IJSES

International Journal of Scientific Engineering and Science

Volume 10, Issue 9, pp. 17-24, September 2026. *ISSN (Online): 2456-7361*

Review of Data-Driven Approaches for Wheat Yield Prediction Under Climate Changing Scenarios

Khushmanpreet Kaur[1], Anantdeep Kaur[2], Priyanka Jarial[3]

[1]Research Scholar, Department of Computer Science and Engineering, Punjabi University Patiala

[2]Assistant Professor, Department of Computer Science and Engineering, Punjabi University Patiala

[3]Assistant Professor, Department of Computer Science and Engineering, Punjabi University Patiala

Abstract— *Accurate forecasting of wheat yield is of crucial importance for agricultural planning and food security, particularly in whe at intensive areas like Punjab, India, where wheat yield is linked to weather fluctuations. The uncertainty of the yield has increased since the irregularities in the temperature, precipitation, humidity, and other atmospheric factors. In response, there has been increasing interest in u sing artificial-intelligence-based methods, which have been shown to be effective in capturing the nonlinear relationships between weather and crop output. This review provides a comparison of recent studies on wheat yield prediction based on machine learning (ML), deep learning (DL), and rule-based crop-simulation models in various climatic conditions. It reviews common ML algorithms that were used in the past such as random forest, support vector regression, decision tree, KNN, LASSO, and XGBoost and compares them with DL architectures such as MLP, DNN, R NN, GRU, and LSTM and then reviews well-established simulation platforms such as DSSAT, CERES-Wheat, APSIM, AquaCrop, and WOFOST modeling crop growth and soil-plant-atmosphere dynamics. Results show that the ensemble models such as Random Forest and XGBoost are very effective on the structured data sets as they are robust to handling the nonlinearity and mixed data input types, while having not so good results on han dling temporal patterns. Recurrent architectures, particularly LSTM and GRU, are able to make better use of sequential meteorologi cal data and typically produce the best accuracy, with GRU being the closest with less computational expense. Rule-based simulation models are still useful as they are physiologically grounded and are easy to understand, but require calibration and region -specific parameters. There is an overall trend that data-driven ML and DL methods are superior to simulation-based methods in terms of performance, especially in the cases of large and complex datasets in which the patterns are learned directly from the da ta. This review summarises current research trends, highlights the pros and cons of each family of models, and emphasises the need for accurate, interpretable, and scalable prediction models t o enable climate-resilient wheat production and evidence-based agricultural decision making.*

Keywords— *Climate Change: Deep Learning: LSTM: Machine Learning: Meteorological Parameters: Precision Agriculture: Random Forest: Time-Series Analysis: Wheat Yield Prediction.*

I. INTRODUCTION

Wheat is one of the most important staple crops in the world, essential for food security for a significant proportion of the world's population and nowhere is this truer than in India. Wheat cultivation and consumption is a major crop in Punjab, a state where farming is the economic mainstay of most families. However, its production is increasingly threatened by the variable climatic conditions that affect wheat growth, such as temperature fluctuations, irregular rainfall, humidity and length of sunshine in the past few years that have become more unpredictable, making yield forecasting difficult and farm level planning difficult (Mausam et al.). This background points to the need for accurate wheat-yield prediction. Wheat goes through six growth stages: Germination, Tillering, Jointing, Flowering, Grain Filling and Harvest and has a different set of weather needs for each. Moderate warmth (20-25°C), light rain and sufficient but not surplus moisture are most favorable for germination, while moderate-to-high sunlight and avoidance of waterlogged conditions are optimum during tillering, and avoidance of water stress and high temperatures are preferred during jointing (18-22°C) and flowering (20-23°C), where the latter is the most critical stage as heavy rain during this period can severely affect the crop; moderate sunlight and moisture levels are optimum during grain filling (20-25°C); and warm and dry conditions with no rain at harvest (25-30°C) is optimal.

Fig. 1.1 Wheat Growing Stages

Different algorithms of machine learning, deep learning or rule-based (Random Forest, Decision Trees, LSTM, GRU, MLP, DSSAT, CERES-Wheat, etc.) that was evaluated in relation to meteorological drivers (temperature, rainfall, humidity, and sunshine hours) under changing climate for the purposes of this review. Previously, classical statistical approaches were more commonly used for the yield forecast, but were less suited to capturing the non-linear relationships between weather and yield. This changed with the advent of machine learning (ML) and deep learning (DL), which were able to capture this complexity and enhance accuracy. For nonlinear problems, algorithms like Random Forest, Support Vector Regression (SVR) and XGBoost performed well and deep architectures like MLP, GRU, LSTM and Bi-LSTM took it one step further by exploiting the temporal structure in sequential data, an area where conventional ML often struggles. Until the advent of widespread use of ML and DL, rule-based crop-simulation platforms were primarily used, including

