Journal of Artificial Intelligence and Metaheuristic

♦ ISSN: 2833-5597 Vol: 11, No. 02, 2026 ♦ PP. 94–114

Article DOI: https://doi.org/10.54216/JAIM.110207

FULL LENGTH ARTICLE

Global Crop Yields Forecasting Using Informer Architecture Optimized with Football-Inspired Metaheuristic

Toufik Mzili[1,*]

[1] Faculty of Sciences, Chouaib Doukkali University, El Jadida, Morocco

Email: mzili.t@ucd.ac.ma

Received: January 03, 2026 Revised: March 01, 2026 Accepted: May 04, 2026 ⋆ Corresponding author

ABSTRACT

Accurate forecasting of agricultural production is essential for managing food supply chains, guiding policy decisions, and building resilience against climate variability. However, modeling long-range, country-level crop production remains challenging due to temporal complexity, nonlinear dependencies, and the need for highly generalizable prediction systems. This study addresses these challenges by developing a hybrid forecasting framework that combines deep learning architectures with metaheuristic hyperparameter optimization. A global dataset spanning tomato and potato production from 1961 to 2021 was used to evaluate multiple forecasting models, including Informer, N-BEATS, LogTrans, N-HITS, EALSTM, TST, and LSTM. The Informer model achieved the best baseline performance (RMSE = 0.0799; NSE = 0.9070) and was selected for optimization. A comparative analysis was conducted using several metaheuristic algorithms, with particular focus on the Football Optimization Algorithm (FbOA), a novel strategy inspired by cooperative team dynamics. FbOA delivered the highest gains across all evaluated metrics, reducing RMSE to 0.00109, MSE to 1.19×10^{-6}, and increasing NSE to 0.9260, R^2 to 0.9600, and the correlation coefficient r to 0.9550. These results confirmed that metaheuristic tuning substantially enhances the forecasting capability of deep models, particularly when guided by domain-inspired search logic. The proposed framework demonstrates strong potential for integration into real-time, AI-driven agricultural decision support systems, offering scalable solutions for food security planning, climate-smart agriculture, and long-term sustainability forecasting.

Keywords: Crop Yield Prediction ▪ Neural Ordinary Differential Equation (NODE) ▪ Metaheuristic Optimization ▪ Football Optimization Algorithm (FbOA) ▪ Large Language Models (LLMs)

1. INTRODUCTION

The predictive agricultural production is at the center of the world sustainability, economic stability, and food security initiatives. In a world where climatic volatility and population increase are becoming the new norm and resource competition is a reality, then skills in accurate prediction of crop yields are no longer optional or recommended but increasingly mandatory. Among the myriad of crops cultivated worldwide, *Solanum tuberosum* (potato) and *Solanum lycopersicum* (tomato) are of exceptional strategic importance[1, 2]. Potatoes form a staple of caloric diet in both normal and developing economies, and tomatoes are both a market product and a processed product traded on the world market. Their economic, nutritional, and agronomic interdependence makes them a perfect choice for an item of a high-resolution, data-driven forecasting system[3, 4].

With the collection by decades of structural agricultural data by organizations such as the Food and Agriculture Organization (FAO), there has been a wealth of new opportunities

created in the context of analytical innovation[5]. The data used in the work spans the period from 1961 to 2021, covering the national production of tomatoes and potatoes across various geopolitical and agro-ecological regions of the country[6]. It has categorical descriptions, including country, item, element options, year, unit, and value time variables. This rich repository of history is indicative of the variety and heterogeneity of agriculture in the world - both can be used to the advantage of ML capabilities and it is also a problem because it is multidimensional, inconsistent, and lacks the capacity to generalize it.

In this regard, machine learning has been found to have a revolutionary time-series modeling approach in the agricultural field[7]. In contradiction to more traditional statistical methods that tend to use restricted linear formulations, ML algorithms are designed to learn non-linear trends, cross-temporal relationships, and infer to new high-dimensional feature domains. Learned models like recurrent neural networks (RNNs), transformers with attention mechanisms, and even feedforward forecasting models have all shown excellent results in modeling the rich dynamics and multiscale structure of agricultural data sources [8, 9]. Nevertheless, the ML models have powerful algorithms, yet are hard to make insensitive to their hyperparameters. Hyperparameters like the learning rate, window size, hidden layer size, dropout rate and batch size can have quite a significant impact on the performance, stability and convergence characteristics. Consequently, hyperparameter optimization is not a marginal issue, but rather one of the aspects that fundamentally define the efficiency of the model][10].

To make this task even harder is the fact that the data is very dimensional. Agricultural data that are normally present usually have redundant or loosely informative features, Multicollinearity and unequal coverage in time, which in sum create noise in the process of learning. Moreover, the categorical heterogeneity of aspects like country and item of crop may further complicate the issues of encoding, normalization, and alignment. Therefore, preprocessing is not just a pre-process but a very vital step that determines future success of model training. This paper is thus a new approach to preprocessing informed by the DeepSeek-V3 large language model (LLM). The LLM was leveraged to design a preprocessing pipeline that includes encoding schemes, imputation methods, and scaling techniques tailored to the structural characteristics of the dataset, thereby enhancing model readiness and reducing noise propagation[11, 12].

Alongside preprocessing and model selection there is the task of hyperparameter optimization by means of metaheuristic algorithms. Metaheuristics are population-based adaptive algorithms that are based on natural, biological or sociocultural evolution and processes. They have the advantage of being able to search non-convex and high-dimensional search spaces without any gradient information or continuity assumptions. This paper will apply a variety of metaheuristic algorithms to the hyperparameters of the chosen forecasting model to systematize the optimization. These are not limited to Football Optimization Algorithm (FbOA), Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Biogeography-Based Optimizer (BBO), Whale Optimization Algorithm (WOA) and many more-there are some different in exploring-exploitation trade-offs and convergence behavior.

However, as far as the recent developments have occurred, there are still several significant gaps in the literature. Current agricultural forecasting methods seldom combine the best of both in one system that combines the strengths of both LLM-directed preprocessing and a metaheuristically based hyperparameter optimization. Besides, the state-of-the-art deep learning architectures have not been benchmarked on long-range, high-resolution agricultural data, especially in the context of decades of national production data. Not many studies compare both the classical and the modern time-series architecture systematically employing a homogeneous preprocessing pipeline, and they do not thoroughly evaluate the benefits of meticulously optimizing the metaheuristics.

Based on this loopholes, the following research objectives guide this study:

- to construct a reliable and reproducible data pipeline for agricultural production forecasting by leveraging large-scale FAO datasets.
- to benchmark a suite of state-of-the-art ML models for their efficacy in long-range crop yield prediction.
- to implement a novel preprocessing protocol using the DeepSeek-V3 large language model.
- to optimize the hyperparameters of the best-performing model through the deployment of diverse metaheuristic algorithms, thereby enhancing accuracy, convergence, and generalization.

Four works in total are the main contributions of this work. To start with, it proposes a solid and scalable preprocessing architecture that incorporates the recommendation of LLM into the agricultural modeling process. Second, it provides a comparative performance account of some of the best ML architectures used on global data of tomatoes and potatoes production. Third, it presents a multi-purpose metaheuristic optimization layer, which reviews a range of algorithms in terms of their capabilities to tune deep learning models. Fourth, it gives a theoretical and technical roadmap of how preprocessing, modeling and optimization can be incorporated into a unified forecasting system, which can be generalized to other crops, fields or areas with limited modification.

