algorithm has become one of the key methods in forecasting land yields, gaining recognition for its ability to efficiently analyze diverse datasets and model complex relationships in the agricultural ecosystem. Accurate yield forecasting is absolutely essential for the rational use of agrotechnical practices and for ensuring broad food security, which is becoming an increasingly pressing issue in a changing climate. Nowadays, the RF method is widely used and adapted to lo-kal demands for accurate predictive models. The purpose of Huang et al.'s (2024) study was to interpret and evaluate the accuracy of the RF model in forecasting potato yields in China and to quantify the main sources of uncertainty using C.T. de Wit's triplet diagram [93]. A dataset of 2182 plot and year combinations was extracted from 63 potato field eco-experiments conducted in nine Chinese provinces over a three-year period. Model performance was evaluated using 10-fold cross-validation (CV), block validation (LBOCV), location-free validation (LSOCV) and year-free validation (LYOCV). The resulting root mean square error (RMSE) values were 3.5, 8.3, 9.9, and 10.3 t ha-1, respectively, while the model efficiency coefficient (MEC) was 0.92, 0.64, 0.52, and 0.43 for 10-fold CV, LBOCV, LSOCV, and LYOCV. Sunshine duration and the topographic position index emerged as the most important explanatory variables, while fertilizer variables were identified as the least important in modelling yields. The standard deviation of yield variability from replication, estimated using a linear model, accounted for 32% of the RMSE for LSOCV. The introduction of measured nutrient uptake rates and yields from fertilization as additional explanatory variables, on average, reduced the RMSE for LSOCV by 2.3 t ha-1. The fitted models could explain up to 92% of the variability in potato yields in China, although a significant residual error was noted when extrapolating to other areas or years.

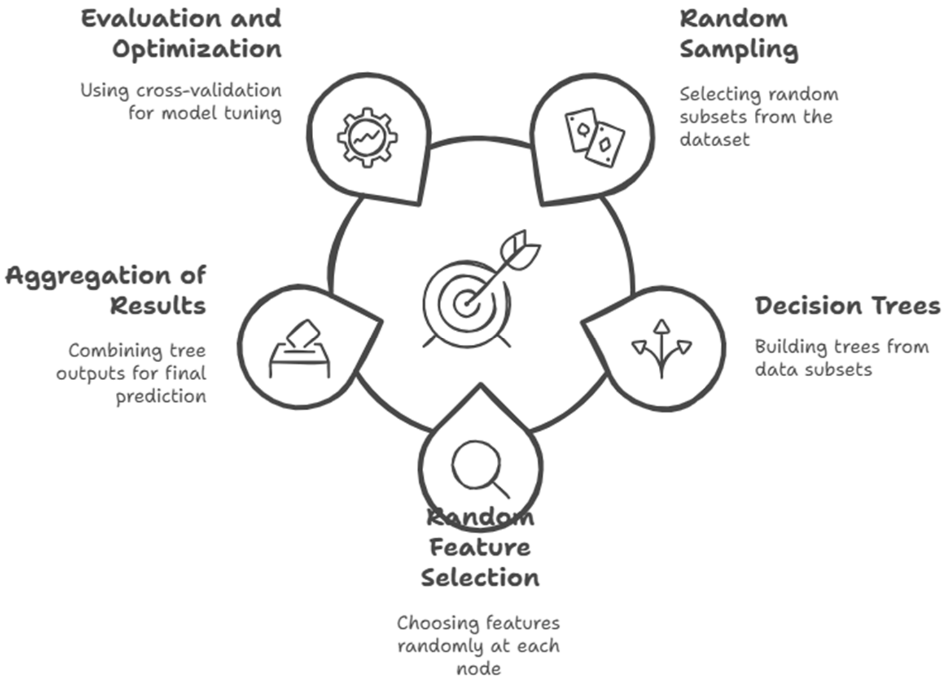

Figure 3. Components of the Random Forest algorithm in machine learning. Own work.

Other studies have compared various predictive models, including Random Forest and other state-of-the-art machine learning algorithms such as K-nearest neighbors (KNN), gradient boosting, and XGBoost. The results of the study by El-Kenawy et al. (2024) [80] indicated that while gradient boosting and XGBoost were effective at predicting potato yields, deep learning algorithms such as Random Forest, graph-based neural networks (GNNs), and LSTM not only achieved higher accuracy rates, but could also effectively capture complex spatial and temporal data patterns. For example, gradient boosting achieved an MSE of 0.03438 and an R^2 of 0.49168, while XGBoost achieved an MSE of 0.03583 and an R^2 of 0.35106. GNNs, on the other hand, showed an MSE of 0.02363 and an R^2 of 0.51719, making them the best model in terms of overall performance. No less promising were LSTM, which achieved an MSE of 0.03177, and GRU with an MSE of 0.0315.

The development of UAV technology has made it possible to collect high-resolution images, a key element in accurate yield forecasting. A study by Tatsumi and Usami (2024) in 2022 focused on using these technologies to assess and forecast tuber weight, tuber number, and fresh weight per tuber. A total of 648 statistical variables were collected, and the RF model achieved high prediction accuracy, with R^2 values of 0.57 for fresh tuber weight, 0.45 for tuber number, and 0.49 for fresh weight per tuber. These results indicate the significant potential of RF in yield prediction based on advanced UAV image analysis [94].

At the same time, climate change is affecting agricultural productivity around the world, as evidenced by a study on potato and maize yield forecasting in Rwanda. In this case, RF proved to be the most successful model, achieving an RMSE of 510.8 for potatoes and 129.9 for maize, with R^2 values of 0.875 and 0.817, respectively. These results underline RF's ability to predict yields in the context of changing weather conditions, which can provide key information for farmers when planning future harvests [95].

From a forward-looking perspective, the use of Random Forest in potato yield forecasting, combined with remote sensing data, offers great potential to improve production efficiency. Research points to the need for greater integration of data on varieties, soil properties, and management techniques to create more complex predictive models that can better adapt to local conditions and a changing climate.

High-resolution yield maps are key to identifying patterns of yield variation and providing guidance for management in pre-precision agriculture. Differences among potato (*Solanum tuberosum* L.) varieties can significantly affect yield forecasting using remote sensing technology. The goal of the Li et al. (2023) study was to improve potato yield forecasting using remote sensing from drones (UAVs) by incorporating variety information into machine learning models, especially Random Forest Regression (RFR). In 2018 and 2019, researchers conducted experiments with different potato varieties and nitrogen (N) application rates, while applying models such as Random Forest Regression (RFR) and Support Vector Regression (SVR) and collecting multispectral UAV images [96]. The results showed that data from the early growth stage, at the tuber initiation stage (in late June), were more correlated with market yields. The performance of RFR and SVR models from remote sensing data was unsatisfactory (R^2 = 0.48–0.51), but when variety information was included, the R^2 increased to 0.75–0.79. Combining high-resolution UAV images and variety information using RFR significantly improved yield forecasting. Further research is needed to include more detailed variety data, soil variables, and management techniques.

In a study of potato yield forecasting in Kazakhstan, machine learning methods, including Random Forest, also showed very good results. An analysis of weather data from 1990–2023 proved that the Random Forest algorithm achieved a coefficient of determination (R^2) value of 0.97865, indicating high forecast accuracy. RMSE values ranged from 0.25 to 0.46, suggesting relatively low error rates [97].

Further research should focus not only on improving the RF algorithm, but also on exploring other machine learning models and integrating them with UAV data and meteorological information. The joint use of these resources can support farmers in making informed decisions that will contribute to sustainable agriculture and increased yields. Ultimately, advanced predictive models such as RF can become key tools in tackling food security and climate change issues, which are becoming increasingly pressing challenges for modern agriculture [98].

4.1.2. Neural Networks

In recent years, the use of neural networks in potato yield forecasting has gained prominence due to their ability to analyze complex patterns in agricultural data. Neural networks can effectively model nonlinear relationships between variables, which is important in the context of potato farming, where yield performance is affected by many factors such as soil conditions, climatic variables, and management techniques.

In addition, multilayer perceptron (MLP) networks have also gained recognition in yield forecasting. MLPs, by virtue of their architecture, can model complex relationships between multiple variables, making them useful for highly complex data. The use of MLPs in combination with other machine learning techniques can lead to further improvements in the accuracy of potato yield forecasts [99–102].

MLPs have gained popularity in this context because of their ability to model complex, nonlinear relationships between various variables that affect yields. These networks are relatively easy to implement, require fewer computational resources than more advanced deep learning models, and can efficiently process a wide variety of data types, making them an attractive choice in the agricultural context.

The MLP architecture consists of at least three layers of neurons: an input layer, one or more hidden layers, and an output layer. The neurons in each layer are connected to each other, and these connections are assigned weights, which are modified during the learning process. The operation of an MLP network can be described in several steps (Figure 4).