International Journal of Scientific Engineering and Science
Volume 10, Issue 9, pp. 17-24, September 2026. ISSN (Online): 2456-7361

DSSAT, CERES-Wheat, APSIM, AquaCrop and WOFOST, which were developed with agronomic and physiological rules that represent interactions between soil, phenology, irrigation, fertilization and climate. The simulators have good interpretability and long-term climate impact information, but require intensive calibration, expert input and region-specific data, thus restricting their scalability. The recent push towards the research of ML and DL methods that can learn complex relationships directly from large datasets has only come as a result of such constraints. This paper summarizes the most recent (2012–16) wheat yield prediction studies, reviews and compares the accuracy, advantages and disadvantages of ML, DL, and rule-based approaches, and identifies areas of research need and contrasts results presented in the reviewed literature. The review does not aim to create a new predictive model, but rather makes conclusions based on the work already done in order to determine the most appropriate approach for yield estimation under different environmental conditions. The review also aims to identify models that are most adaptable to climate resilient agriculture, based on their ability to cope with temporal variability, and identifies gaps, such as the lack of low-cost, weather-only models, and the lack of comprehensive comparisons between deep-learning models. The results are meant to help researchers and policy makers select appropriate predictive tools for sustainable farm management, so that farmers will be able to make decisions about crop investment based on advance yield estimates.

II. LITERATURE REVIEW

2.1 Studies combining Machine Learning and Deep Learning algorithms for wheat yield prediction

Author & Year	Title	Data Used	Models Used	Key Findings	Limitations
Raza et al. (2025)	Improving Wheat Yield Prediction with Multi-Source Remote Sensing Data and Machine Learning in Arid Regions	Remote sensing + weather	RF, Gradient Boosting, DTs, BTs	Fusing multiple remote-sensing and weather sources boosted accuracy	Computationally intensive
Attwal et al. (2025)	Integrated machine learning model for wheat yield prediction using agronomic and meteorological factors: A case study from Punjab, India	Agronomic + weather data	RF, J48 (Decision Tree), SVM, MLP	Combining agronomic and meteorological inputs raised model accuracy	Needs large, diverse data inputs
Aravind et al. (2025)	Development of multistage crop yield estimation model using machine learning and deep learning techniques	Multi-stage crop data	ANN, DNN, RF, SVR	DL techniques enhanced early-season yield forecasts	Limited interpretability
Sarowar et al. (2025)	Crop yield prediction using machine learning: an extensive and systematic literature review	Meteorological data, vegetation indices, yield data	RF, SVM, DT, LR, LSTM, CNN, DNN, RNN	LSTM and CNN surpassed conventional ML approaches	Limited data volume; vegetation indices rely on costly satellite feeds
Wang et al. (2024)	Progress in research on deep learning based crop yield prediction	Weather data, soil quality, crop varieties, environmental factors	RF, DT, LR, DNN, CNN, LSTM, RNN	DNN and LSTM outperformed ML techniques in accuracy	Demands extensive data and computing power; sensitive to climate variability
Iqbal et al. (2024)	Analysis of wheat yield prediction using machine learning models under climate change scenarios	Climate projections + weather	RF, MLR, ANN, Ensemble models	ANN models were most accurate under climate-change projections	Uncertainty in future climate projections
Nishu et al. (2021)	Deep learning based wheat crop yield prediction model in Punjab region of North India	Climate variables (temp, rainfall)	ANN, RF, MLR, LSTM, RNN	LSTM and RF delivered the strongest results	Region-specific; DL models are computationally demanding

2.2 Studies using only Machine Learning algorithms for wheat yield prediction

Author & Year	Title	Data Used	Models Used	Key Findings	Limitations
Mausam et al. (2024)	Weather based wheat yield prediction using machine learning	Weather data (Punjab districts)	SVR, LASSO, ML models	Models relying solely on weather variables reached strong accuracy	No deep-learning comparison included

Author & Year	Title	Data Used	Models Used	Key Findings	Limitations
Morales and Villalobos (2023)	Using machine learning for crop yield prediction in the past or the future	Weather, crop simulation and wheat datasets	ANN, ML models	ML improved yield estimates across shifting climate conditions	Heavy reliance on historical records; limited regional transfer
Jin et al. (2023)	Developing machine learning models for wheat yield prediction using ground-based data, satellite-based ET and vegetation indices	Meteorological + satellite data	Random Forest, XGBoost	XGBoost edged out Random Forest marginally	Complexity rises when merging multiple data sources
Bansal et al. (2023)	Winter Wheat Crop Yield Prediction on Multiple Heterogeneous Datasets using Machine Learning	Weather and soil datasets	RF, CatBoost, XGBoost	Merging varied datasets boosted predictive performance	Cross-region performance varies due to data heterogeneity
Ashfaq et al. (2024)	Accurate wheat yield prediction using machine learning and climate-NDVI data fusion	Climate + NDVI data	RF, XGBoost	XGBoost and RF produced the top accuracy scores	Reliant on satellite availability