2. LITERATURE REVIEW

Integration of recent computational methods, satellite observation, and machine learning (ML) has led to increased ability to forecast and predict agricultural production, especially of essential crops including wheat, potato, tomato, and rice. Computational. In recent years, the importance of diverse approaches such as unmanned aerial systems (UAS) or deep learning architecture in contributing to better decision-making in contemporary farming operations has been proven.

Application of UAS in acquiring images has proved helpful in determining the survival of crops and necrosis. As an example, multi-band imaging combined with UAS allowed the rapid assessment of winter wheat survival and potato necrosis, where high correlations with ground-truth were obtained (up to $r = 0.93$). These findings help to emphasize the future

feasibility of UAS in delivering quantitative, timely, spatially accurate data to help decision-making at the application level[13].

LSTM models have been a popular method of deliberately predicting rice yield. Another study involving ICAR-NRRI, Odisha also recorded low error scores (RMSE = 0.21), high accuracy values (F1 = 0.984), and demonstrated that the use of deep learning models can indeed be effective for yield forecasting In a similar manner, artificial intelligence (AI) techniques such as artificial neural networks (ANNs) and deep neural networks (DNNs) applied to forecasting rice and potato yields in West Bengal, India. Despite the QTLs not being mapped with the same accuracy across environments, the optimized DNN obtained better accuracy (R^2 = 0.98 in rice and 0.97 in potato), which implies the robustness of the predictive performance across different environments[2, 14].

The remote sensing methods and GIS methods are also becoming popular. Vegetation indices based on Land sat and Sentinel-2 images were used to predict potato tuber yields in Saudi Arabia. The models yielded prediction errors ranging from 3.8% to 13.5%, with coefficients of determination (R^2) between 0.39 and 0.65, demonstrating that satellite-based indices can capture field productivity variations and support site-specific agricultural management [15]. Complementary, recent research applied neural networks, histogram-based gradient boosting, and support vector machines to potato selection using yield, size, appearance, and processing-quality traits. The non-linear models consistently outperformed logistic regression across robustness evaluations [16].

The use of Sentinel-2 data to predict weather using satellites also enhanced the results when feature selection was added. For example, Regression Quantile Lasso and Leap Backwards models achieved RMSE values as low as 10.94% and R^2 up to 0.93, while random forest (RF) models successfully predicted potato yields months prior to harvest in Spain [17]. Similarly, a Potato Productivity Index (PPI), when combined with Sentinel-2 spectral bands, enhanced yield prediction accuracy ($R^2 = 0.77$, RMSE = 15.42%), surpassing conventional vegetation indices such as NDVI [18]. Spectral imaging by UAV has found application in predicting the yield of potato crops particularly in scenarios that involve the integration of cultivar-specific data. Incorporating cultivar traits into RF and support vector regression (SVR) models improved prediction accuracy from $R^2 = 0.48$ to $R^2 = 0.79$, underscoring the value of genetic and management-related data in precision agriculture [19].

The use of the optimization techniques has become an effective instrument in predicting agriculture. An example of this is the Balance Dynamic Biruni Earth Radius Optimization Algorithm (BDBER) coupled with LSTM was used to optimize potato production forecasts, with an exceptionally low RMSE of 0.008999, and better than traditional ML methods did[20]. Similarly, to optimize the ARIMA-based models to predict the potato crop, the Waterwheel Plant Algorithm (WWPA) was employed to enhance the accuracy of the prediction (RMSE = 0.0001) significantly [21]). Comparative analysis of ARIMA and ETS model on potato production in China, India and USA indicated that the ETS was more efficient than the ARIMA especially in long term national predictions to forecast production in China at 100, 417 thousand tonnes in 2027 [22].

Tomato-focused studies reveal parallel advances. Machine learning models, including ANN, RF, and decision trees (DT), have been applied to predict tomato yields under different irrigation regimes. RF achieved the highest accuracy ($R^2 =$ 0.96, RMSE = 3.5 t/ha), demonstrating the ability of ML to provide actionable insights under water-limited conditions [3]. Tomato market forecasting has also been studied through ML-based models such as LSTM and ARIMA, which leverage historical pricing, seasonal trends, and weather conditions to better inform supply-chain stakeholders [4].

The deep learning has provided considerable potential in the prediction of yield among crops. In potato and tomato yield prediction in Portugal, bidirectional LSTM models made with a control group of 4 years had an average of R^2 = 0.97 to 0.99 and performed better than either convolutional neural networks (CNN) or classical ML[23]. Moreover, crop simulation models have been modified to recreate climate change conditions using crop simulation models, e.g. SUBSTOR-Potato. Projections at the global scale suggested that yield losses of up to 26 percent might occur by 2085 when the RCP 8.5 scenario is considered, as adaptive management strategies were badly needed at that time, as well[24].

To be able to offer a brief but sufficiently detailed summary of the reviewed works, a summary table has been created (Table 1). The table brings into the limelight the area of focus, the methodology used and the key works or results of each of the works referred to. This comparative viewpoint makes possible the improved comprehension of the methodological multiplicity and the comparative merits of various methodologies.

3. MATERIALS AND METHODS

3.1 Dataset Description

The data used in the present paper is based on the publicly accessible archives of the Food and Agriculture Organization (FAO) of the United Nations that have one of the most extensive longitudinal data sets of the world agricultural output. The chosen data set includes national-level production data of two strategically important crops- tomatoes and potatoes over a continuous time period between the year 1961 and 2021. This 61-year intervals provide a unique historical perspective, the metamorphosis of agricultural activities, technological interventions, climatic disturbance and change in policy in various geopolitical and agro ecological settings.

The dataset consists of records which represent agricultural production entries with multi-attributes structure. The main properties are: Area, a property that represents the geographical or national area of production; Country, a nominal property that indicates the country of the FAO country coding system; Item, which indicates the type of crop type, such as tomato or potato; Element, which is a property that shows the nature of the reported element of agriculture (e.g. production quantity, harvested area); Year, a property that represents the time period; Unit, a property that represents the measurement unit of the reported production quantity (e.g. metric tons); and Value, a property that represents the numeric value of

Temporal quality of the dataset, its geographic and categorical variety makes it very realistic in terms of the real-world

Table 1. Comparative summary of literature reviewed

Reference	Focus Area	Methodology	Key Findings and Contributions
[13]	Wheat survival & potato necrosis	UAS-based multi-band imagery	Strong correlation with field data ($r = 0.60$–0.93); rapid assessment tool.
[25]	Rice yield prediction	LSTM model (Odisha NRRI data)	High accuracy (F1=0.984); RMSE = 0.21 for test data.
[14]	Rice & potato yields (India)	ANN, DNN, SVM models	DNN outperformed others; $R^2 = 0.97$–0.98, RMSE ≤ 0.95 t/ha.
[15]	Potato yield (Saudi Arabia)	Landsat-8 & Sentinel-2 indices (NDVI, SAVI)	Yield prediction error 3.8–13.5%; R^2 up to 0.65.
[16]	Potato selection and yield traits	Neural network, HGBC, SVM	Non-linear models outperformed logistic regression across robustness evaluations.
[17]	Potato yield (Spain)	Sentinel-2 bands, Regression Quantile Lasso, RF	Feature selection improved accuracy; RF predicted yields 1–2 months ahead.
[20]	Potato production	LSTM optimized by BDBER algorithm	RMSE = 0.00899; superior to traditional ML approaches.
[18]	Potato yield index	Sentinel-2 bands + Potato Productivity Index (PPI)	PPI improved performance ($R^2 = 0.77$, RMSE = 15.42%) vs NDVI.
[3]	Tomato yield (irrigation regimes)	ANN, RF, DT	RF best ($R^2 = 0.96$, RMSE = 3.50 t/ha).
[19]	Potato yield (cultivar effects)	UAV spectral data + cultivar info (RFR, SVR)	Accuracy improved from $R^2 = 0.48$ to 0.79 with cultivar traits.
[21]	Potato forecasting	ARIMA optimized by WWPA	Feature selection via WWPA; RMSE = 0.0001, very high precision.
[22]	Global potato production	ARIMA vs ETS models	ETS consistently better; projected 2027 outputs: China = 100,417 kt, India = 61,882 kt, USA = 18,229 kt.
[4]	Tomato price forecasting	LSTM, ARIMA models	Accurate real-time forecasts; useful for supply-chain optimization.
[23]	Potato & tomato yields	Bidirectional LSTM, GRU, CNN, MLP	Bi-LSTM best ($R^2 = 0.97$–0.99); CNN second-best.
[24]	Climate change impacts	SUBSTOR-Potato simulation	Global yield reductions projected: –2% to –26% by 2085 under RCP 8.5.