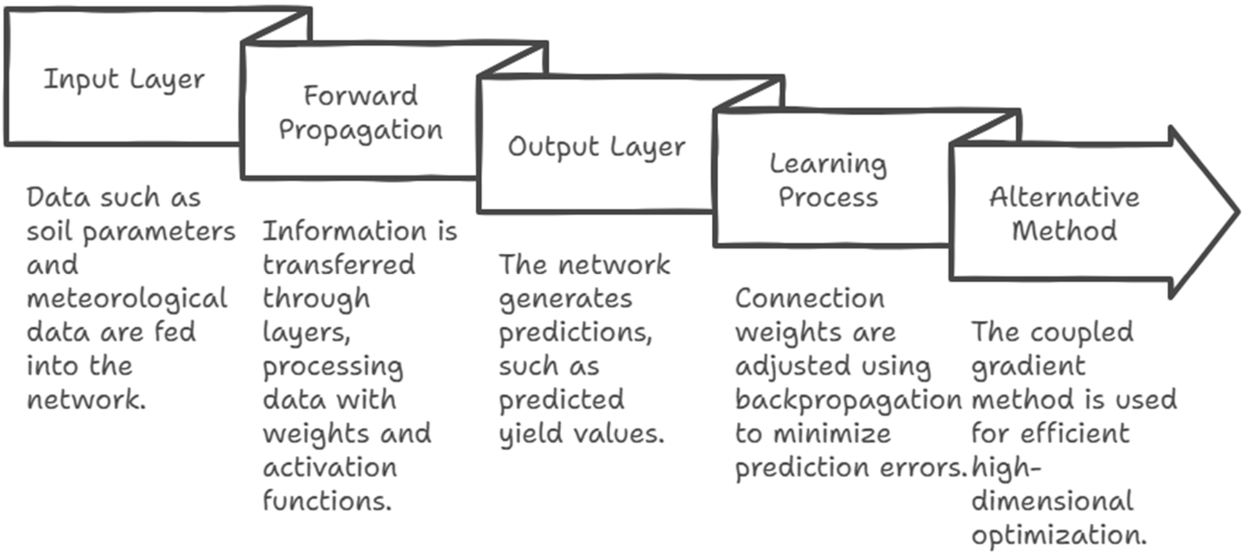

Figure 4. Operation of MLP neural network.

In potato yield forecasting, neural networks, and in particular the algorithms of artificial neural networks (ANNs), are gaining importance due to their ability to model complex relationships between diverse agricultural data. As a result, they can significantly improve the accuracy of yield forecasts, thereby demonstrating many advantages in the context of agronomic data analysis.

One study on the use of ANNs in the prediction of earth potato yields focused on the prediction of potato yields of three early varieties: Arielle, Riviera, and Viviana. Analyzing data from 2010–2017, linear and non-linear models were created to predict yields as of June 20. MLP-based models achieved an MAPE (mean absolute percentage error) of no more than 15%, indicating the high quality of the predictions. The NY1 neural model had better results in terms of quality measures and forecast errors compared to the traditional RY1 regression model, confirming the effectiveness of the application of artificial intelligence methods in agriculture [99].

A study in China that analyzed the impact of meteorological factors showed that the optimized BP model, using the WOA optimization algorithm, significantly outperformed other models. Its coefficient of determination (R^2) was 0.9764, demonstrating exceptional yield forecasting accuracy, based on data from Internet of Things (IoT) systems collected from 2010 to 2022 [103].

In a study on the effect of soil tillage systems on tuber yield, the ANN model obtained a correlation coefficient of 0.97 and an RMSE of 0.077, suggesting that the neural network was able to effectively determine the relationship between potato yields and soil properties and tillage systems [100].

Another study focused on the relationship between energy and potato yields in Saudi Arabia. The ANN model with a 6-15-22-1 architecture, used to predict yield, performed highly, with an R^2 of 0.704 and a mean absolute error of 2.36 tons/ha with a relative error of 5.59%. Causal analysis showed that energy from irrigation had the greatest impact on yields, highlighting the importance of efficient resource management in the production process [104].

The objective of the Fortin et al. (2010) experiment was to optimize a neuron network (NN) model for predicting potato tuber growth in eastern Canada, taking into account climatic data and the cumulative and maximum leaf area index (LAI) [105]. Analyses were carried out using data from 2010–2017 from e-field experiments. The NN model, based on comprehensive data analysis, achieved an MAE (mean absolute error) of 209 kg DM ha^{-1}, which was less than 4% of the average yield. The study showed that the most important variables for tuber growth were solar radiation and cumulative rainfall. The NN model integrated with LAI significantly improved prediction accuracy compared to traditional regression. In the context of the validation data, the R^2 was 0.93 for tuber growth. In addition, the study showed that the use of the Whale Optimization Algorithm (WOA) resulted in an R^2 coefficient of determination of 0.9764. A comparison of the regression results showed that traditional regression models, such as MLR, were less effective, with their RMSE not exceeding 0.29 t/ha in the analyses. The use of neural networks significantly contributed to better modelling of temporal variation in tuber growth. The juxtaposition of MAE and R^2 values underscores the importance of nonlinear methods in pre-precision potato yield forecasting.

4.1.3. Deep Learning

Deep learning, combined with image analysis, is ideal for assessing the quality of crops, including potatoes. Inspection methods that are fast and harmless, such as those based on deep learning and image processing, can be extremely useful in monitoring the quality of potatoes and other agricultural products. Deep learning encompasses a range of architectures and techniques that are commonly used in machine learning, including in the context of crop yield prediction. Key methods in this field are convolutional neural networks, which are particularly effective in image and computer vision analysis. CNNs use convolution operations to extract features from visual data for accurate pattern recognition, such as plant health and soil analysis. Recurrent neural networks (RNNs), especially improved versions such as LSTMs and GRUs (Gated Recurrent Units), are also used in potato yield prediction. These networks enable analysis of sequential data, which is important in the context of temporal analyses, such as changes in yield in response to different climatic or seasonal conditions. With their ability to store past information, LSTMs and GRUs can handle gradient degradation, making them useful for predicting future performance from past data. In addition, generative networks, including generative adversarial models (GANs), can be used to generate synthetic data that support yield modelling and forecasting, especially in situations where actual data are limited. Autoencoders can be used to compress input data and reduce dimensions to facilitate analysis. Integrating these methods in the context of crop yield prediction allows for a comprehensive approach to data analysis and develops the ability to accurately predict yields under changing environmental conditions [106]. The use of deep learning in this field brings new perspectives and significantly contributes to the development of sustainable agriculture.

In a study by Samatha et al. (2023), a set of images of infected potatoes was used as input to a DNN model in the training and testing phase [92]. In addition, researchers used a modified support vector machine (MSVM) and a convolutional neural network for segmentation and feature extraction of potato images. The proposed model was tested on modern versions of standard datasets maintained by the departments of agriculture of Canada and the United States. Experimental results showed that the deep learning model outperformed both traditional methods and other machine learning models. It was found that some types of images achieved an accuracy of more than 99%, while others achieved an accuracy of 97%. These results underscore the effectiveness of using zaa-advanced image analysis techniques in assessing plant health and their usefulness in agricultural practices.

One of the most promising approaches is the use of LSTM-type neural networks. They can effectively capture both sequential and temporal aspects of the data, allowing analysis of temporal data such as temperature and rainfall to accurately predict potato yields based on long-term trends. Studies have shown that LSTM models can achieve higher forecast accuracy compared to traditional methods, as evidenced by the values of mean square error (MSE) and coefficient of determination (R^2) obtained in analyses [107].

Another approach is convolution-based neural networks (CNNs), which are adapted for image analysis. In the context of remote sensing, CNNs can utilize UAV data, allowing the use of multispectral imagery for yield forecasting. Research shows that convolutional networks are effective in identifying key plant traits and their health, which directly affects yield forecasts [108,109].

The use of the YOLO (You Only Look Once) network in potato yield forecasting is becoming increasingly relevant due to its effectiveness in object detection and image segmentation. YOLO can be effectively applied to the analysis of images acquired using remote sensing technologies such as drones (UAVs) and satellites. Through rapid image processing, the model can detect and classify various aspects of a potato crop, such as the presence of tubers, leaf health, and potential pest infestations. All these elements affect the final tuber yield, which is why ongoing monitoring of plantations is so important.

Accurate detection of potato seedlings is key to obtaining information on crop health and increasing yields, and their evaluation is important for building future yields. The goal of the Wang et al. (2024) study was to improve potato seedling detection in drone images using a new lightweight model, VBGS-YOLOv8n, which is an improved version of YOLOv8n. A dataset with images of potato seedlings was used to train the model. The model used a lighter VanillaNet architecture as a backbone and introduced a weighted two-way pyramid network for feature extraction, which reduced information loss between layers and improved detection performance. The model achieved 1,524,943 parameters and 4.2 billion FLOPs, uz-achieving an accuracy of 97.1%, an average accuracy of 98.4%, and an inference time of just 2.0 ms. Compared to other models, such as YOLOv8, its parameters decreased by 51.7% and accuracy increased by 1.4%. The study confirmed that VBGS-YOLOv8n is an effective tool for detecting potato seedlings in remote imagery, offering valuable insights for future applications in precision agriculture [110].

In conclusion, the use of deep learning methods such as LSTMs and CNNs in potato yield forecasting offers promising opportunities to increase the accuracy and precision of predictive models. Integrating these state-of-the-art techniques with remote sensing data and detailed varietal information could be a key step toward increasing potato production efficiency and sustainable resource management in agriculture. Future research should focus on further refinement of these models and their implementation into practical solutions for farmers.

4.1.4. Support Vector Machines

The use of the support vector machine (SVM) method in potato yield forecasting is becoming increasingly popular due to its effectiveness in analyzing complex, non-linear relationships in agricultural data. The SVM is a classification and regression method that works by finding an optimal hyperplane that maximizes the distance between different classes of data. In the context of potato yield forecasting, the SVM can be used to analyze a variety of variables such as soil parameters, climatic data, and variety information [111].

One of the main strengths of the SVM is its ability to cope with high dimensionality of data, making it an ideal tool when dealing with complex datasets from multiple sources. Studies have shown that the SVM, especially when combined with optimization techniques, can achieve high yield forecasting accuracy. In one study, the SVM model was found to achieve a coefficient of determination (R^2) of 0.817 for potato yield datasets, indicating a high quality of forecasts [112].

The SVM method can also be used with a variety of kernel functions, allowing complex non-linear relationships to be modelled effectively. This allows the model to be better adapted to the specific local conditions and variables that affect potato yield. Furthermore, the inclusion of SVM in the forecasting process, especially in combination with other algorithms, can lead to significant improvements in the accuracy and efficiency of forecasts [113].