2.3 Studies using only Deep Learning algorithms for wheat yield prediction

Author & Year	Title	Data Used	Models Used	Key Findings	Limitations
Joshi et al. (2025)	An explainable Bi-LSTM model for winter wheat yield prediction	Meteorological data, crop growth data	Bi-LSTM (Deep Learning)	Bi-LSTM captured temporal patterns well; explainability methods added transparency	Computationally costly; needs sizable datasets
Babbar et al. (2024)	Prediction of Wheat Yield by Novel SDC-LSTM Framework	Weather and historical yield data	SDC-LSTM	SDC-LSTM raised temporal-forecast accuracy	Needs large training sets and substantial compute
Srivastava et al. (2022)	Winter wheat yield prediction using Convolutional neural networks from environmental and phenological data	Weather, soil, and crop phenology data	CNN, DNN	CNN surpassed baseline ML approaches	Requires large historical datasets and heavy computation
Abdalraheem (2024)	A Hybrid CNN-LSTM Framework for Spatio-Temporal Crop Yield Prediction Integrating Satellite Imagery and IoT Sensor Data	Satellite imagery + IoT sensor data	CNN-LSTM	Merging spatial and temporal signals improved outcomes	High compute demand; extensive image preprocessing
Zhang et al. (2025)	Wheat Yield Prediction Based on Parallel CNN-LSTM-Attention with Transfer Learning Model	Climate and wheat yield datasets	CNN-LSTM-Attention	Attention mechanisms sharpened feature extraction and forecasting	Added complexity lengthens training

2.4 Studies using Rule-based crop-simulation approaches for wheat yield prediction

Author & Year	Title	Data Used	Models Used	Key Findings	Limitations
Agrawal et al. (2025)	Calibration and validation of CERES-Wheat and APSIM models for different wheat varieties in Raipur district of Chhattisgarh	Wheat variety, sowing date, weather and field data	CERES-Wheat, APSIM	Both simulators reproduced growth and yield accurately across sowing dates	Needs calibration and localized inputs
Kheir et al. (2025)	Hybridization of process-based models, remote sensing, and machine learning for enhanced spatial predictions of wheat yield and quality	DSSAT outputs, remote sensing, climate and field datasets	DSSAT (CERES, CROPSIM, NWheat)	Hybrid process-based/ML/remote-sensing approach sharpened yield and nutrient estimates	High complexity; depends on multiple data streams
Lal and Niwas (2024)	Yield Prediction by DSSAT Model of Wheat Crop: A Review	Irrigation, nitrogen, sowing date and climate datasets	DSSAT	DSSAT reliably simulated grain yield and growth under variable climate	Needs precise calibration and validation

Author & Year	Title	Data Used	Models Used	Key Findings	Limitations
Al-Shammari et al. (2024)	Incorporation of mechanistic model outputs as features for data-driven models for yield prediction: a case study on wheat and chickpea	Crop simulation outputs, weather and yield datasets	Mechanistic crop models	Feeding mechanistic outputs into data-driven models strengthened robustness	Integrating simulation and ML frameworks is complex
Nouri and Veysi (2025)	Mitigating crop modeling uncertainties through machine learning in drylands	Meteorological and dryland wheat datasets	CSM-CERES-Wheat	CERES-Wheat captured yield variability well in dryland settings	Sensitive to inaccuracies in weather inputs

2.5 Datasets used across the reviewed studies

Dataset Type	Data Included	Purpose in Yield Prediction	Representative Studies
Meteorological Data	Temperature, rainfall, humidity, sunshine hours, wind speed, solar radiation	Links climate variability to yield outcomes	Mausam et al., Iqbal et al., Nishu et al., Agrawal et al., Nouri and Veysi
Remote Sensing Data	NDVI, ARVI, satellite imagery	Tracks crop health and vegetation to sharpen spatial forecasts	Raza et al., Ashfaq et al., Jin et al., Kheir et al.
Agronomic Data	Soil type, fertilizer use, irrigation practices	Refines accuracy using field-level conditions	Attwal et al., Godara et al.
Historical Yield Data	Previous years' wheat yield records	Reveals long-term yield trends; supports calibration of simulation models	Joshi et al., Babbar et al., DSSAT and CERES-Wheat studies
Phenological Data	Crop growth stages and development data	Links growth-stage timing to yield variation	Srivastava et al., Joshi et al.
Climate Projection Data	Future climate scenarios and weather projections	Forecasts yield under prospective climate conditions	Iqbal et al., Morales and Villalobos, Nouri and Veysi, DSSAT-based studies
Soil Data	Soil texture, moisture, nitrogen content, fertility	Models soil-crop interaction and nutrient availability	APSIM, DSSAT-based studies
Crop Management Data	Sowing date, irrigation, fertilizer application	Assesses how management practices shape yield	Lal and Niwas, Agrawal et al.
Crop Simulation Outputs	Growth stages, biomass, evapotranspiration	Strengthens robustness and climate-impact assessment	Kheir et al., Al-Shammari et al.