agricultural dynamics. It incorporates both the developed and developing countries which differ in their profiles of agronomy, technological capability, as well as climate. Consequently, the dataset represents a diverse data environment which can be best utilized in testing the scalability, versatility, and efficiency of sophisticated machine learning prediction algorithms. The time- series models with long memory dependencies that are based on the longitudinal consistency and the international breadth of the data also have a solid foundation of training and when gauging the performance of generalization across unseen temporal slices as well as regional variations.

Notably, the structure of the dataset requires certain attention to be paid to preprocessing. Combining numeric and categorical characteristics, hierarchical and temporal relationships are challenging in nature. As an example, the field called Element can contain various measures of production explosion that need to be harmonised semantically before modelling. Similarly, the difference in the frequency of national reporting and the presence of non-reported values also require the well-planned types of imputation. Also, since the dataset is consolidated on the annual basis, it adds the opportunity to verify the consistency of time steps of a country with another one, which is critical to construct identical and time-equivalent input sequences used by the machine learning algorithms.

The main attribute to be predicted in this study is the value attribute that regards the quantity of production as a continuous variable. This is aimed at predicting such a production quantity in future years as per the historical trends and pattern derived of the other contextual features. This variable is selected based on its policy applicability and its importance as a leading indicator of agricultural output, potential of trade, and resiliency of food supply.

As a strategy to prepare the dataset to the modeling, a temporally stratified approach to the dataset was applied to divide it into the training, validation, and testing subsets. In particular, previous periods were used to be trained and validated to pass the long-term trends, whereas the last few years were used to undergo final testing. This is to make sure that the forecasting exercise reflects too closely a realistic implementation environment, where it is required to make predictions in the future using only unavailable information. Also, the partitioning design provides the possibility of the temporal generalization analysis and, therefore, eliminates the issue of the data leakage, thereby maintaining the integrity of the predictive analysis.

Overall, the data has heterogeneous and extensive foundation on which one can create, educate and experiment machine learning designs in the agricultural predicting domain. It

is spatially, temporally, and categorically appropriate to be used as an example of the things, which the baseline and optimized modeling pipelines can (and cannot) do in order to quantify the change in global agricultural systems over time. The general workflow that was adopted in this research is as follows. The structured framework of forecasting is shown in 1. potato and tomatoes growth on a global level with hybrid deep learning and optimization techniques. It starts with the data acquisition and preprocessing and contains the processing of missing data. scaling, temporal structuring and feature encoding as well as values. The dataset is then divided into subsets of training and testing to have strong evaluation. Multiple baseline models They are used to train such models as Informer, N-BEATS, LogTrans, N-HITS, EALSTM, TST, and LSTM. comparative analysis. In order to further improve predictive performance, a set of optimization. model tuning is based on application of algorithms, such as the Football Optimization Algorithm (FOA). Lastly, the performance evaluation measures are calculated to determine accuracy, stability, and capability to generalize. The strength of this coordinated workflow is the integration of. deep learning models based on data and metaheuristic optimization methods.

3.2 Data Preprocessing

Data preprocessing is a critical component in any data-centric modeling package, as its effectiveness can determine whether the machine learning models can be used or not. It is also true in the agricultural example given, where vast quantities of super-dimensional, longitudinal data may be quite heterogeneous across dimensions, the missing values may be incongruent, and the values themselves may be scaled in totally different manners, all of which may comprise significant noise or bias in the model-training procedure. The pipeline of data preprocessing was built by and run under the supervision of the DeepSeek-V3 large language model (LLM), which provided strategic guidance based on data topology and semantics. To ensure procedural consistency, context, and structural flexibility within a domain extending well beyond six decades and representing a global distribution of production practices, the choice to adopt a language model in this phase was motivated by these considerations. To reveal the best practices for addressing missing data, numerical variable encoding, and variable scaling, the DeepSeek-V3 model[26], distinguished by its strong general-purpose reasoning performance, has been referred to. The responses were coded in reproducible procedures of processing the data, and implemented on all the data. The treatment of missing values is one significant preprocessing task that has a critical impact, so its significance was taken into account. Some entries had incomplete or null fields of Value due to discrepancies in national reporting or due to lack of records in the past. With the suggestions of the LLM and due to the nature of the temporal variation of the dataset, the forward-fill imputation strategy was adopted. This process uses historical continuity between country-crop pairs to use the most recent non-missing value to project the value into the future. Mean imputation into the missingness at the start of a temporal sequence with the respective crop country average has been adopted. The hybrid approach maintained the temporal patterns and reduced the instances of artificial deformations of the series. Then categorical variables were considered, including: country and item. Since there are many categorical values that are unique and more so, there are dozens of categorical values in the Country field, a direct one-hot encoding strategy was considered computationally inefficient and possibly sparse. Rather, a hybrid of label encoding and embedding-based transformation was taken. The LLM suggested a label encoding step as a transitory encoding, which will be replaced by low-dimensional vector embeddings when training the model, especially in those architectures that can operate on embedded inputs (e.g., transformer-based models). This facilitated the memory performance as well as the semantic continuity between categorical representations. In the case of the numerical properties, including the field of Value and any time-related characteristics (e.g., normalized year index), feature scaling was vital, to prevent the instability of the gradient in the model training. According to the recommendations of the LLM, the min-max normalization was used to the attribute named Value, bringing all the production quantities to the range of [0,1]. It was informed by the fact that deep learning models are placed in the training pipeline, known to be better when the input values are limited to a small numerical space. Min-max scaling did not lose the interpretability of proportional differences across samples, which is fundamentally needed to do downstream comparative analysis. The other step in the preprocessing phase entailed time structuring of the data into a format of supervised learning. Since the dataset initially is in a panel format with separate entries per country and crop in a year, the data was rearranged into a rolling-window format. In every input sequence, the amount of historical production values was a fixed-length window, which was synchronized with the corresponding categorical indicators, and was used to predict the next future time step. The choice of window length was informed by model specific constraints of input and guided by knowledge about the production cycle memory. The rolling-window methodology also established that temporal dependences are preserved, meaning that the machine learning models learn through changing patterns as time goes on. Lastly, statistical tests and data distribution visualization were used to confirm all preprocessing actions to provide that no distortions, scaling effects, or encoding errors were applied. Such checks validated the integrity of the procedures used in constructing the LLM-governing pipeline and its compliance with the structural aspects of the machine learning models used during the forecasting step. This multiphase, interpolating preprocessing approach based on the use of the LLM converted the dataset into a uniform, numerically stable, and semantically consistent format that can be used in the process of training and evaluation of rigorous models. DeepSeek-V3 added a level of logical consistency and expert judgment that would be hard to achieve solely using the heuristics of a human being. The production dynamics of potato and tomato have experienced great fluctuations between countries in the last twenty years. To come into a better perspective of such trends, the production trends of top 10 countries of production are presented in the form of a figure, as shown in the figure below, Figure 2. ten manufacturing nations in 2000-2021. This character gives a comparative context, which brings out how some. Countries, like China and India, show a pronounced increasing trend whereas other countries depict a rather stable picture or are comparatively