SVM is a classification and regression method that relies on the identification of an optimal hyperplane that maximizes the margin between different classes of data, which, in the context of potato yield forecasting, becomes extremely important when taking into account the many factors that influence crop performance [114].

The principle of the SVM is shown in Figure 5. The first step in applying the SVM in yield forecasting is to collect relevant data, which can range from climatic data to soil parameters and previous crop performance. The accuracy of the forecasts is closely related to the quality of the collected data and its representativeness for specific agrotechnical conditions. Data collection is followed by pre-processing, which includes normalization and standardization, ensuring that all features have equal weight in the analysis. The selection of a kernel function that transforms the data to higher dimensions is another key step. Different forms of kernel, such as a linear kernel, a Radial Basis Function (RBF) kernel, or a polynomial kernel, can be used depending on the specifics of the data, allowing non-linear relationships to be modelled efficiently. Once the kernel function is defined, the SVM model is trained on a training dataset. In this stage, the model adjusts the connection weights to optimize the classification error. The use of optimization techniques, such as cross-validation, allows the model's parameters to be tuned, resulting in better prediction accuracy. The training stage is followed by prediction, in which the SVM model generates predictions on a test set. The performance of the model is then assessed using measures such as the coefficient of determination, MAE, and mean square error (MSE).

The literature indicates that SVM models can achieve high performance, with R^2 values reaching 0.85 and above, indicating their ability to accurately predict potato yields considering local conditions [115].

In conclusion, the SVM is a powerful tool in potato yield forecasting that allows for accurate modelling of complex data relationships. Its use, especially when combined with data analytics techniques, has the potential to significantly improve crop efficiency and sustainable resource management in agriculture. Further research should focus on the optimization of SVM models and their integration with other machine learning techniques.

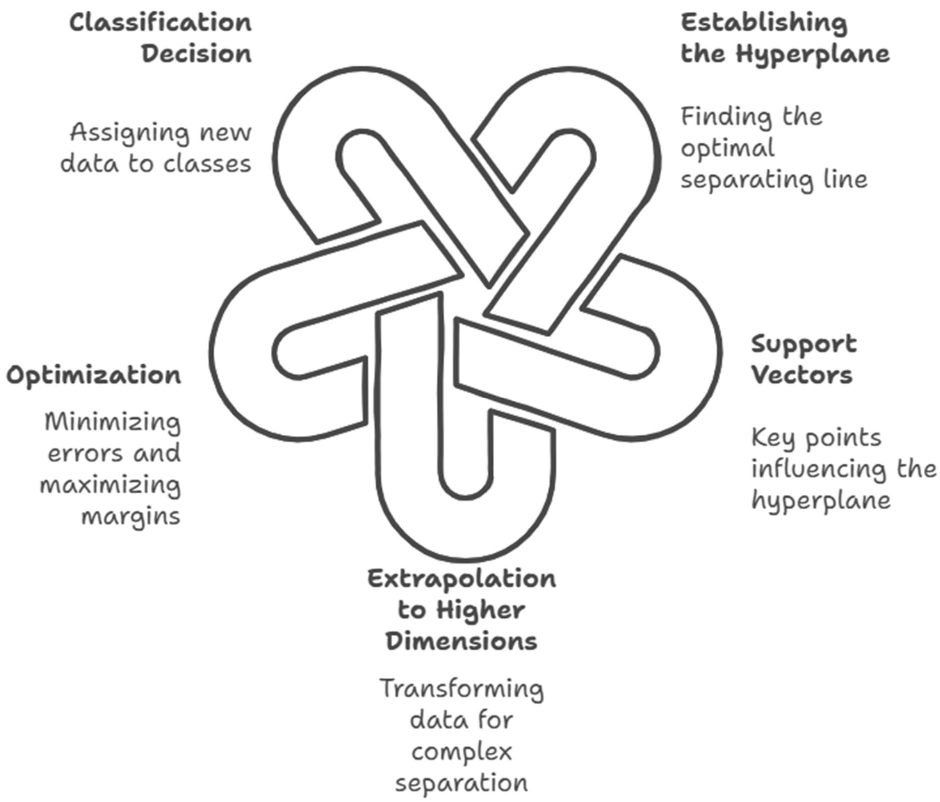

Figure 5. Principle of the SVM model.

A study by Salvador et al. (2020), conducted in Mexico, used SVMs to estimate potato yields at the local level, based on meteorological data from the ERA5 set, satellite images from the TERRA platform, and field data [116]. In the winter cycle, the Random Forest model performed best (%RMSE = 18.9, R^2 = 0.757), but in the summer cycle the SVM with a polynomial kernel (svmP) achieved higher accuracy (R^2 = 0.858, %RMSE = 14.9). These results demonstrated the potential of SVMs in yield modelling, highlighting the importance of early seasonal data collection.

A study by Kurek et al. (2022), conducted in Poland, included five years of data from 36 commercial potato fields, integrating agronomic, climatic, soil, and satellite remote sensing data [41]. The use of advanced data processing techniques, such as principal component analysis (PCA) and SVM outlier detection, resulted in the hybrid model achieving the lowest MAPE error (5.85%). This study highlighted the effectiveness of a multidisciplinary approach combining multiple data sources, leading to more accurate yield forecasts.

A study by Singh and Swain (2020) looked at forecasting rice and potato yields in the Tarakeswar region of India, considering a variety of variables such as climatic factors and agricultural practices [117]. While the SVM performed well for the training data, significant over-fitting problems during validation indicate the need for further optimization of the model. Nevertheless, the DNN model achieved the highest prediction accuracy for potatoes (R^2 = 0.97, RMSE = 0.95 t/ha), demonstrating the benefits of combining the SVM with other machine learning techniques.

In summary, the SVM presents significant potential in potato yield forecasting, especially when using advanced data processing techniques and integrating different sources of information. Despite the challenges of over-fitting, the SVM, in combination with other machine learning methods, can support accurate yield forecasting, influencing sustainable agricultural development and strategic agronomic decision-making.

4.2. Integrating Remote Sensing and Big Data in Potato Yield Forecasting

The integration of remote sensing and big data in potato yield forecasting is a key approach that combines cutting-edge technology with innovative agricultural management strategies. Remote sensing, using technologies such as satellite imagery and data from unmanned aerial vehicles, provides high-resolution images that allow monitoring of crop health, soil conditions, and various environmental factors over large areas. With these technologies, it is possible to achieve real-time analysis of variables such as temperature, soil moisture, and vegetation cover, which are key to analyzing potato growth dynamics [41,118,119].

In combination with remote sensing data, big data analysis plays a fundamental role in the processing and interpretation of huge agronomic datasets. The use of machine learning methods, such as Random Forest regression models, SVMs, and neural networks, enables more accurate yield forecasting from different data sources. Advanced ML algorithms can analyze complex relationships between climatic, soil, and agronomic data, allowing for more accurate crop yield predictions. Examples of applications, such as modelling potato growth behavior in response to changing conditions, perfectly illustrate how ML can be used to optimize agricultural practices [80,120]. Firstly, satellite data can be combined with meteorological and soil sensor data to provide a more complete picture of potato growing conditions. This approach enables better resource management by farmers by providing information on when and where it is best to apply fertilizers or crop protection products. Secondly, with advanced data analysis techniques such as machine learning and artificial intelligence, it is possible to identify patterns that might otherwise be overlooked. This allows the prediction not only of potential yields, but also the identification of threats, such as diseases or pests, before they become a measurable problem. Thirdly, examples of applications of this integration include the design of precision irrigation systems that use soil and meteorological data to optimize water management, and decision support programs for farmers that suggest the best action based on analyzed patterns [121,122].

The integration of remote sensing and big data also makes it possible to identify spatial patterns of yield variability, which supports management tailored to specific local soil and climatic conditions. By using advanced data analysis techniques such as principal component analysis and outlier detection, the yield forecasting process can be optimized [123]. Studies have shown that this combination is able to achieve a low mean absolute error, which significantly improves the accuracy of the forecasts [41].

Through the synergy of remote sensing, big data, and machine learning, agricultural management systems can better adapt operational practices to changing conditions, leading to more sustainable use of resources and minimization of losses. Application examples, such as UAV-based plant health monitoring and ML algorithms for yield prediction, are gaining importance in the context of precision agriculture [116]. The integration of remote sensing, big data, and machine learning in potato yield forecasting offers promising solutions to increase production efficiency and ensure food security. Future research should focus on the further development of these models and their implementation in practical solutions to fully exploit the potential of these technologies in agriculture.

4.3. Integration of the NDVI and Other Spectral Indices with Weather Data Assimilation Models in Potato Yield Prediction

The integration of the NDVI (Normalized Difference Vegetation Index) and other spectral indices into weather data assimilation models represents a significant step forward in yield forecasting, particularly in the context of potato cultivation. The NDVI is widely used to assess plant health and monitor plant growth; analysis of red and near-infrared reflectance yields values ranging from -1 to 1. In potatoes, optimum NDVI values are most often recorded at critical growth stages such as tuber initiation and flowering, where values above 0.5 indicate healthy plants [28,124].

In addition to the NDVI, other important spectral indices, such as the EVI (Enhanced Vegetation Index) and LAI, also play a key role in yield forecasting. The EVI, which is less sensitive to changes in soils, can reach values in the region of 0.2–0.8 depending on vegetation status [28]. The LAI, on the other hand, which measures the total leaf area per unit area of soil, can take values from 2 to 5 for healthy potato plants in full growth [125].