III. ALGORITHMS USED FOR YIELD PREDICTION IN THE REVIEWED STUDIES

Based on the above-mentioned studies, the performance of different modelling families can be compared in terms of accuracy, robustness and climatic variability suitability. The table below is a summary of the most commonly reported machine learning and deep learning methods, as well as the most commonly reported rule-based methods.

3.1 Machine Learning Algorithms

Model	Performance	Strengths	Weakness	Best Use Case
Random Forest	High	Robust; handles nonlinearity well	Weaker at capturing temporal patterns	Heterogeneous agricultural data
SVR	Moderate	Effective on small datasets	Sensitive to parameter tuning	Small, nonlinear datasets
LASSO	Low-Moderate	Simple and interpretable	Cannot represent complex patterns	Feature selection
Decision Tree	Moderate	Easy to interpret	Prone to overfitting	Interpretable predictions
KNN	Moderate	Simple to implement	Computationally expensive at scale	Similarity-based prediction
XGBoost	High	High accuracy; handles complex data	Requires careful tuning	Complex nonlinear data
CatBoost	High	Handles categorical and heterogeneous data well	Longer training time	Categorical agricultural data
Bagging Trees	High	Reduces variance; improves stability	Computationally expensive	Ensemble-based prediction
MLR	Moderate	Simple and interpretable	Cannot capture nonlinear relationships	Linear yield prediction
Gradient Boosting	Very High	High accuracy on complex data	Requires parameter tuning	Nonlinear agricultural datasets

International Journal of Scientific Engineering and Science
Volume 10, Issue 9, pp. 17-24, September 2026. ISSN (Online): 2456-7361

3.2 Deep Learning Algorithms

Model	Performance	Strengths	Weakness	Best Use Case
MLP	High	Captures nonlinear patterns	Needs large training data	General nonlinear modelling
GRU	Very High	Efficient time-series modelling	Slightly less expressive than LSTM	Efficient time-series prediction
LSTM	Very High	Strong at long-term dependencies	High computational cost	Long-term time-series forecasting
DNN	Very High	Learns highly complex patterns	Risk of overfitting; high compute cost	Large, complex datasets
CNN	Very High	Excellent for spatial and satellite-image analysis	Needs large datasets and GPU resources	Satellite and spatial data
Bi-LSTM	Very High	Learns forward and backward temporal patterns	More compute-intensive than LSTM	Bidirectional sequence learning
CNN-LSTM	Very High	Combines spatial and temporal feature learning	Complex architecture	Spatio-temporal prediction
SDC-LSTM	Very High	Strong temporal sequence learning	Requires large sequential datasets	Sequential climate data
CNN-LSTM-Attention	Extremely High	Attention mechanism sharpens feature extraction	Very high computational complexity	High-accuracy forecasting
ANN	High	Captures nonlinear climatic variables	Needs sufficient training data	Nonlinear yield prediction
RNN	High	Captures sequential patterns	Vanishing-gradient problem	Basic time-series prediction

3.3 Rule-Based Crop Simulation Models

Model	Performance	Strengths	Weakness	Best Use Case
DSSAT	High	Simulates crop growth under varying climate conditions	Needs extensive input data and calibration	Climate-impact assessment
CERES-Wheat	High	Effective for wheat growth and yield simulation	Requires region-specific calibration	Weather-based yield prediction
APSIM	Very High	Models crop-soil-climate interactions effectively	Complex model configuration	Long-term agricultural planning
AquaCrop	Moderate-High	Efficient water-productivity simulation	Less effective for complex climatic interactions	Drought and irrigation analysis
WOFOST	High	Strong phenology and crop-growth simulation	Requires detailed environmental inputs	Crop-growth monitoring under climate variability

3.4 Comparative Performance of Algorithms Across Reviewed Studies

IV. Discussion

In the literature analysed, climate variables have proved to be a key influence in yield-prediction results. As stated by Iqbal et al. (2024), projected yields have a much stronger association with changes in weather patterns, such as the effect of warming, drought and increased temperatures, on wheat production, which is especially sensitive in South Asia, and a comparatively lesser influence from rainfall. Ensemble ML methods have emerged as the most promising methods for yield forecasting in the context of climate change and have been found to outperform single-model methods, which capture nonlinear and complex relationships between climate and yield. The study reveals that the grain filling stage is the most critical period for temperature and water factor impact on the crop, and also confirms that using ML in conjunction with remote sensing is more beneficial for accurate crop monitoring, in improving farm-management practices and enhancing food security in arid regions under climatic stress, with RF-ARVI model accuracy outperforming the accuracy of other models tested, and that the Remote Sensing Index, Atmospherically Resistant Vegetation