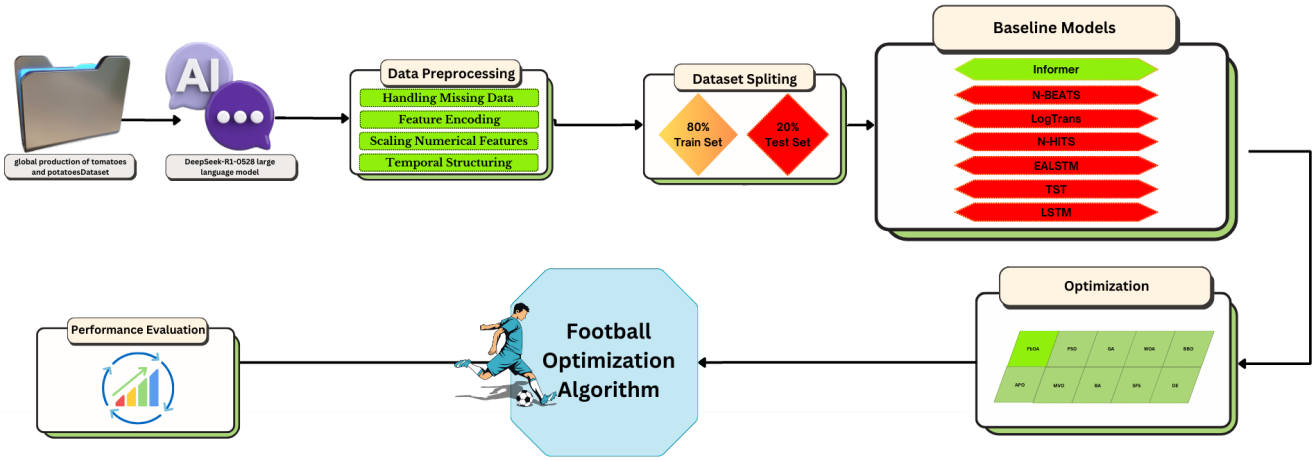

Figure 1. Proposed framework for hybrid deep learning and optimization-based forecasting.

stable. varying production trends. These understandings are necessary to put the world agricultural productivity into perspective, evaluate. regional donations and establish possible causes of these trends.

In order to further examine the association that exists between land use efficiency and productivity, figure Figure 3 shows the yield of the potato with respect to the area that is harvested. in the case of specific high-producing nations. This graph gives a valuable insight. on the proportion of yield to hectare, and of cultivated land, in the various countries. The United Kingdom and Turkiye countries are characterized by high productivity. that with relative concentration of harvested areas, and others such as Uganda and Tanzania. reflect low yields in larger or more diffused land parcels. Another dimension is provided by the bubble sizes, which are proportional to the production values. to determine the relationship between yield and area harvested and the overall output.

In-depth understanding of the dynamics of long-term production of potatoes is presented in. the historical production, is given in Figure 4. trends of four chosen countries (Turkiye, Uganda, the United Kingdom, and Zimbabwe) from 1961 to 2021. This chart is smoothed by the use of a three year rolling average. smaller changes in the short term and pays less attention to smaller structural changes in production. The figure indicates a different trend: Turkiye demonstrates stable growth until Uganda, the beginning of 2000s, reflects steep fluctuations that are associated with agricultural. instability. The United Kingdom on the other hand measures gradual decline. compared to historically high levels, but Zimbabwe puts a lot of emphasis on an enormous decline. After the early 2000s in production. These are contradictory paths which highlight. the non-uniformity of production of potatoes in various areas.

To understand the input of various features that are involved in the predictive modeling process, The SHAP (SHapley Additive exPlanations) is shown in Figure 5. summary plot. This number shows the relative significance of the features of driving. model outputs, and hence providing transparency and interpretability in the predictive framework. Interestingly enough, the most significant variable is the variable RollingMean3, Then there will be `Prev Year Value` and Countryitemean. Other features, such as GrowthRate, DiffFromMean and YearlyRank, are contributory. a smaller but still give useful context in improving the performance of the model. Color scale also shows the impact of high and low values of these features on the. predictions, and this provides a more subtle insight into their interactions.

3.3 Machine Learning Models

The agricultural prognostication exercise developed in this paper requires the application of machine learning (ML) frameworks that are able to capture long-term temporal relationships and control heterogeneous input variables, as well as short-term variations and long-term trends in the structure. Since the dataset is sequential and both the dimensions are categorical and continuous, a collection of advanced ML models was chosen to represent both ends of the state-of-the-art forecasting paradigms, such as recurrent, transformer-based, and feedforward neural network models.

Both models were selected because of their empirical power as claimed in recent time-series literature, as well as the fact they are architecturally appropriate to agricultural production data. Such a variety in the choice of model enables a strict comparative structure to decide which inductive biases and representational strategies most accurately reflect the statistical characteristics of the underlying data.

The following classifiers were used and trained on the standardized dataset with the mentioned preprocessing pipeline:

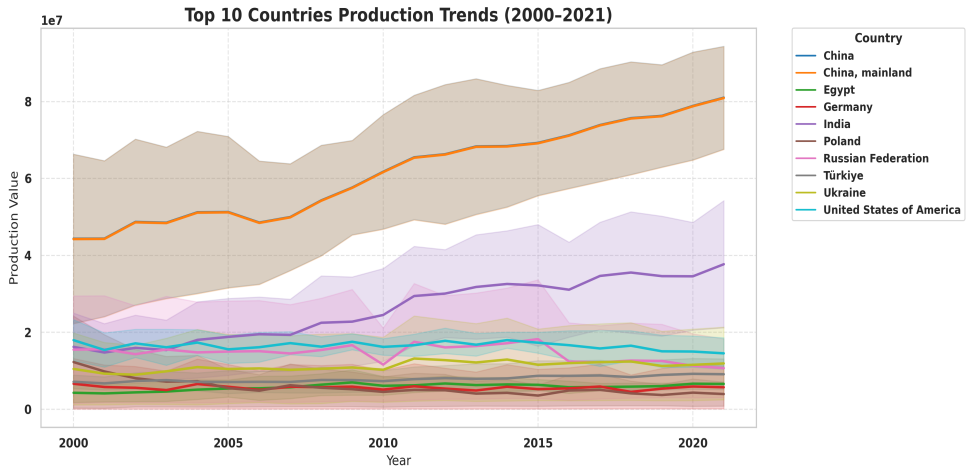

Figure 2. Top 10 countries production trends (2000–2021).