Various sources can be used to obtain values for vegetation indices. Satellite images from platforms such as Sentinel-2 or Landsat provide high-resolution data and allow the calculation of the NDVI and other spectral indices. In the case of UAVs, equipment equipped with multispectral cameras allows even more accurate local monitoring of plant health. In addition, meteorological data from local weather stations are crucial for analyzing the conditions under which potatoes grow.

In the context of meteorological data, information on air temperature, rainfall, insolation, and soil moisture is most used. Evapotranspiration, the process by which water evaporates from the soil surface and transpires through the plants, is a key parameter that should be included in predictive models, especially where crop irrigation is used. Accurate determination of parameters related to evapotranspiration allows for a better understanding of the impact of water on crop productivity [4,99,124].

The use of these tools and methods allows for more effective forecasting of potato yields and adaptation of agronomic practices to changing conditions. By synergizing the NDVI, other spectral indices, and weather data with ML and DL techniques, farmers can make more informed decisions, leading to more sustainable crop management and increased production efficiency in the face of increasing food demands.

5. Challenges and Directions in Potato Yield Forecasting

Predicting potato yields using modern methods such as machine learning is becoming a key element in ensuring food security and optimizing agricultural management processes. However, it faces several challenges that can affect the accuracy and effectiveness of predictive models. The first significant challenge is the high cost and complexity of data processing. To implement effective predictive models, farmers and researchers need to invest significant resources in technology and expertise. This requires sophisticated infrastructure for data collection, storage, and processing, which can be a barrier for many smaller farms [126].

Integrating heterogeneous data sources into a coherent system is another challenge. The collation of data from different sources, such as remote sensing, meteorological measurements, and agronomic data, is crucial for effective modelling. Problems arise from the variety of formats, types, and quality of data, which complicates the integration process and reduces the quality of forecasts. Data availability and quality are also significant barriers. Many regions face limited access to reliable, high-resolution data, which can significantly affect the accuracy of forecasts. In the context of potato yield forecasting, difficult-to-access meteorological data and a lack of historical yield data can limit the models' capabilities [127].

Climate variability introduces additional uncertainty in yield models. Uncontrolled weather conditions, associated with the effect of climate change, can significantly affect crop yields. Extreme weather conditions, such as droughts or heavy rainfall, can distort accepted models, making yield prediction difficult [128,129].

Interpretability of models is also a challenge. The increase in the use of advanced ML algorithms, such as SVMs and neural networks, comes with the need for transparency in their operation. Users need to be able to understand how models make decisions in order to have confidence in the predictions issued [130].

Finally, the scalability of models is a critical challenge. Models need to be flexible enough to adapt to different local conditions, crop sizes, and a variety of management techniques, while maintaining forecast accuracy.

To improve potato yield forecasting, several key areas need attention. Hybrid models are becoming increasingly popular, combining traditional statistical models with modern machine learning techniques. This type of approach can harness the power of each approach, leading to better forecasting results [41].

Explainable AI is another important development. The development of interpretable ML models should improve decision-making and increase confidence in forecast outcomes. Examples of such initiatives include techniques for visualizing the impact of variables on forecast outcomes and building models that can be easily interpreted by users. Real-time prediction systems that integrate IoT devices and data streams represent the future in yield forecasting. The use of real-time data allows farming practices to be dynamically adapted to current conditions [131].

Open data initiatives, which are gaining prominence in the agricultural context, offer the opportunity to share information, enabling better understanding and use of data in yield forecasting. Such initiatives create platforms where farmers, researchers, and policy makers can collaborate and share their insights and data. This makes it possible to build more comprehensive predictive models that consider local conditions as well as different cultivation techniques. Open data can include a variety of information, such as meteorological datasets, soil analyses, as well as the results of field experiments. By sharing such data, patterns and trends can be better identified, which is particularly important in the context of climate variability. Examples of data sharing include collaborations between research institutions, government agencies, and non-profit organizations that collect, compare, and publish data to support evidence-based agriculture. These initiatives not only increase the availability of information, but also promote innovative solutions that can lead to more sustainable resource management. They also enable the development of data-driven applications that can be used by farmers in their daily crop management. As technology develops, open data provide the foundation for future innovations in yield forecasting and precision agriculture, contributing to efficient food production in a sustainable way [132,133].

6. Conclusions

Statistical models remain fundamental to analyzing and forecasting potato yields due to their simplicity, interpretability, and ability to leverage historical data effectively. These models serve as the foundation for more advanced methods, offering a clear framework to identify key relationships between variables such as soil properties, climatic factors, and crop management practices. By quantifying these relationships, statistical models provide crucial insights into the factors driving yield variability, which can be instrumental for agricultural decision-making.

Furthermore, statistical models are increasingly being integrated with mechanistic crop models, enhancing their predictive capabilities. Mechanistic models simulate the biological processes underlying crop growth, such as photosynthesis, respiration, and nutrient uptake. When combined with statistical models, these hybrid approaches allow for more accurate yield predictions by accounting for both empirical data trends and biological mechanisms.

Statistical models also play a key role in simulating scenarios, particularly in the context of climate change. By incorporating projected climate variables, these models can estimate how future temperature, precipitation, and CO_2 levels might influence potato yields. This capability supports policymakers and agricultural stakeholders in planning adaptive strategies to mitigate potential negative impacts.

As technological advancements continue—such as the proliferation of high-resolution satellite imagery, IoT-based soil sensors, and real-time weather data—statistical models are expected to evolve further. These innovations will improve the granularity and timeliness of input data, thereby enhancing model accuracy. Moreover, the development of user-friendly computational tools and platforms will make statistical yield forecasting more accessible to a broader range of users, from researchers to farmers.

In conclusion, while statistical models are inherently limited by their reliance on historical data and simplified assumptions, their adaptability and compatibility with emerging methodologies ensure their continued relevance. As integration with advanced technologies progresses, statistical models will likely achieve greater precision and broader applicability, solidifying their role as a cornerstone of agricultural forecasting.

Process-based models (PBMs) are indispensable in potato yield forecasting due to their capacity to integrate diverse dimensions of agricultural systems, including plant physiology, soil dynamics, climatic factors, and management practices. These models simulate the underlying biological and physical processes that govern crop growth and development, offering a comprehensive framework for understanding how various factors interact to shape yields. This integrative approach not only enhances predictive accuracy but also provides insights into the causal mechanisms driving yield variability, which are crucial for effective decision-making.

One of the most significant strengths of PBMs lies in their ability to simulate physiological processes such as photosynthesis, respiration, nutrient uptake, and water use efficiency. By modelling these processes, PBMs allow researchers and practitioners to investigate how changes in environmental conditions, such as temperature, precipitation, and atmospheric CO_2 levels, affect crop performance. For potatoes, which are particularly sensitive to soil moisture and nitrogen availability, PBMs can identify critical growth stages where interventions may have the most significant impact.

Moreover, PBMs support scenario-based and long-term analyses, making them invaluable for addressing complex challenges like climate change. By incorporating climate projections, these models can simulate future conditions and evaluate potential impacts on potato yields across different regions. This capability allows researchers to test various adaptation strategies, such as adjusting planting dates, adopting drought-resistant cultivars, or modifying irrigation and fertilization practices, to mitigate adverse effects and capitalize on opportunities presented by changing climates.

The growing integration of genetic and phenotypic information further enhances the utility of PBMs. Advances in genomics and high-throughput phenotyping provide detailed data on crop traits, such as drought tolerance, nitrogen use efficiency, and disease resistance. By incorporating these traits into PBMs, researchers can simulate the performance of specific genotypes under varying environmental and management scenarios. This facilitates targeted breeding programs aimed at developing potato varieties optimized for specific conditions, thereby bridging the gap between theoretical modelling and practical application.

PBMs are also instrumental in evaluating innovative management strategies. For example, they can simulate the effects of precision agriculture techniques, such as variable-rate irrigation and fertilization, on yield and resource efficiency. Similarly, PBMs can assess the potential benefits of integrating cover crops or catch crops into potato rotations to reduce nitrate leaching and enhance soil health. These evaluations enable stakeholders to adopt evidence-based practices that improve productivity while ensuring environmental sustainability.

Despite their strengths, PBMs require extensive data input, including detailed soil properties, weather data, and crop-specific parameters, which can be challenging to obtain. However, advancements in remote sensing, IoT-based field monitoring, and machine learning are progressively addressing these limitations by providing high-resolution, real-time data to support model calibration and validation.

In summary, process-based models are essential for advancing potato yield forecasting, as they offer a robust platform for integrating diverse sources of knowledge and addressing both current and future challenges. Their ability to analyze complex interactions, predict outcomes under various scenarios, and guide innovation positions PBMs as central tools in the pursuit of sustainable and resilient potato production systems.

Machine learning plays a key role in potato yield prediction, introducing innovative approaches to the analysis of agricultural data. One of the main strengths of ML models is their ability to model complex, non-linear relationships between different factors affecting potato growth and yield, such as temperature, soil moisture, and weather conditions. Models such as Random Forest, gradient boosting, or neural networks show high efficiency in analyzing these interactions, making it possible to obtain accurate yield forecasts.

The integration of diverse data sources is another key aspect of applying machine learning in this field. Combining satellite data, soil sensor measurements, and meteorological information allows for the creation of comprehensive models that support complex analytics and enable the generation of personalized forecasts at the field or specific fertilizer input levels. It is also worth highlighting the importance of transfer learning, which allows models trained on different datasets to be used to forecast yields in new contexts. This practice is particularly useful in regions where large local datasets are lacking, thus increasing the versatility of models and their applicability under different agronomic conditions.

With accurate yield forecasts, machine learning enables farmers to better plan activities such as irrigation, fertilization, or harvesting times. This information supports more informed decision-making, leading to more efficient use of resources and increased production profitability. In addition, ML models can help identify crop risks, such as pest threats or climate change. Responding early to these potential problems contributes to better risk management and greater production stability.