IJSES

International Journal of Scientific Engineering and Science
Volume 10, Issue 9, pp. 17-24, September 2026. ISSN (Online): 2456-7361

Index (ARVI), is most beneficial in arid areas with dust and haze, with the RF-ARVI model producing the highest overall accuracy. Attwal et al. (2025) found that introducing meteorological and agronomic information improves prediction accuracy, with a Block-wise Average Yield (BAY) model using climate data, and a Yield Class (YC) model using agronomic data, while the combination of both into a Final Yield Prediction (FYP) model yielded the better results, with J48 (Decision Tree) and Random Forest being the best models. Aravind et al. (2025) state that stage-specific, weather-indexed approaches are crucial for climate-resilient forecasting: RF and SVR are ideally suited for nonlinear climatic conditions, whereas DL models require a large amount of data to be effective, but struggle with limited data. Weather data, farming techniques, and fertilizer usage, when used in a meaningful way, can sensibly improve yield, but the additional information can reduce the gain, emphasising the importance of balancing the data with a climate-conscious approach – the best model in this instance was a 1-D CNN, according to Godara et al., 2025. Similarly, Sarowar et al. (2025) highlight the need to integrate environmental and agricultural aspects, point out that while Linear Regression is extensively used for its simplicity, it is not always the most effective model, and mention that Random Forest, SVM, Gradient Boosting, KNN, and Decision Trees are well-performing models, while CNN and LSTM often give better results than traditional ML models. The authors of Joshi et al. (2025) demonstrate the strong influence of vegetation indices, temperature and precipitation data on wheat yield and the efficacy of their Bi-LSTM model, superior in terms of R^2 and lower MAE, compared to LSTM, 1-D CNN and Random Forest, attributing this to its capability to capture both historical and future temporal relationships. They also used explainable-AI (XAI) techniques to render wheat forecasts more understandable. The models of Mausam et al. (2024) implicitly include the effects of climate through sensitive variables like temperature, relative humidity, evapotranspiration, growing degree days (GDD), photo-thermal index (PTI), and heat-use efficiency (HUE), underscoring the importance of climate variability, heat stress, and changing water requirements. Wang et al. (2024) point out that models trained on past data can lose their accuracy over time due to the shifting climate context, and suggest models that are 'adaptive' by consuming new climate data, as both ML and DL models. The yield is found as a key parameter sensitive to water availability, as climate change alters the rainfall pattern, droughts and floods; adding remote-sensing data helps overcome the dependency on meteorological inputs only, as satellite images capture the actual situation of crops which can fill the gaps in meteorological data. Adding remote-sensing data helps overcome the dependency on meteorological inputs only, as satellite images capture the actual situation of crops which can fill the gaps in meteorological data, and yield is found as a key parameter sensitive to water availability, which is reshaped by climate change through its changes in rainfall pattern, droughts and floods. As revealed in Nishu et al. (2021), more advanced models are needed in a changing climate, as multivariate linear regression is unable to account for non-linear weather yield relationships, while RNN and LSTM are well suited to learning from the history to model time dependency. When satellite and ground data are paired, Jin et al. (2023) conclude that XGBoost and Random Forest are effective. Bansal et al. (2022) demonstrate how heterogeneous weather and soil information enhances the robustness of the models. Morales and Villalobos (2023) validate that ML can be used to predict yield in future climate scenarios. The authors of Babbar et al. (2024) show that the SDC-LSTM architecture improves prediction accuracy and is capable of learning long-range relationships in the weather and yield data. CNN-LSTM models are able to integrate spatiotemporal climate information, as demonstrated by Abdulraheem (2024). By combining CNN-LSTM with attention mechanisms and transfer learning, Zhang et al. (2025) further enhance the accuracy. Hybrid use of environmental and phenological data results in CNN-based models outperforming conventional ones, according to Srivastava et al. (2022). In the rule-based literature, the added value of crop-simulation approaches is also emphasized: Agrawal et al. (2025) demonstrate that CERES-Wheat and APSIM adequately simulate the growth and yield of wheat under various sowing conditions; Kheir et al. (2025) demonstrate that combining the DSSAT outputs with ML and remote sensing enhances robustness and spatial accuracy; Lal and Niwas (2024) show that DSSAT is useful for assessing irrigation, fertilization, and climatic effects on productivity; and Nouri and Veysi (2025) show that DSSAT based crop-simulation models are effective in capturing the variability in dryland production under uncertain weather. These kinds of simulation models have high interpretability and provide some insight into the effects of climate, but they are not as flexible as current ML/ DL models because of the requirement to heavily calibrate the models and the inputs that are specific to the domain.