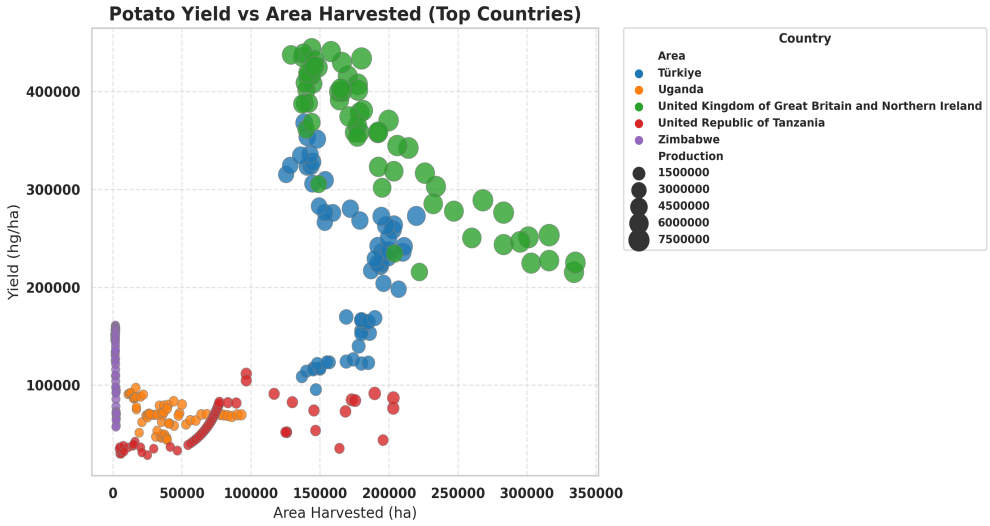

Figure 3. Potato yield vs. area harvested (top countries).

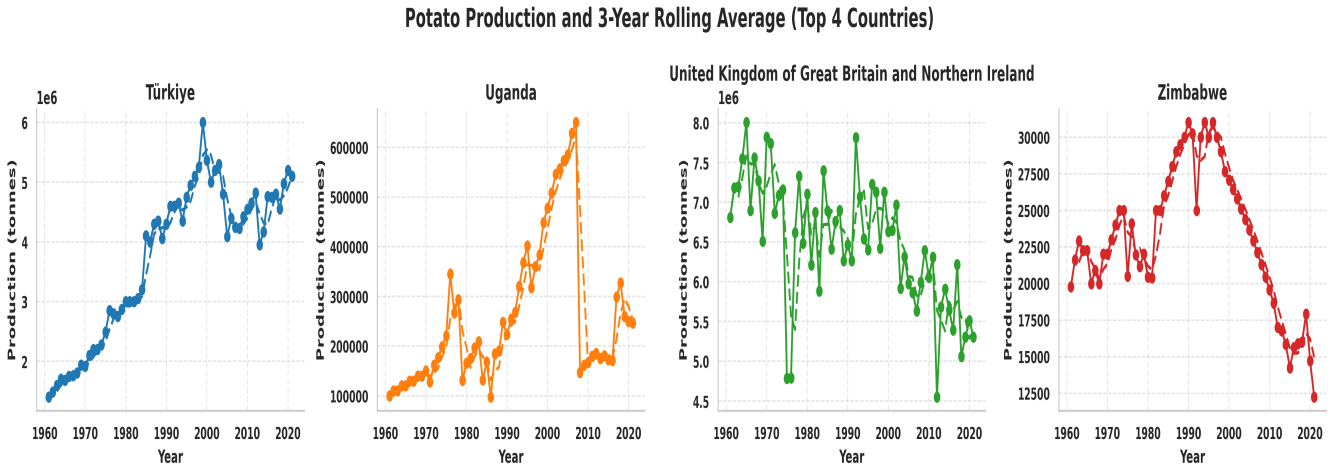

Figure 4. Potato production and three-year rolling average trends for selected top countries.

The Informer model describes a transformer-based approach that was optimized to perform long sequence time-series predictions[27]. It proposes the ProbSparse self-attention, which has a smaller complexity than regular attention, in $\mathcal{O}(L^2)$ to $\mathcal{O}(LlogL)$. This novelty enables the application of the model to cover much larger time spans, particularly in agricultural data with multi-decade spans. Informer is equipped with generative-style decoder stages, as well as self-attention distillation that boosts performance and efficiency.

N-BEATS (Neural Basis Expansion Analysis for Time series): N-BEATS is a deep feed-forward forecasting model that does not follow the traditional assumptions of autoregressive models. It deconstructs both time series into trend and seasonality components through backward residual learning block and

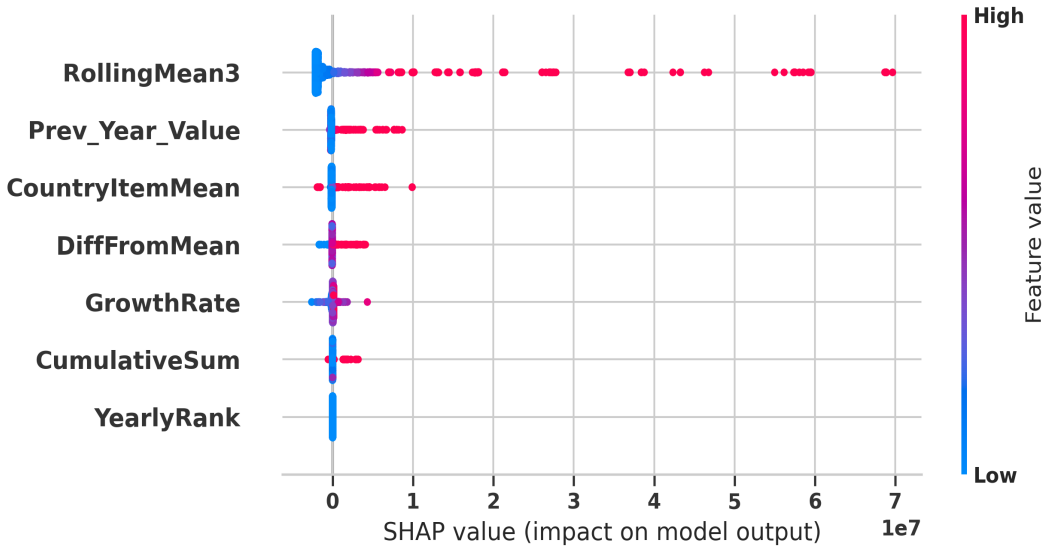

Figure 5. SHAP summary plot showing feature importance for the predictive model.

forward residual learning block[28]. This also enables the extraction of deterministic signals from noisy or irregular data, which is typical of national-scale agricultural reporting. In addition, it is better understood due to its interpretable application in policy-relevant contexts.

It is a variation of standard transformer architecture with a logarithmic attention mechanism to enable focus on short-term relationships without the need to model long-term structure in a computationally costly manner, as well as [29]. Modeling with high-resolution, large-scale time-series data, e.g., multi-country agricultural outputs over time, is a candidate application area because of its training stability and low memory requirement.

N-HITS (Neural Hierarchical Interpolation of Time Series): N-HITS builds on signal decomposition notions to i) multiscale blocks of neural forecasting, and ii) hierarchical interpolation of the temporal dimension[30]. It employs interpolation-based residual learning to preserve patterns across a variety of scales, offering a flexible framework that can fit with either fast or slow-changing and slow-evolving data. In forecasting in agriculture, where production shocks can occur from one year to the next on top of longer run trends, it can be beneficial.

EALSTM: The model combines the use of the encoder and the long short-term memory with attention and context representations, powering them by static context embedding[31]. The architecture enables the model to adjust the weight of each time step on a per-exampleper-example basis based on both sequential data and non-sequentialnon-sequential categorical data,, such as country or crop type. This hybridization enhances the context-sensitivity of predictions, especially within environments having diverse routes of input.

TST (Temporal Fusion Transformer): TST is a model integrated into a sequence-to-sequence learning framework, which combines in a single model differentiated components of attention with variable selection layers and gating functions[32]. It proves to be effective with multivariate time series whose observations are irregular, and thus during training, it will automatically identify relevant variables and time periods. TST has proved to be an attractive option due to its adaptability and has shown good performance over a wide range of settings.