There are a number of challenges and opportunities in potato yield forecasting that are relevant to both agricultural practice and research. Key challenges include the variability of climatic conditions, which significantly affect plant growth and yield, as well as the difficulty of obtaining accurate data on soil properties, irrigation, and agronomic methods. In addition, the complexity of agricultural ecosystems and interactions between different environmental factors can make effective forecasting difficult. Despite these difficulties, there are also significant opportunities. Technological advances, including the development of mathematical models and tools based on data acquired from satellites and sensors, offer new prospects for more accurate yield forecasting. Collaboration between researchers and agricultural practitioners, combined with the application of modern technologies, can contribute to a better understanding of crop specifics and the adaptation of agronomic strategies to changing conditions. Therefore, an in-depth analysis of the challenges and opportunities associated with potato yield forecasting is essential to more effectively support sustainable agricultural development.

Future research should focus on using Internet of Things (IoT) technology to collect real-time data to better understand crop conditions. It will also be important to analyze the impact of biological variables, such as different genotypes, on yield under varying environmental conditions. An interdisciplinary approach is recommended, integrating knowledge from ecology, agronomy, and meteorology, which could lead to the development of new predictive models that take into account complex interactions. In the context of the increasing use of machine learning methods, a key aspect will be not only the validation and calibration of existing models based on data from real crop conditions, but also the exploration of innovative algorithms that can significantly improve the quality of predictions. The introduction of techniques such as deep learning could make it possible to analyze large datasets and identify subtle patterns that are difficult to capture by traditional analytical methods. In addition, research should consider the impact of sustainable agricultural practices on yields to understand how different management methods can improve both productivity and ecosystem health. Implementation of the above recommendations, along with the development of machine learning methods, has the potential to significantly advance knowledge in the area of potato yield forecasting and contribute to the development of more efficient and sustainable agricultural practices.

In conclusion, the application of machine learning in potato yield prediction is an area that bridges technology and agriculture, offering practical solutions that can significantly improve productivity and production efficiency in the face of growing global challenges.

Author Contributions: Conceptualization, M.P. and G.N.; methodology, M.P. and G.N.; software, G.N.; validation, G.N., formal analysis, M.P. and G.N.; investigation, M.P. and G.N.; resources, M.P. and G.N.; data curation, M.P. and G.N.; writing—original draft preparation, M.P. and G.N.; writing—review and editing, M.P. and G.N.; visualization, M.P.; supervision, G.N., project administration, M.P.; funding acquisition, G.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study.

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

References

1. Vreugdenhil, D.; Bradshaw, J.; Gebhardt, C.; Govers, F.; MacKerron, D.; Taylor, M.; Ross, H. *Potato Biology and Biotechnology: Advances and Perspectives*; Elsevier: Amsterdam, The Netherlands, 2007; ISBN 9780444510181.
2. Niedbała, G.; Nowakowski, K.; Rudowicz-Nawrocka, J.; Piekutowska, M.; Weres, J.; Tomczak, R.J.; Tyksiński, T.; Pinto, A.Á. Multicriteria prediction and simulation of winter wheat yield using extended qualitative and quantitative data based on artificial neural networks. *Appl. Sci.* **2019**, *9*, 2773. [CrossRef]
3. Roy, T.; George, K.J. Precision Farming: A Step Towards Sustainable, Climate-Smart Agriculture. In *Global Climate Change: Resilient and Smart Agriculture*; Springer: Singapore, 2020; pp. 199–220.
4. Hara, P.; Piekutowska, M.; Niedbała, G. Selection of Independent Variables for Crop Yield Prediction Using Artificial Neural Network Models with Remote Sensing Data. *Land* **2021**, *10*, 609. [CrossRef]
5. Fisher, R.A. Biological monographs and manuals. In *Statistical Methods for Research Workers*, 3rd ed.; Oliver and Boyd: Edinburgh, UK; London, UK, 1925.
6. Gómez, D.; Salvador, P.; Sanz, J.; Casanova, J.L. New spectral indicator Potato Productivity Index based on Sentinel-2 data to improve potato yield prediction: A machine learning approach. *Int. J. Remote Sens.* **2021**, *42*, 3426–3444. [CrossRef]
7. Dobson, A.J.; Barnett, A.G. *An Introduction to Generalized Linear Models*, 3rd ed.; CRC Press: Boca Raton, FL, USA, 2008.
8. Hara, P.; Piekutowska, M.; Niedbała, G. Prediction of Pea (*Pisum sativum* L.) Seeds Yield Using Artificial Neural Networks. *Agriculture* **2023**, *13*, 661. [CrossRef]
9. Bharti; Das, P.; Banerjee, R.; Ahmad, T.; Devi, S.; Verma, G. Artificial Neural Network Based Apple Yield Prediction Using Morphological Characters. *Horticulturae* **2023**, *9*, 436. [CrossRef]
10. Cochran, W.G.; Cox, G.M. *Experimental Designs*, 2nd ed.; John Wiley & Sons, Inc.: New York, NY, USA, 1957.
11. Snedecor, G.W.; Cochran, W.G. *Statistical Methods*, 8th ed.; Wiley-Blackwell, Ltd.: Hoboken, NJ, USA, 1989.
12. Gelman, A.; Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*; Cambridge University Press: Cambridge, UK, 2006; ISBN 9780521867061.
13. Cressie, N.A.C. *Statistics for Spatial Data*; Wiley Series in Probability and Statistics; Wiley: Hoboken, NJ, USA, 1993; ISBN 9780471002550.
14. Webster, R.; Oliver, M.A. *Geostatistics for Environmental Scientists*, 2nd ed.; John Wiley and Sons: Hoboken, NJ, USA, 2008; ISBN 9780470028582.

15. Cătuna Petrar, T.M.; Brașovean, I.; Racz, C.-P.; Mîrza, C.M.; Burduhos, P.D.; Mălinaș, C.; Moldovan, B.M.; Odagiu, A.C.M. The Impact of Agricultural Inputs and Environmental Factors on Potato Yields and Traits. *Sustainability* **2024**, *16*, 8759. [CrossRef]
16. Zhang, N.; Luo, H.; Li, H.; Bao, M.; Liu, E.; Shan, W.; Ren, X.; Jia, Z.; Siddique, K.H.M.; Zhang, P. Maximizing potato tuber yields and nitrogen use efficiency in semi-arid environments by precision fertilizer depth application. *Eur. J. Agron.* **2024**, *156*, 127147. [CrossRef]
17. Salan, M.S.A.; Hossain, M.M.; Sumon, I.H.; Rahman, M.M.; Kabir, M.A.; Majumder, A.K. Measuring the impact of climate change on potato production in Bangladesh using Bayesian Hierarchical Spatial-temporal modeling. *PLoS ONE* **2022**, *17*, e0277933. [CrossRef]
18. Rötter, R.P.; Carter, T.R.; Olesen, J.E.; Porter, J.R. Crop–climate models need an overhaul. *Nat. Clim. Change* **2011**, *1*, 175–177. [CrossRef]
19. Xu, Q.; Liang, H.; Wei, Z.; Zhang, Y.; Lu, X.; Li, F.; Wei, N.; Zhang, S.; Yuan, H.; Liu, S.; et al. Assessing Climate Change Impacts on Crop Yields and Exploring Adaptation Strategies in Northeast China. *Earth's Future* **2024**, *12*, e2023EF004063. [CrossRef]
20. Soil Survey Staff. *Keys to Soil Taxonomy*, 13th ed.; USDA-Natural Resources Conservation Service: Olympia, WA, USA, 2022.
21. Haverkort, A.J.; Struik, P.C. Yield levels of potato crops: Recent achievements and future prospects. *Field Crop. Res.* **2015**, *182*, 76–85. [CrossRef]
22. Fan, Y.; Liu, Y.; Yue, J.; Jin, X.; Chen, R.; Bian, M.; Ma, Y.; Yang, G.; Feng, H. Estimation of potato yield using a semi-mechanistic model developed by proximal remote sensing and environmental variables. *Comput. Electron. Agric.* **2024**, *223*, 109117. [CrossRef]
23. Raymundo, R.; Asseng, S.; Cammarano, D.; Quiroz, R. Potato, sweet potato, and yam models for climate change: A review. *Field Crop. Res.* **2014**, *166*, 173–185. [CrossRef]
24. Duan, Y.; Duan, S.; Xu, J.; Zheng, J.; Hu, J.; Li, X.; Li, B.; Li, G.; Jin, L. Late Blight Resistance Evaluation and Genome-Wide Assessment of Genetic Diversity in Wild and Cultivated Potato Species. *Front. Plant Sci.* **2021**, *12*, 710468. [CrossRef]
25. Dorigo, W.A.; Zurita-Milla, R.; de Wit, A.J.W.; Brazile, J.; Singh, R.; Schaepman, M.E. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *Int. J. Appl. Earth Obs. Geoinf.* **2007**, *9*, 165–193. [CrossRef]
26. Mulla, D.J. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosyst. Eng.* **2013**, *114*, 358–371. [CrossRef]
27. Marshall, M.; Belgiu, M.; Boschetti, M.; Pepe, M.; Stein, A.; Nelson, A. Field-level crop yield estimation with PRISMA and Sentinel-2. *ISPRS J. Photogramm. Remote Sens.* **2022**, *187*, 191–210. [CrossRef]
28. Mukiibi, A.; Machakaire, A.T.B.; Franke, A.C.; Steyn, J.M. A Systematic Review of Vegetation Indices for Potato Growth Monitoring and Tuber Yield Prediction from Remote Sensing. *Potato Res.* **2024**. [CrossRef]
29. Draper, N.R.; Smith, H. *Applied Regression Analysis*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2014; ISBN 9780471170822.
30. Jolliffe, I.T. *Principal Component Analysis*; Springer Series in Statistics; Springer: New York, NY, USA, 2002; ISBN 0-387-95442-2.
31. Alahmad, T.; Neményi, M.; Nyéki, A. Applying IoT Sensors and Big Data to Improve Precision Crop Production: A Review. *Agronomy* **2023**, *13*, 2603. [CrossRef]
32. Park, S.J.; Hwang, C.S.; Vlek, P.L.G. Comparison of adaptive techniques to predict crop yield response under varying soil and land management conditions. *Agric. Syst.* **2005**, *85*, 59–81. [CrossRef]
33. Seber, G.A.F.; Wild, C.J. *Nonlinear Regression*; Wiley Series in Probability and Statistics; Wiley: Hoboken, NJ, USA, 1989; ISBN 9780471617600.
34. Gorzelany, J.; Belcar, J.; Kuźniar, P.; Niedbała, G.; Pentoś, K. Modelling of Mechanical Properties of Fresh and Stored Fruit of Large Cranberry Using Multiple Linear Regression and Machine Learning. *Agriculture* **2022**, *12*, 200. [CrossRef]
35. Martre, P.; Yin, X.; Ewert, F. Modeling crops from genotype to phenotype in a changing climate. *Field Crop. Res.* **2017**, *202*, 1–4. [CrossRef]
36. Awad, M.M. Toward Precision in Crop Yield Estimation Using Remote Sensing and Optimization Techniques. *Agriculture* **2019**, *9*, 54. [CrossRef]
37. Awad, M.M. An innovative intelligent system based on remote sensing and mathematical models for improving crop yield estimation. *Inf. Process. Agric.* **2019**, *6*, 316–325. [CrossRef]
38. Bojacá, C.R.; García, S.J.; Schrevens, E. Analysis of Potato Canopy Coverage as Assessed Through Digital Imagery by Nonlinear Mixed Effects Models. *Potato Res.* **2011**, *54*, 237–252. [CrossRef]
39. Hjelkrem, A.-G.R.; Eikemo, H.; Le, V.H.; Hermansen, A.; Nærstad, R. A process-based model to forecast risk of potato late blight in Norway (The Nærstad model): Model development, sensitivity analysis and Bayesian calibration. *Ecol. Model.* **2021**, *450*, 109565. [CrossRef]
40. Garhwal, A.S.; Pullanagari, R.R.; Li, M.; Reis, M.M.; Archer, R. Hyperspectral imaging for identification of Zebra Chip disease in potatoes. *Biosyst. Eng.* **2020**, *197*, 306–317. [CrossRef]
41. Kurek, J.; Niedbała, G.; Wojciechowski, T.; Świderski, B.; Antoniuk, I.; Piekutowska, M.; Kruk, M.; Bobran, K. Prediction of Potato (*Solanum tuberosum* L.) Yield Based on Machine Learning Methods. *Agriculture* **2023**, *13*, 2259. [CrossRef]
42. Hastie, T.; Tibshirani, R. Generalized Additive Models. *Stat. Sci.* **1986**, *1*, 297–310. [CrossRef]