V. CONCLUSION

All the literature reviewed highlights the major role of climate variability, particularly temperatures, on wheat yield predictions. Yield is significantly impacted by rising temperatures and heat stress at the time of grain filling, and more frequent and severe droughts, especially in South Asia, with rainfall also being important but comparatively less. Ensemble ML and advanced DL methods outperform conventional statistical methods in all cases from a modelling perspective. Iqbal et al. show that ensemble models are good at capturing the nonlinear interaction between climate and yield, and Joshi et al. make use of the Bi-LSTM and LSTM for yield time-series forecasting, as they are known for capturing temporal dependencies. Typical methods like Multivariate Linear Regression are widely used due to their simplicity (Sarowar et al.), but they may fail to account for the nonlinear relationship between climate and crops and hence provide large errors. Random Forest, on the other hand, is found to be one of the most reliable ML models in studies (Aravind et al., Mausam et al., Wang et al.) particularly in heterogeneous data. Another theme is that combining multiple data streams, such as combining meteorological variables (temperature, rainfall, humidity, sunshine), remote-sensing based indices (like ARVI developed by Sarowar et al., Raza et al., Ashfaq et al.) and agronomic data (soil type and fertilizer application), generally increases accuracy. The combination of climate and satellite data is found to provide significant improvements in the accuracy (Wang et al.), and the application of XAI methods (Joshi et al.) enhances transparency and trust in model outputs. Despite this, performance is still heavily dependent on data availability and quality: DL models like ANN and DNN work best with large datasets, but their performance is not as strong in small datasets; ML models like Random Forest and SVR perform better in small datasets than DL models, but DL models with sequence-based architectures outperform them in capturing long-range temporal dependencies, such as those in ANN. All the ML models, such as SVR, Random Forest, XGBoost, CatBoost, and LASSO models, performed well in the modelling of the nonlinear relationship between climate and yield, and it was observed that the ensembles of models and heterogeneous datasets improved their accuracy, as reported by Mausam et al. (2024), Ashfaq et al. (2024), Jin et al. (2023), and Bansal et al. (2022). When large climatic and satellite datasets were available, DL approaches such as LSTM, Bi-LSTM, CNN-LSTM and attention-based approaches further captured temporal and sequential structure as demonstrated by Joshi et al. (2025), Babbar et al. (2024), Abdalraheem (2024), Zhang et al. (2025), and Srivastava et al. (2022). While good for interpretability and physiologically based simulation, rule-based simulators like DSSAT, CERES-Wheat, APSIM, AquaCrop and WOFOST can be less efficient for large, non-linear, sequential datasets, and are limited to a set of rules, requiring extensive calibration. Some recent studies indicate that a combination of rule-based and AI based models can result in more powerful and climate resilient forecasting systems. All of the reviewed studies include one or more climate-sensitive meteorological variables, although not all explicitly focus on climate-change scenarios, and the studies remain relevant to climate-driven yield prediction. Finally, there is no single best model; the model that is best for a particular application depends on the type of data, the variability of the climate, and the application context. This section presents future directions and research gaps.

VI. Research Gaps and Future Directions

There several gaps, such as few studies are based on multi-source and expensive input data like NDVI, satellite imagery, and agronomic data, while studying the computational efficiency of LSTM and GRU are relatively limited, and simple, scalable, low-cost DL models using only weather data are not sufficiently developed. This review concentrates on literature published from 2021–2025 and primarily uses ML and DL techniques with a particular emphasis on datasets, climatic contexts, and evaluation metrics, all of which make direct comparisons across studies difficult. Some avenues are worth pursuing in the wake of these gaps. For future, firstly if the models were robust, low cost, and scalable, they would expand their use by those areas that do not have high-resolution satellite or NDVI data. Second, hybrid models with both nonlinear-relationship modelling and time-series capability, such as combining Random Forest or XGBoost with LSTM or GRU, could boost the accuracy and robustness of the model under changing climatic conditions. Third, there would be a great benefit in having current, often updated, weather data feeds so that forecasts would be timelier and more useful to farmers and policy makers for use in decision-support systems. Further research on the efficiency and scalability of DL systems with large agricultural data sets is also required. Research further should explore the integration of crop simulators with rule-based approaches and ML and DL models that integrate the advantages of accuracy with the interpretability of DSSAT, CERES-Wheat and APSIM for the purpose of providing improved understanding of crop-climate interactions, and develop ways to reduce the calibration needs of the crop simulation models and to enhance their adaptability across regions. Lastly, the need for model transparency, using XAI methods, would enable trust to be established with farmers and stakeholders to facilitate acceptance of these tools.