LSTM (Long Short-Term Memory): LSTM is a classic time-series deep learning framework that utilizes a recurrent network architecture with memory cells to consider long-term dependencies[33]. Although there is no parallelism as in transformer-based models, nor the hierarchical innovations of more modern approaches, it is a necessary benchmark because it has been well-documented as robust and generally applicable. Its addition allows for the measurement of possible gains in performance relative to a known standard.

All models were trained using similar preprocessing, temporal windowing, and evaluation protocols, which allowed the fairness and reproducibility of comparative analyses. By covering a wide structural gamut, including recurrent and transformerized, all-fully-connected models, the study will provide a robust basis against which to decide what families of models can best generalize across several decades of agricultural data globally. This also sets up the prerequisite of choosing the most promising candidate for downstream hyperparameter optimization using metaheuristic algorithms.

3.4 Metaheuristic Optimization

Hyperparameter optimization is recognized as one of the most influential determinants of model performance in the context of the deep learning model development. It entails the identification of optimal setups of model-level and training-level parameters that cannot be learned using gradient-based techniques. These parameters include learning rate, dropout rate, embedding dimensions, batch size and sequence length and determine the dynamics of the learning process and its performance out of the training dataset. The search space of such parameters is however generally non-convex and high-dimensional and not necessarily continuous. It is not

computationally feasible or statistically reliable, especially training deep architectures on large-noise time-series (including multi-decadal crop production data) in such spaces, using conventional search schemes like grid search or random sampling.

This research entails the use of population-based and stochastic optimization paradigm, which is based on the metaheuristic algorithms to overcome these limitations. These computational methods are specifically aimed at exploring objective functions that are complex using natural, physical or social simulations. Unlike other deterministic methods, the metaheuristic methods are not constrained by gradient information or absence of convexity and, hence, will tend to perform the global searches on the solution space and narrow down high potential regions with each iteration. Their strength can be characterized by the ability to reach a balance between exploration and exploitation by means of iterative and collaborative learning in groups of potential solutions.

The foundation of the optimization framework proposed in this paper is the Football Optimization Algorithm (FbOA), a new metaheuristic whose dynamic and strategic decision-making behavior is motivated by the dynamics of a competitive football. The model of each candidate solution with FbOA can be observed as an agent, such as a football player, whose location is a point in the hyperparameter set of the search space. These agents behave and imitate the fluid plays of a football team, switching between offense and defense. There are three key actions to describe the agent'ss interaction with the search space: short passin,gwhichich is local refinem;nt; lob pass,iwhichwhicmid-rangerange u;pdate; and through ball p,awhichg, which is global exploration. These metaphorical behaviors are transformed into mathematical operators that drive the agents around the search space by acting with evolving velocities, accelerations and position changes[34].

The FbOA exploration is achieved as velocity-driven dynamics. Agents navigate the search space by leveraging adaptive movement procedures guided by a multi-force update model. Each iteration, an agent's velocity is modified by both external forces and past solutions that have proven advantageous, thereby extending the stochastic exploration of the solution space to a greater expansion. The cosine modulated term in the velocity equation is used to scale the extent of the search at different points in time; a large-scale search in initial steps and gradually narrowing down as convergence is approached. The velocity update formula is developed as:

$$V_n = F_{\max}\Big(b_x a_i (F_{\text{ext}} - F_{\min}) + r b_y a_j (F_{\text{best}} - F_{\min}) \cos\Big(\frac{\pi}{\text{Iteration}}\Big)\Big)$$

This formulation ensures that exploration is not random but strategically biased toward promising regions, guided by both stochastic variables and memory-dependent forces. The randomness term r, combined with directional coefficients and oscillatory decay, introduces structured diversity that helps agents escape local minima and probe underexplored regions.

The global best solution across the population is continuously updated to drive the collective behavior of the agents. This is achieved through an exponentially weighted averaging mechanism:

$$F_{\text{best}} = \frac{1}{K}\sum_{n=0}^{K}\left(\frac{F_{\max}^{n^2}}{(2n+1)^2}\right)$$

This formulation enhances convergence by amplifying high-performing solution components while maintaining normalization. The value of K, an iteration-sensitive exponential factor, increases over time to transition the algorithm smoothly from exploration to exploitation.

As the optimization process advances, FbOA naturally shifts toward an *exploitation*-dominant regime. This transition is embedded in the algorithm's iterative logic through convergence-aware positional updates. In the exploitation phase, agents reduce their movement radius and focus their search around high-fitness regions identified during earlier exploration. This behavior is implemented through a sinusoidal refinement mechanism:

$$Fb(S_{t+1}) = F_i + z_3 \cdot Fb(S_t) + K \cdot \sin\Big(\frac{\pi}{\text{Iteration}}\Big)$$

Here, the parameter K, which scales with the number of iterations, ensures that as the algorithm progresses, movement becomes increasingly fine-grained. The use of sinusoidal modulation prevents stagnation by preserving low-amplitude oscillations that maintain minimal diversity, even in later stages.

To prevent the algorithm from converging prematurely or being trapped in suboptimal regions, FbOA introduces a mutation operator designed to inject variability at critical junctures. This operator perturbs agent positions using a cosine-rational function that scales with iteration progress:

$$S(t) = K \cdot aq\left(\frac{2n+1}{x}\right) + K \cdot \cos\Big(\frac{\pi}{\text{Iteration}}\Big)$$

This strategy promotes long-term adaptability and reinforces the optimizer's ability to recover from premature convergence. The controlled oscillation ensures that mutation strength decays gradually, maintaining global search capacity without destabilizing late-stage convergence.

The complete procedure of FbOA is captured in the following formal pseudocode:

To contextualize FbOA's performance within the broader optimization literature, a comparative analysis was conducted using several well-established metaheuristic algorithms. These included Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Whale Optimization Algorithm (WOA), Biogeography-Based Optimizer (BBO), Multiverse Optimization (MVO), Bat Algorithm (BA), Stochastic Fractal Search (SFS), Differential Evolution (DE), and Artificial Protozoa Optimizer (APO). Each optimizer was independently applied to the same forecasting model using identical initialization protocols, parameter bounds, and evaluation budgets to ensure fairness and reproducibility.

All optimization procedures were executed within a normalized hyperparameter space tailored to the architectural specifications of the target model. The evaluation criterion for

Algorithm 1 Football Optimization Algorithm (FbOA)

Require: Objective function $f(x)$, population size N, maximum iterations T
Ensure: Best-found solution x^*
Initialize population of agents with random positions and velocities
Evaluate initial fitness and determine global best F_{best}
for $t = 1$ to T **do**
  for each agent i **do**
    Update velocity V_n using dynamic force model
    Update agent position for exploration or exploitation
    Apply mutation to introduce diversity when needed
    Evaluate new fitness and update F_{best} if improved
  end for
end for
return F_{best}

each optimizer was a composite loss function measured over a validation set, ensuring that hyperparameter configurations led to models that generalized well beyond the training data. Fixed iteration counts and population sizes were used on all of the optimizers to allow a fair comparison.

By integrating a new team-strategy-based optimizer (FbOA) and comparing it with a well-established body of existing metaheuristics, this paper presents a methodologically sound framework for intelligent hyperparameter tuning in deep learning-based agricultural forecasting. The lessons learned from comparing this type of structure would not be limited to selecting methodologies for specific optimization fields. Still, they would also inform the advancement of machine learning in optimization.

3.4.1 Comparative Metaheuristic Algorithms

To provide a broader perspective on the performance of FbOA within the optimization literature, a comparative study was conducted with several established metaheuristic algorithms. Each was used separately in the same forecasting model with the same initialization protocols, hyperparameter bounds, population sizes and iteration limits so that the methods were not methodologically unfair and brought some reproducibility.