43. Hu, T.; Zhang, X.; Khanal, S.; Wilson, R.; Leng, G.; Toman, E.M.; Wang, X.; Li, Y.; Zhao, K. Climate change impacts on crop yields: A review of empirical findings, statistical crop models, and machine learning methods. *Environ. Model. Softw.* **2024**, *179*, 106119. [CrossRef]
44. Monteith, J.L.M. Climate and the efficiency of crop production in Britain. *Philos. Trans. R. Soc. B Biol. Sci.* **1977**, *281*, 277–294. [CrossRef]
45. De Vries, F.W.T.P.; Brunsting, A.H.M.; Van Laar, H.H. Products, requirements and efficiency of biosynthesis a quantitative approach. *J. Theor. Biol.* **1974**, *45*, 339–377. [CrossRef] [PubMed]
46. van Ittersum, M.K.; Donatelli, M. Modelling cropping systems—Highlights of the symposium and preface to the special issues. *Eur. J. Agron.* **2003**, *18*, 187–197. [CrossRef]
47. Boogaard, H.; Wolf, J.; Supit, I.; Niemeyer, S.; van Ittersum, M. A regional implementation of WOFOST for calculating yield gaps of autumn-sown wheat across the European Union. *Field Crop. Res.* **2013**, *143*, 130–142. [CrossRef]
48. Jones, J.W.; Hoogenboom, G.; Porter, C.H.; Boote, K.J.; Batchelor, W.D.; Hunt, L.A.; Wilkens, P.W.; Singh, U.; Gijsman, A.J.; Ritchie, J.T. The DSSAT cropping system model. *Eur. J. Agron.* **2003**, *18*, 235–265. [CrossRef]
49. Šťastná, M.; Toman, F.; Dufková, J. Usage of SUBSTOR model in potato yield prediction. *Agric. Water Manag.* **2010**, *97*, 286–290. [CrossRef]
50. Hodges, T.; Johnson, S.L.; Johnson, B.S. A Modular Structure for Crop Simulation Models: Implemented in the SIMPOTATO Model. *Agron. J.* **1992**, *84*, 911–915. [CrossRef]
51. Kooman, P.; Rabbinge, R. An Analysis of the Relation between Dry Matter Allocation to the Tuber and Earliness of a Potato Crop. *Ann. Bot.* **1996**, *77*, 235–242. [CrossRef]
52. Kooman, P.L.; Haverkort, A.J. Modelling development and growth of the potato crop influenced by temperature and daylength: LINTUL-POTATO. In *Potato Ecology and Modelling of Crops Under Conditions Limiting Growth, Proceedings of the Second International Potato Modeling Conference, Wageningen, The Netherlands, 17–19 May 1994*; Haverkort, A.J., MacKerron, D.K.L., Eds.; Springer: Dordrecht, The Netherlands, 1995; pp. 41–59.
53. Keating, B.A.; Carberry, P.S.; Hammer, G.L.; Probert, M.E.; Robertson, M.J.; Holzworth, D.; Huth, N.I.; Hargreaves, J.N.; Meinke, H.; Hochman, Z.; et al. An overview of APSIM, a model designed for farming systems simulation. *Eur. J. Agron.* **2003**, *18*, 267–288. [CrossRef]
54. Brisson, N.; Gary, C.; Justes, E.; Roche, R.; Mary, B.; Ripoche, D.; Zimmer, D.; Sierra, J.; Bertuzzi, P.; Burger, P.; et al. An overview of the crop model stics. *Eur. J. Agron.* **2003**, *18*, 309–332. [CrossRef]
55. Rosenzweig, C.; Jones, J.W.; Hatfield, J.L.; Ruane, A.C.; Boote, K.J.; Thorburn, P.; Antle, J.M.; Nelson, G.C.; Porter, C.; Janssen, S.; et al. The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agric. For. Meteorol.* **2013**, *170*, 166–182. [CrossRef]
56. Jägermeyr, J.; Müller, C.; Ruane, A.C.; Elliott, J.; Balkovic, J.; Castillo, O.; Faye, B.; Foster, I.; Folberth, C.; Franke, J.A.; et al. Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. *Nat. Food* **2021**, *2*, 873–885. [CrossRef] [PubMed]
57. Wolf, J. Comparison of two potato simulation models under climate change. I. Model calibration and sensitivity analyses. *Clim. Res.* **2002**, *21*, 173–186. [CrossRef]
58. Hassanpanah, D.; Anzabi, S.P.; Giglou, P.S.; Gorji, A.M.; Anzabi, E.P.; Hassanpanah, H.; Jelodar, Y.J.; Anzabi, F.P.; Shadbahr, M. The Effect of Matrix Nutrition by Sulfonated Silicon With S8 on Tuber Yield and Its Components of Potato Cultivars Under Water Deficit Stress Condition. *J. Agric. Sci.* **2023**, *15*, 99. [CrossRef]
59. Boote, K.J.; Jones, J.W.; White, J.W.; Asseng, S.; Lizaso, J.I. Putting mechanisms into crop production models. *Plant. Cell Environ.* **2013**, *36*, 1658–1672. [CrossRef]
60. Asseng, S.; Ewert, F.; Rosenzweig, C.; Jones, J.W.; Hatfield, J.L.; Ruane, A.C.; Boote, K.J.; Thorburn, P.J.; Rötter, R.P.; Cammarano, D.; et al. Uncertainty in simulating wheat yields under climate change. *Nat. Clim. Chang.* **2013**, *3*, 827–832. [CrossRef]
61. Soltani, A.; Sinclair, T.R. *Modeling Physiology of Crop Development, Growth and Yield*; CABI: Delémont, Switzerland, 2012.
62. Rosenzweig, C.; Elliott, J.; Müller, C.; Deryng, D.; Chryssanthacopoulos, J.; Boote, K.J.; Büchner, M.; Foster, I.; Glotter, M.; Heinke, J.; et al. The Global Gridded Crop Model Intercomparison: Data and modeling protocols for Phase 1 (v1.0). *Geosci. Model Dev.* **2015**, *8*, 261–277. [CrossRef]
63. Wallach, D.; Makowski, D.; Jones, J.W.; Brun, F. *Working with Dynamic Crop Models: Methods, Tools and Examples for Agriculture and Environment*; Academic Press: Cambridge, MA, USA, 2018.
64. Liu, Q.; Yang, M.; Mohammadi, K.; Song, D.; Bi, J.; Wang, G. Machine Learning Crop Yield Models Based on Meteorological Features and Comparison with a Process-Based Model. *Artif. Intell. Earth Syst.* **2022**, *1*, 1–34. [CrossRef]
65. Daccache, A.; Weatherhead, E.K.; Stalham, M.A.; Knox, J.W. Impacts of climate change on irrigated potato production in a humid climate. *Agric. For. Meteorol.* **2011**, *151*, 1641–1653. [CrossRef]