References

[1] Aravind, K.S., Vashisth, A., Krishnan, P. et al. Development of multistage crop yield estimation model using machine learning and deep learning techniques. Int J Biometeorol 69, 499–515 (2025). https://doi.org/10.1007/s00484-024-02829-9

[2] Nishu Bali & Anshu Singla (2021) Deep Learning Based Wheat Crop Yield Prediction Model in Punjab Region of North India, Applied Artificial Intelligence, 35:15, 1304-1328, DOI: 10.1080/08839514.2021.1976091

[3] "Weather based wheat yield prediction using machine learning", MAUSAM, vol. 75, no. 3, pp. 639–648, Jul. 2024, doi: 10.54302/mausam.v75i3.5606.

[4] Iqbal, N.; Shahzad, M.U.; Sherif, E.-S.M.; Tariq, M.U.; Rashid, J.; Le, T.-V.; Ghani, A. Analysis of Wheat-Yield Prediction Using Machine Learning Models under Climate Change Scenarios. Sustainability 2024, 16, 6976. https://doi.org/10.3390/su16166976

[5] Wang, Y.; Zhang, Q.; Yu, F.; Zhang, N.; Zhang, X.; Li, Y.; Wang, M.; Zhang, J. Progress in Research on Deep Learning-Based Crop Yield Prediction. Agronomy 2024, 14, 2264. https://doi.org/10.3390/agronomy14102264

[6] Sarowar Morshed Shawon, Falguny Barua Ema, Asura Khanom Mahi, Fahima Lokman Niha, H.T. Zubair, Crop yield prediction using machine learning: An extensive and systematic literature review. Smart Agricultural Technology, Volume 10, 2025, 100718, ISSN 2772-3755, https://doi.org/10.1016/j.atech.2024.100718.

[7] Raza, A.; Shahid, M.A.; Zaman, M.; Miao, Y.; Huang, Y.; Safdar, M.; Maqbool, S.; Muhammad, N.E. Improving Wheat Yield Prediction with Multi-Source Remote Sensing Data and Machine Learning in Arid Regions. Remote Sens. 2025, 17, 774. https://doi.org/10.3390/rs17050774

[8] Joshi A, Pradhan B, Chakraborty S, Varatharajoo R, Alamri A, Gite S and Lee C-W (2025) An explainable Bi-LSTM model for winter wheat yield prediction. Front. Plant Sci. 15:1491493. doi: 10.3389/fpls.2024.1491493

[9] M. Ashfaq, I. Khan, A. Alzahrani, M. U. Tariq, H. Khan and A. Ghani, "Accurate Wheat Yield Prediction Using Machine Learning and Climate-NDVI Data Fusion," in IEEE Access, vol. 12, pp. 40947-40961, 2024, doi: 10.1109/ACCESS.2024.3376735.

[10] R. Attwal, "Integrated agronomic and weather-based crop yield prediction using machine learning," Journal of Agricultural Science, vol. 17, no. 1, 2025.

[11] Godara, S., Batra, K., Bana, R.S. et al. AgriFact framework for modelling the impact of farmers' information demand on nationwide wheat productivity in India. Sci Rep 15, 33621 (2025). https://doi.org/10.1038/s41598-025-19133-0

[12] Maximilian Zachow, Stella Ofori-Ampofo, Harald Kunstmann, Rıdvan Salih Kuzu, Xiao Xiang Zhu, Senthold Asseng, Wheat yield forecasts with seasonal climate models and long short-term memory networks, Computers and Electronics in Agriculture, Volume 239, Part B, 2025, 110965, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2025.110965.

[13] Peipei Chen, Yue Li, Xiaojun Liu, Yongchao Tian, Yan Zhu, Weixing Cao, Qiang Cao, Improving yield prediction based on spatio-temporal

International Journal of Scientific Engineering and Science
Volume 10, Issue 9, pp. 17-24, September 2026. ISSN (Online): 2456-7361