The optimization algorithms that were chosen to be compared are as follows:

- **Particle Swarm Optimization (PSO):** probably the most well-known swarm intelligence algorithm, which derives its ideas from a flock of birds or a group of fish, characterized by a simple concept and exemplary performance in continuous optimization[35].
- **Genetic Algorithm (GA):** An evolutionary solution in which a solution is represented as a genome organised into chromosomes and in which crossover, mutation and selection are used to search and explore the solution space.
- **Whale Optimization Algorithm (WOA):** a nature-inspired algorithm that models the bubble-net feeding behavior of humpback whales, and specifically, in searching high-dimensional spaces[36].
- **Biogeography-Based Optimizer (BBO)** is an algorithm that models the migration and mutation of species to different habitats, making it best suited for problems that require the sharing of solutions and diversity maintenance[37].
- **Multiverse Optimization (MVO):** Founded on the concepts of black holes, white holes and wormholes with multiverse cosmology, this optimizer enables the exchange of probabilistic solutions across populations[38].
- **Bat Algorithm (BA):** An ear-based optimizer which was modelled after the echolocation behavior of bats and which dynamically adapts movement based on frequency and loudness parameters[39].
- **Stochastic Fractal search (SFS):** This search algorithm uses a fractal-based approach to diffuse search agents according to a stochastic process, promising acceptable performance when the landscape is noisy and chaotic[40].
- **Differential Evolution (DE)**:A robust population-based search algorithm based on vector-based mutation and crossover operators, optimising well in a continuous numerical population[41].
- **Artificial Protozoa Optimizer (APO):** A newer bio-inspired optimizer that simulates the behavior of protozoa-attractiveness, repelling and random movement to ensure exploration over complex landscapes[42].

The algorithms ran within a bounded, unitless search space, optimized to the structural limitations of the model. The composition of the validation loss function was used to define the evaluation objective as a combination of several prediction measures.. To alleviate the differences in the comparisons, all optimizers were tested with identical numbers of iterations and populations, and selected at random with identical seeds. This systematic benchmarking process not only presents a reasonable performance benchmark but also provides an enhanced view of the trade-offs and merits presented by each of the metaheuristic strategies.

3.5 Evaluation Metrics

As a way to evaluate the performance of forecasting models created within the scope of this research, a wide range of statistical standard metrics of evaluation was used. Such metrics were chosen to capture various facets of the quality of prediction, and they are an absolute error magnitude, directional bias, correlation with ground truth, and overall predictive agreement.

The measures are divided into three broad categories: (i)absolute and squared error metrics i.e.; Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE); (ii) bias and correlation measures i.e.; Mean Bias Error (MBE), Pearson correlation coefficient (r) and coefficient of determination (R^2), and (iii) relative and

efficiency-based metrics i.e.; Relative Root Mean Squared Error (RRMSE), Nash-Sutcliffe Efficiency (NSE) and the Willm Collectively, these nine measures induce a balanced understanding of how models perform in time-series regression and make it simple to fairly compare models pre- optimization with their optimized counterparts.

Table 2 presents the mathematical formulation for each metric used in this study.

These metrics were used throughout the study to benchmark the forecasting performance of all models and to guide the optimization process under various metaheuristic strategies.

4. EMPIRICAL RESULTS

4.1 Baseline Model Comparison

The first step of the empirical analysis is dedicated to the benchmarking of the performance of various machine learning models on the forecasting task before any hyperparameter optimization or metaheuristic intervention. This baseline comparison is aimed at determining the best predictive architecture to model the long-range agricultural production time-series data, as well as provision of a performance benchmark to which the effect of further optimization can be compared.

All the models were trained on the preprocessed data discussed in Section 2, and data were uniformly split, encoded, and aligned in time. In order to make the evaluations fair, all the models were started with default or generally accepted hyperparameter configurations where available in the literature. The input sequences of both models were organized with the application of a standard sliding-window mechanism and the prediction horizon remained similar as well among all experiments. None of the model-specific tuning was performed, and all training runs were run in the same computational environment.

The models that have been evaluated in this comparison are Informer, N-BEATS, LogTrans, N-HITS, EALSTM, TST, and LSTM. These architectures are a varied slice of deep learning techniques, and include attention transformers, feedforward-based residual models, hierarchical interpolators, and traditional recurrent models. Their inclusion was driven by the previous empirical effectiveness of time-series forecasting, and their varying inductive biases, which give information on the impact of architectural decisions on forecasting performance in agricultural fields.

Table 3 show the predictive performance of each model as measured on the hold out test set using the entire set of evaluation metrics described in Section 3. The outcomes indicate the performance of the model on unknown data and they are applied to determine the generalization capacity of each model.

In order to investigate further the distributional qualities of model evaluation measures, Figure 6 provides a mixed visualization of a combination of. boxplots and curve density estimation (KDE) of each metric. This representation gives a statistical summary (through boxplots) as well as a smooth probability distribution (through KDE), which provides a more detailed. knowledge of metric variation. The central tendencies are pointed out in the plots. spread, and density distributions match each other in terms of MSE, RMSE, MAE, etc. MBE, r, R^2, RRMSE, NSE, and WI. The density estimates and combination of density estimates was found to be effective. boxplots makes it possible to identify both the normal levels of performance and potential. outliers, thus helping to make a more multifaceted understanding of model reliability.

In order to compare the results of the baseline deep learning architectures to the performance of the Informer model, A heatmap of normalized measures of evaluation is shown in Figure 7. Such models as N-BEATS, LogTrans, N-HITS, and others can be seen in this visualization as a comparative overview of the models. EALSTM, TST and LSTM with respect to Informer. The normalization (on a 0-1 scale) permits metrics (MSE, RMSE, MAE, MBE, r, R^2, RRMSE, NSE, and WI) will be compared directly. across models. The findings indicate that Informer is performing well in relation to other Informer. most of the indicators, and models such as LSTM and N-BEATS show competitive results in a selected. metrics. The heatmap is useful in creating and identifying the strengths and weaknesses of every model. underlying a subtle comparison of architectures.

In order to give detailed comparing analysis of model performance on model error measures, The figure below, Figure 8, demonstrates the values of MSE, RMSE and MAE. all the deep learning architectures, their average, and standard deviation. This bar plot does not just show the absolute performance of models like Informer, N-BEATS, and LogTrans, N-HITS, EALSTM, TST, and LSTM there are also the variability of their. results. The assessment of the mean and standard deviation bars allows the inclusion of bars. both stability and central tendency, provides information on model consistency. As it can be observed, Informer and LogTrans can obtain relatively low error values with smaller. as compared to TST and LSTM, which feature a larger error magnitude with larger dispersion. The visualization aids in the determination of models that are well balanced in accuracy and robustness.

To provide a disaggregated perspective on model performance, Figure 9 presents a facet grid visualization of evaluation metrics across all tested models. Each subplot corresponds to a specific metric—including MSE, RMSE, MAE, MBE, r, R^2, RRMSE, NSE, and WI—thereby enabling a detailed comparison of how different models perform under multiple criteria. This layout facilitates the identification of trade-offs, such as models achieving lower error values but slightly reduced correlation coefficients, or vice versa. The facet grid structure also highlights consistent patterns across metrics, helping to pinpoint models that demonstrate balanced and robust performance across diverse evaluation dimensions.

In order to offer disaggregated view of model performance, A facet grid is shown in Figure 9 . of assessment measures among all the tested models. Each subplot corresponds to a given metric such as MSE, RMSE, MAE, MBE, r, R^2, RRMSE, this is the case with NSE, and WI–thus making it possible to compare in detail how various models. compete on a variety of standards. The identification is made possible by this layout. of trade-offs, e.g. models with smaller error values but with a small decrease. correlation coefficients, or the other (way around). The facet grid structure also points

Table 2. Evaluation metrics used to assess time-series forecasting performance.