66. Wolf, J.; Van Oijen, M. Model simulation of effects of changes in climate and atmospheric CO_2 and O_3 on tuber yield potential of potato (cv. Bintje) in the European Union. *Agric. Ecosyst. Environ.* **2003**, *94*, 141–157. [CrossRef]
67. Franke, A.C.; Muelelwa, L.N.; Steyn, J.M. Impact of climate change on yield and water use efficiencies of potato in different production regions of South Africa. *S. Afr. J. Plant Soil* **2020**, *37*, 244–253. [CrossRef]
68. Spitters, C.; van Keulen, H.; van Kraalingen, D. A simple and universal crop growth simulator: SUCROS87. In *Simulation and Systems Management in Crop Protection*; Springer: Berlin/Heidelberg, Germany, 1989.
69. de Wit, A.; Boogaard, H.; Fumagalli, D.; Janssen, S.; Knapen, R.; van Kraalingen, D.; Supit, I.; van der Wijngaart, R.; van Diepen, K. 25 years of the WOFOST cropping systems model. *Agric. Syst.* **2019**, *168*, 154–167. [CrossRef]
70. Ahmed, M.; Fatima, Z.; Iqbal, P.; Kalsoom, T.; Abbasi, K.S.; Shaheen, F.A.; Ahmad, S. Potato Modeling. In *Systems Modeling*; Springer: Singapore, 2020; pp. 383–401.
71. Maiorano, A.; Martre, P.; Asseng, S.; Ewert, F.; Müller, C.; Rötter, R.P.; Ruane, A.C.; Semenov, M.A.; Wallach, D.; Wang, E.; et al. Crop model improvement reduces the uncertainty of the response to temperature of multi-model ensembles. *Field Crop. Res.* **2017**, *202*, 5–20. [CrossRef]
72. Wu, L.; McGechan, M.B.; McRoberts, N.; Baddeley, J.A.; Watson, C.A. SPACSYS: Integration of a 3D root architecture component to carbon, nitrogen and water cycling—Model description. *Ecol. Model.* **2007**, *200*, 343–359. [CrossRef]
73. Wu, L.; Misselbrook, T.H.; Feng, L.; Wu, L. Assessment of nitrogen uptake and biological nitrogen fixation responses of soybean to nitrogen fertiliser with SPACSYS. *Sustainability* **2020**, *12*, 5921. [CrossRef]
74. Jégo, G.; Martínez, M.; Antigüedad, I.; Launay, M.; Sanchez-Pérez, J.M.; Justes, E. Evaluation of the impact of various agricultural practices on nitrate leaching under the root zone of potato and sugar beet using the STICS soil–crop model. *Sci. Total Environ.* **2008**, *394*, 207–221. [CrossRef]
75. Montoya, F.; Camargo, D.; Domínguez, A.; Ortega, J.F.; Córcoles, J.I. Parametrization of Cropsyst model for the simulation of a potato crop in a Mediterranean environment. *Agric. Water Manag.* **2018**, *203*, 297–310. [CrossRef]
76. Alva, A.K.; Marcos, J.; Stockle, C.; Reddy, V.R.; Timlin, D. A Crop Simulation Model for Predicting Yield and Fate of Nitrogen in Irrigated Potato Rotation Cropping System. *J. Crop Improv.* **2010**, *24*, 142–152. [CrossRef]
77. Lenz-Wiedemann, V.I.S.; Klar, C.W.; Schneider, K. Development and test of a crop growth model for application within a Global Change decision support system. *Ecol. Model.* **2010**, *221*, 314–329. [CrossRef]
78. Khan, M.S.; Yin, X.; van der Putten, P.E.L.; Struik, P.C. An ecophysiological model analysis of yield differences within a set of contrasting cultivars and an F1 segregating population of potato (*Solanum tuberosum* L.) grown under diverse environments. *Ecol. Model.* **2014**, *290*, 146–154. [CrossRef]
79. Parent, B.; Tardieu, F. Temperature responses of developmental processes have not been affected by breeding in different ecological areas for 17 crop species. *New Phytol.* **2012**, *194*, 760–774. [CrossRef] [PubMed]
80. El-Kenawy, E.-S.M.; Alhussan, A.A.; Khodadadi, N.; Mirjalili, S.; Eid, M.M. Predicting Potato Crop Yield with Machine Learning and Deep Learning for Sustainable Agriculture. *Potato Res.* **2024**. [CrossRef]
81. Elavarasan, D.; Vincent, P.M.D. Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications. *IEEE Access* **2020**, *8*, 86886–86901. [CrossRef]
82. Durai, S.K.S.; Shamili, M.D. Smart farming using Machine Learning and Deep Learning techniques. *Decis. Anal. J.* **2022**, *3*, 100041. [CrossRef]
83. Cedric, L.S.; Adoni, W.Y.H.; Aworka, R.; Zoueu, J.T.; Mutombo, F.K.; Krichen, M.; Kimpolo, C.L.M. Crops yield prediction based on machine learning models: Case of West African countries. *Smart Agric. Technol.* **2022**, *2*, 100049. [CrossRef]
84. Rashid, M.; Bari, B.S.; Yusup, Y.; Kamaruddin, M.A.; Khan, N. A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction. *IEEE Access* **2021**, *9*, 63406–63439. [CrossRef]
85. Meghraoui, K.; Sebari, I.; Pilz, J.; Ait El Kadi, K.; Bensiali, S. Applied Deep Learning-Based Crop Yield Prediction: A Systematic Analysis of Current Developments and Potential Challenges. *Technologies* **2024**, *12*, 43. [CrossRef]
86. Kulyal, M.; Saxena, P. Machine Learning approaches for Crop Yield Prediction: A Review. In Proceedings of the 2022 7th International Conference on Computing, Communication and Security (ICCCS), Seoul, Republic of Korea, 3–5 November 2022; pp. 1–7.
87. Padhiary, M.; Saha, D.; Kumar, R.; Sethi, L.N.; Kumar, A. Enhancing precision agriculture: A comprehensive review of machine learning and AI vision applications in all-terrain vehicle for farm automation. *Smart Agric. Technol.* **2024**, *8*, 100483. [CrossRef]
88. Jajja, A.I.; Abbas, A.; Khattak, H.A.; Niedbała, G.; Khalid, A.; Rauf, H.T.; Kujawa, S. Compact Convolutional Transformer (CCT)-Based Approach for Whitefly Attack Detection in Cotton Crops. *Agriculture* **2022**, *12*, 1529. [CrossRef]
89. Siddique, A.; Sun, J.; Hou, K.J.; Vai, M.I.; Pun, S.H.; Iqbal, M.A. SpikoPoniC: A Low-Cost Spiking Neuromorphic Computer for Smart Aquaponics. *Agriculture* **2023**, *13*, 2057. [CrossRef]
90. Tanaka, Y.; Watanabe, T.; Katsura, K.; Tsujimoto, Y.; Takai, T.; Tanaka, T.; Kawamura, K.; Saito, H.; Homma, K.; Mairoua, S.; et al. Deep learning-based estimation of rice yield using RGB image. *Res. Sq.* 2021; *preprint*.