deep learning approaches for winter wheat: A case study in Jiangsu Province, China, Computers and Electronics in Agriculture, Volume 213, 2023, 108201, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2023.108201.
[14] Nabila Chergui, Durum wheat yield forecasting using machine learning, Artificial Intelligence in Agriculture, Volume 6, 2022, Pages 156-166, ISSN 2589-7217, https://doi.org/10.1016/j.aiia.2022.09.003.
[15] Ashfaq, M., Khan, I., Shah, D. et al. Predicting wheat yield using deep learning and multi-source environmental data. Sci Rep 15, 26446 (2025). https://doi.org/10.1038/s41598-025-11780-7
[16] Lou Z and Sun D (2025) Yield estimation of winter wheat in the Huang-Huai-Hai region using MODIS and meteorological data: spatio-temporal analysis and county-level modeling. Front. Plant Sci. 16:1721972. doi: 10.3389/fpls.2025.1721972
[17] Bijlwan, A., Ranjan, R., Singh, M. et al. Predicting crop disease severity using real time weather variability through machine learning algorithms. Sci Rep 15, 34767 (2025). https://doi.org/10.1038/s41598-025-18613-7
[18] Muhammad Haseeb, Zainab Tahir, Syed Amer Mahmood, Aqil Tariq, Winter wheat yield prediction using linear and nonlinear machine learning algorithms based on climatological and remote sensing data, Information Processing in Agriculture, Volume 12, Issue 4, 2025, Pages 431-444, ISSN 2214-3173, https://doi.org/10.1016/j.inpa.2025.02.004.
[19] Cheng E, Zhang B, Peng D, Zhong L, Yu L, Liu Y, Xiao C, Li C, Li X, Chen Y, Ye H, Wang H, Yu R, Hu J and Yang S (2022) Wheat yield estimation using remote sensing data based on machine learning approaches. Front. Plant Sci. 13:1090970. doi: 10.3389/fpls.2022.1090970
[20] Muhammed Cem Akcapınar, Halit Apaydin, Early yield estimation in wheat using open access global datasets and artificial intelligence, Computers and Electronics in Agriculture, Volume 237, Part C, 2025, 110486, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2025.110486.
[21] Morales A and Villalobos FJ (2023) Using machine learning for crop yield prediction in the past or the future. Front. Plant Sci. 14:1128388. doi: 10.3389/fpls.2023.1128388
[22] Mojtaba Naghdyzadegan Jahromi, Shahrokh Zand-Parsa, Fatemeh Razzaghi, Sajad Jamshidi, Shohreh Didari, Ali Doosthosseini, Hamid Reza Pourghasemi, Developing machine learning models for wheat yield prediction using ground-based data, satellite-based actual evapotranspiration and vegetation indices, European Journal of Agronomy, Volume 146, 2023, 126820, ISSN 1161-0301, https://doi.org/10.1016/j.eja.2023.126820.
[23] Y. Bansal, D. Lillis, and T. Kechadi, "Winter Wheat Crop Yield Prediction on Multiple Heterogeneous Datasets using Machine Learning," in 2022 International Conference on Computational Science and Computational Intelligence (CSCI), Dec. 2022, pp. 206–212, doi: 10.1109/CSCI58124.2022.00040.
[24] N. Babbar, A. Kumar, and V. K. Verma, "Prediction of Wheat Yield by Novel SDC-LSTM Framework," Brazilian Archives of Biology and Technology, vol. 67, 2024, doi: 10.1590/1678-4324-2024230773.
[25] D. A. Abdalraheem, "A Hybrid CNN-LSTM Framework for Spatio-Temporal Crop Yield Prediction Integrating Satellite Imagery and IoT Sensor Data," International Journal of Advances in Signal and Image Sciences, 2024, doi: 10.29284/bdeyv897.
[26] A. K. Srivastava, N. Safaei, S. Khaki, G. Lopez, W. Zeng, F. Ewert, T. Gaiser, and J. Rahimi, "Winter wheat yield prediction using convolutional neural networks from environmental and phenological data," Scientific Reports, vol. 12, no. 1, p. 3215, 2022, doi: 10.1038/s41598-022-06249-w.
[27] Song, C.; Liu, T.; Ning, W.; Xu, T.; Song, S.; Li, Z.; Ouyang, S.; Song, X.; Han, T.; Zhang, Z.; et al. Wheat Yield Prediction Based on Parallel CNN-LSTM-Attention with Transfer Learning Model. Agriculture 2025, 15, 2519. https://doi.org/10.3390/agriculture15232519
[28] Nickey Agrawal, Harsh Vardhan Puranik, Gopi Krishna Das. Calibration and validation of CERES-wheat and APSIM models for different wheat varieties in Raipur district of Chhattisgarh. Int J Res Agron 2025;8(2):191-200. DOI: 10.33545/2618060X.2025.v8.i2c.2573
[29] Ahmed M.S. Kheir, Ajit Govind, Vinay Nangia, Maher A. El-Maghraby, Abdelrazek Elnashar, Mukhtar Ahmed, Hesham Aboelsoud, Rania Gamal, Til Feike, Hybridization of process-based models, remote sensing, and machine learning for enhanced spatial predictions of wheat yield and quality, Computers and Electronics in Agriculture, Volume 234, 2025, 110317, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2025.110317.
[30] Lal, Devi, and Ram Niwas. 2024. "Yield Predication by DSSAT Model of Wheat Crop: A Review". International Journal of Environment and Climate Change 14 (2):519-24. https://doi.org/10.9734/ijecc/2024/v14i23965.
[31] Al-Shammari, D., Chen, Y., Wimalathunge, N.S. et al. Incorporation of mechanistic model outputs as features for data-driven models for yield prediction: a case study on wheat and chickpea. Precision Agric 25, 2531–2553 (2024). https://doi.org/10.1007/s11119-024-10184-3
[32] Nouri, M., Veysi, S. Mitigating crop modeling uncertainties through machine learning in drylands. Sci Rep 15, 42663 (2025). https://doi.org/10.1038/s41598-025-26811-6