91. Mikołajczyk, M.; Kwasigroch, A.; Grochowski, M. Scene analysis using YOLO deep neural networks. *Zesz. Nauk. Wydz. Elektrotechniki I Autom. PG* **2019**, *68*.
92. Samatha, B.; Kumar, G.R.; Rao, P.A.; Rao, D.K.; Syamsundararao, T.; Karyemsetty, N. IoT-Enabled Potato Diseases Prediction using Deep Learning. In Proceedings of the 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, 14–16 March 2023; pp. 878–883.
93. Huang, Q.; Heuvelink, G.B.M.; He, P.; Leenaars, J.G.B.; Schut, A.G.T. Combining production ecology principles with random forest to model potato yield in China. *Field Crop. Res.* **2024**, *319*, 109619. [CrossRef]
94. Tatsumi, K.; Usami, T. Plant-level prediction of potato yield using machine learning and unmanned aerial vehicle (UAV) multispectral imagery. *Discov. Appl. Sci.* **2024**, *6*, 649. [CrossRef]
95. Kuradusenge, M.; Hitimana, E.; Hanyurwimfura, D.; Rukundo, P.; Mtonga, K.; Mukasine, A.; Uwitonze, C.; Ngabonziza, J.; Uwamahoro, A. Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize. *Agriculture* **2023**, *13*, 225. [CrossRef]
96. Li, D.; Miao, Y.; Gupta, S.K.; Rosen, C.J.; Yuan, F.; Wang, C.; Wang, L.; Huang, Y. Improving Potato Yield Prediction by Combining Cultivar Information and UAV Remote Sensing Data Using Machine Learning. *Remote Sens.* **2021**, *13*, 3322. [CrossRef]
97. Mimenbayeva, A.; Issakova, G.; Tanykpayeva, B.; Tursumbayeva, A.; Suleimenova, R.; Tulkibaev, A. Applying machine learning for analysis and forecasting of agricultural crop yields. *Sci. J. Astana IT Univ.* **2024**, *17*, 28–42. [CrossRef]
98. Lin, Y.; Li, S.; Duan, S.; Ye, Y.; Li, B.; Li, G.; Lyv, D.; Jin, L.; Bian, C.; Liu, J. Methodological evolution of potato yield prediction: A comprehensive review. *Front. Plant Sci.* **2023**, *14*, 1214006. [CrossRef]
99. Piekutowska, M.; Niedbała, G.; Piskier, T.; Lenartowicz, T.; Pilarski, K.; Wojciechowski, T.; Pilarska, A.A.; Czechowska-Kosacka, A. The Application of Multiple Linear Regression and Artificial Neural Network Models for Yield Prediction of Very Early Potato Cultivars before Harvest. *Agronomy* **2021**, *11*, 885. [CrossRef]
100. Abrougui, K.; Gabsi, K.; Mercatoris, B.; Khemis, C.; Amami, R.; Chehaibi, S. Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR). *Soil Tillage Res.* **2019**, *190*, 202–208. [CrossRef]
101. Niedbała, G.; Kurasiak-Popowska, D.; Stuper-Szablewska, K.; Nawracała, J. Application of Artificial Neural Networks to Analyze the Concentration of Ferulic Acid, Deoxynivalenol, and Nivalenol in Winter Wheat Grain. *Agriculture* **2020**, *10*, 127. [CrossRef]
102. Niedbała, G.; Piekutowska, M.; Weres, J.; Korzeniewicz, R.; Witaszek, K.; Adamski, M.; Pilarski, K.; Czechowska-Kosacka, A.; Krysztofiak-Kaniewska, A. Application of Artificial Neural Networks for Yield Modeling of Winter Rapeseed Based on Combined Quantitative and Qualitative Data. *Agronomy* **2019**, *9*, 781. [CrossRef]
103. Lei, X.; Xu, X.; Zhou, S. Potato Yield Prediction Research Based on Improved Artificial Neural Networks Using Whale Optimization Algorithm. *Potato Res.* **2024**. [CrossRef]
104. Al-Hamed, S.A.; Wahby, M.F. Prediction of potato yield based on energy inputs using artificial neural networks and c-sharp under Saudi Arabia Conditions. *Biosci. Biotechnol. Res. Asia* **2016**, *13*, 631–644. [CrossRef]
105. Fortin, J.G.; Anctil, F.; Parent, L.-É.; Bolinder, M.A. A neural network experiment on the site-specific simulation of potato tuber growth in Eastern Canada. *Comput. Electron. Agric.* **2010**, *73*, 126–132. [CrossRef]
106. Nassif, A.B.; Elnagar, A.; Shahin, I.; Henno, S. Deep learning for Arabic subjective sentiment analysis: Challenges and research opportunities. *Appl. Soft Comput.* **2021**, *98*, 106836. [CrossRef]
107. Sunitha, G.; Pushpalatha, M.N.; Parkavi, A.; Boyapati, P.; Walia, R.; Kohar, R.; Qureshi, K. Modeling of Chaotic Political Optimizer for Crop Yield Prediction. *Intell. Autom. Soft Comput.* **2022**, *34*, 423–437. [CrossRef]
108. Khaki, S.; Wang, L. Crop yield prediction using deep neural networks. *Front. Plant Sci.* **2019**, *10*, 621. [CrossRef]
109. Khaki, S.; Wang, L.; Archontoulis, S.V. A CNN-RNN Framework for Crop Yield Prediction. *Front. Plant Sci.* **2020**, *10*, 1750. [CrossRef]
110. Wang, L.; Wang, G.; Yang, S.; Liu, Y.; Yang, X.; Feng, B.; Sun, W.; Li, H. Research on improved YOLOv8n based potato seedling detection in UAV remote sensing images. *Front. Plant Sci.* **2024**, *15*, 1387350. [CrossRef]
111. Ayalew, A.T.; Lohani, T.K. Prediction of Crop Yield by Support Vector Machine Coupled with Deep Learning Algorithm Procedures in Lower Kulfo Watershed of Ethiopia. *J. Eng.* **2023**, *2023*, 1–8. [CrossRef]
112. Dang, C.; Liu, Y.; Yue, H.; Qian, J.; Zhu, R. Autumn Crop Yield Prediction using Data-Driven Approaches:- Support Vector Machines, Random Forest, and Deep Neural Network Methods. *Can. J. Remote Sens.* **2021**, *47*, 162–181. [CrossRef]
113. Shammi, S.; Sohel, F.; Diepeveen, D.; Zander, S.; Jones, M.G.K. A survey of image-based computational learning techniques for frost detection in plants. *Inf. Process. Agric.* **2023**, *10*, 164–191. [CrossRef]
114. Lee, W.S.; Alchanatis, V.; Yang, C.; Hirafuji, M.; Moshou, D.; Li, C. Sensing technologies for precision specialty crop production. *Comput. Electron. Agric.* **2010**, *74*, 2–33. [CrossRef]
115. Xie, J.; Xiao, X.; Xu, Y.; Jin, B. An IoT Assisted Early Warning System for Smart Grid. *J. Phys. Conf. Ser.* **2022**, *2218*, 012027. [CrossRef]

116. Salvador, P.; Gómez, D.; Sanz, J.; Casanova, J.L. Estimation of Potato Yield Using Satellite Data at a Municipal Level: A Machine Learning Approach. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 343. [CrossRef]
117. Singha, C.; Swain, K.C. Rice and Potato Yield Prediction Using Artificial Intelligence Techniques. *Internet Things Anal. Agric.* **2022**, *3*, 185–199.
118. Niedbała, G.; Kurek, J.; Świderski, B.; Wojciechowski, T.; Antoniuk, I.; Bobran, K. Prediction of Blueberry (*Vaccinium corymbosum* L.) Yield Based on Artificial Intelligence Methods. *Agriculture* **2022**, *12*, 2089. [CrossRef]
119. Centorame, L.; Gasperini, T.; Ilari, A.; Del Gatto, A.; Foppa Pedretti, E. An Overview of Machine Learning Applications on Plant Phenotyping, with a Focus on Sunflower. *Agronomy* **2024**, *14*, 719. [CrossRef]
120. Pantazi, X.E.; Moshou, D.; Bochtis, D. Utilization of multisensors and data fusion in precision agriculture. In *Intelligent Data Mining and Fusion Systems in Agriculture*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 103–173.
121. Araújo, S.O.; Peres, R.S.; Ramalho, J.C.; Lidon, F.; Barata, J. Machine Learning Applications in Agriculture: Current Trends, Challenges, and Future Perspectives. *Agronomy* **2023**, *13*, 2976. [CrossRef]
122. Hanadé Houmma, I.; El Mansouri, L.; Gadal, S.; Garba, M.; Hadria, R. Modelling agricultural drought: A review of latest advances in big data technologies. *Geomat. Nat. Hazards Risk* **2022**, *13*, 2737–2776. [CrossRef]
123. Ebrahimy, H.; Wang, Y.; Zhang, Z. Utilization of synthetic minority oversampling technique for improving potato yield prediction using remote sensing data and machine learning algorithms with small sample size of yield data. *ISPRS J. Photogramm. Remote Sens.* **2023**, *201*, 12–25. [CrossRef]
124. Mukiibi, A.; Franke, A.C.; Steyn, J.M. Determination of Crop Coefficients and Evapotranspiration of Potato in a Semi-Arid Climate Using Canopy State Variables and Satellite-Based NDVI. *Remote Sens.* **2023**, *15*, 4579. [CrossRef]
125. Villa, P.M.; Sarmiento, L.; Rada, F.J.; Machado, D.; Rodrigues, A.C. Leaf area index of potato (*Solanum tuberosum* L.) crop under three nitrogen fertilization treatments. *Agron. Colomb.* **2017**, *35*, 171–175. [CrossRef]
126. Ribarics, P. Big Data and its impact on agriculture. *Ecocycles* **2016**, *2*, 33–34. [CrossRef]
127. Santana, L.H.Z. Data Integration for Precision Agriculture—Challenges and Opportunities for the Database community. In Proceedings of the Anais da XVII Escola Regional de Banco de Dados (ERBD 2022), Lages, Brazil, 29 June 2022; Sociedade Brasileira de Computação—SBC: Sao Paulo, Brazil, 2022; pp. 123–126.
128. Heilemann, J.; Klassert, C.; Samaniego, L.; Thober, S.; Marx, A.; Boeing, F.; Klauer, B.; Gawel, E. Projecting impacts of extreme weather events on crop yields using LASSO regression. *Weather Clim. Extrem.* **2024**, *46*, 100738. [CrossRef]
129. Niedbała, G.; Kurasiak-Popowska, D.; Piekutowska, M.; Wojciechowski, T.; Kwiatek, M.; Nawracała, J. Application of Artificial Neural Network Sensitivity Analysis to Identify Key Determinants of Harvesting Date and Yield of Soybean (*Glycine max* [L.] Merrill) Cultivar Augusta. *Agriculture* **2022**, *12*, 754. [CrossRef]
130. Alazzai, W.K.; Obaid, M.K.; Abood, B.S.Z.; Jasim, L. Smart Agriculture Solutions: Harnessing AI and IoT for Crop Management. *E3S Web Conf.* **2024**, *477*, 00057. [CrossRef]
131. Gilpin, L.H.; Bau, D.; Yuan, B.Z.; Bajwa, A.; Specter, M.; Kagal, L. Explaining Explanations: An Overview of Interpretability of Machine Learning. In Proceedings of the 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy, 1–3 October 2018; pp. 80–89.
132. Chamorro-Padial, J.; García, R.; Gil, R. A systematic review of open data in agriculture. *Comput. Electron. Agric.* **2024**, *219*, 108775. [CrossRef]
133. Rozenstein, O.; Cohen, Y.; Alchanatis, V.; Behrendt, K.; Bonfil, D.J.; Eshel, G.; Harari, A.; Harris, W.E.; Klapp, I.; Laor, Y.; et al. Data-driven agriculture and sustainable farming: Friends or foes? *Precis. Agric.* **2024**, *25*, 520–531. [CrossRef]