Metric	Mathematical Definition
Mean Squared Error (MSE)	$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
Root Mean Squared Error (RMSE)	$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
Mean Absolute Error (MAE)	$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$
Mean Bias Error (MBE)	$\mathrm{MBE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)$
Pearson Correlation Coefficient (r)	$r = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}}$
Coefficient of Determination (R^2)	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$
Relative RMSE (RRMSE)	$\mathrm{RRMSE} = \frac{\mathrm{RMSE}}{\bar{y}} \times 100$
Nash–Sutcliffe Efficiency (NSE)	$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$
Willmott Index of Agreement (WI)	$\mathrm{WI} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}\left(\lvert\hat{y}_i - \bar{y}\rvert + \lvert y_i - \bar{y}\rvert\right)^2}$

Table 3. Baseline performance of machine learning models before optimization.

Model	MSE	RMSE	MAE	MBE	r	R^2	RRMSE	NSE	WI
Informer	0.0064	0.0799	0.0434	0.0302	0.8780	0.8906	6.00	0.9070	0.9042
N-BEATS	0.0389	0.1973	0.0531	0.0475	0.8618	0.8744	8.14	0.8847	0.8549
LogTrans	0.0476	0.2182	0.0618	0.0533	0.8383	0.8509	8.42	0.8608	0.8563
N-HITS	0.0536	0.2314	0.0704	0.0896	0.8367	0.8493	8.62	0.8503	0.8439
EALSTM	0.1064	0.3261	0.0965	0.1010	0.8217	0.8433	8.76	0.8351	0.8336
TST	0.2320	0.4816	0.1423	0.1923	0.8134	0.8225	9.18	0.8142	0.8300
LSTM	0.3784	0.6152	0.1652	0.2532	0.7933	0.8030	9.60	0.7920	0.7975

out similar trends in metrics which aid in identifying models that indicate. consistent and strong work performance in various evaluation aspects.

To examine the probability distribution of model assessment measures, Density plots of the kernel density are given in Figure 10. estimation (KDE) superimposes all the metrics used in this research. The subplots are associated with a particular measure such as MSE, RMSE, MAE, etc. MBE, r, R^2, RRMSE, NSE, WI offering information about the form, disperse, and concentration of distributions. The smooth KDE curves supplement the density histograms, indicating the skew existence. or rough normality where there is need. As illustrated, the majority of the metrics display. stable model behavior, which is manifested by well-defined unimodal distributions, as others indicate greater variance that indicates more diverse results.

In terms of overall performance, the Informer architecture showed the highest performance in almost all the measures with the lowest MSE and RMSE and the highest correlation coefficient and Nash-Sutcliffe Efficiency (NSE). The transformer-based attention mechanism in this model seems to be especially appropriate to address long-range temporal correlations and nonlinear trends of national-scale agricultural production data. Competitive performance was also demonstrated by other architectures like N-BEATS and LogTrans, which in turn have higher absolute errors and lower agreement indices. Classical recurrent architectures like LSTM were outperformed by more recent attention-based or residual architectures in more complex multivariate time-series forecasting problems.

This baseline comparison serves as the foundation for subsequent optimization analysis. Based on its superior generalization performance, the Informer model is selected as the primary candidate for hyperparameter optimization using the metaheuristic strategies described in Section 2. The next subsection (Section 4.2) presents the results of these optimization procedures and compares their impact on model accuracy, stability, and computational efficiency.

4.2 Optimized Model Analysis

Following the baseline model comparison, the next phase of the empirical investigation involved applying metaheuristic algorithms to optimize the hyperparameters of the best-

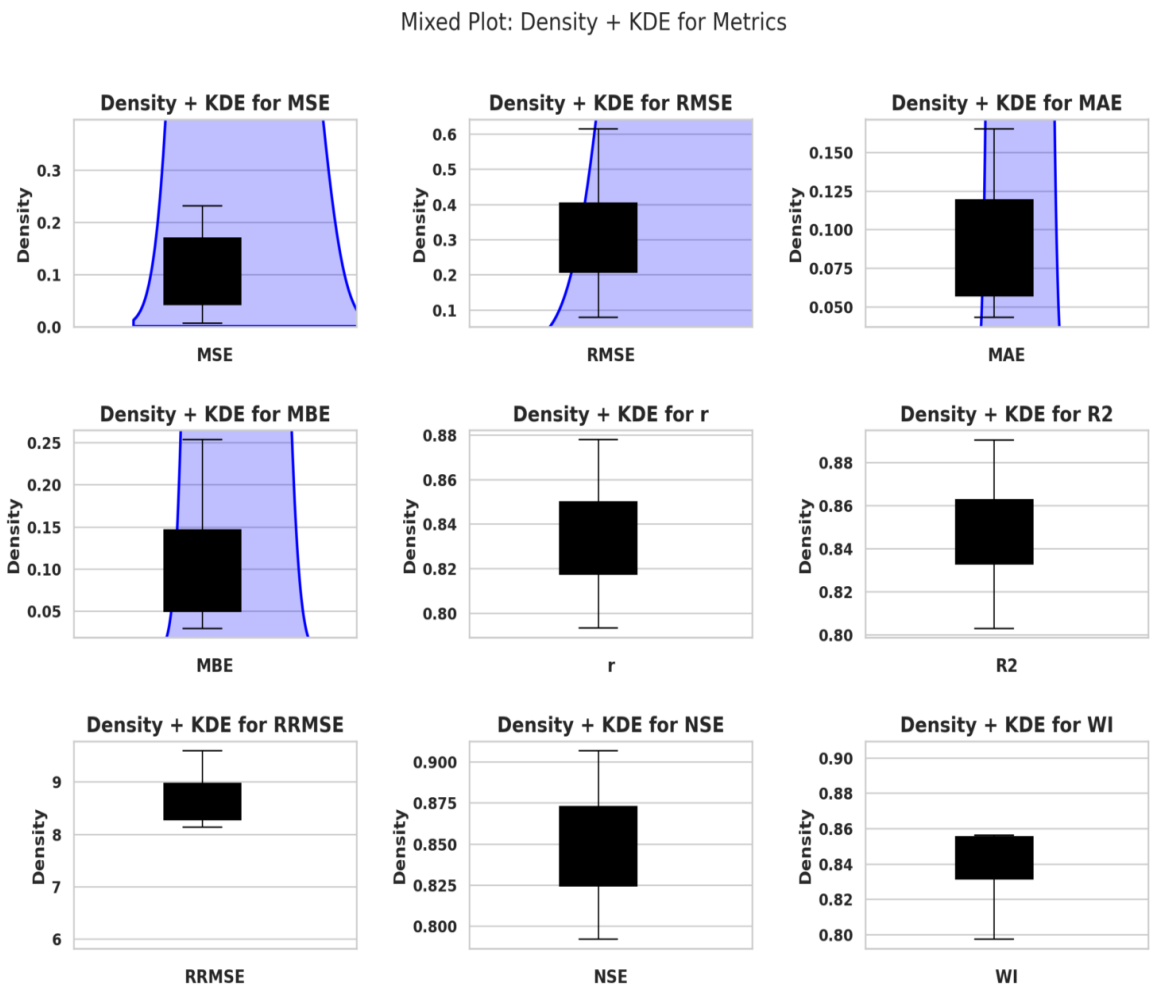

Figure 6. Mixed plots combining boxplots with KDE for all evaluation metrics.

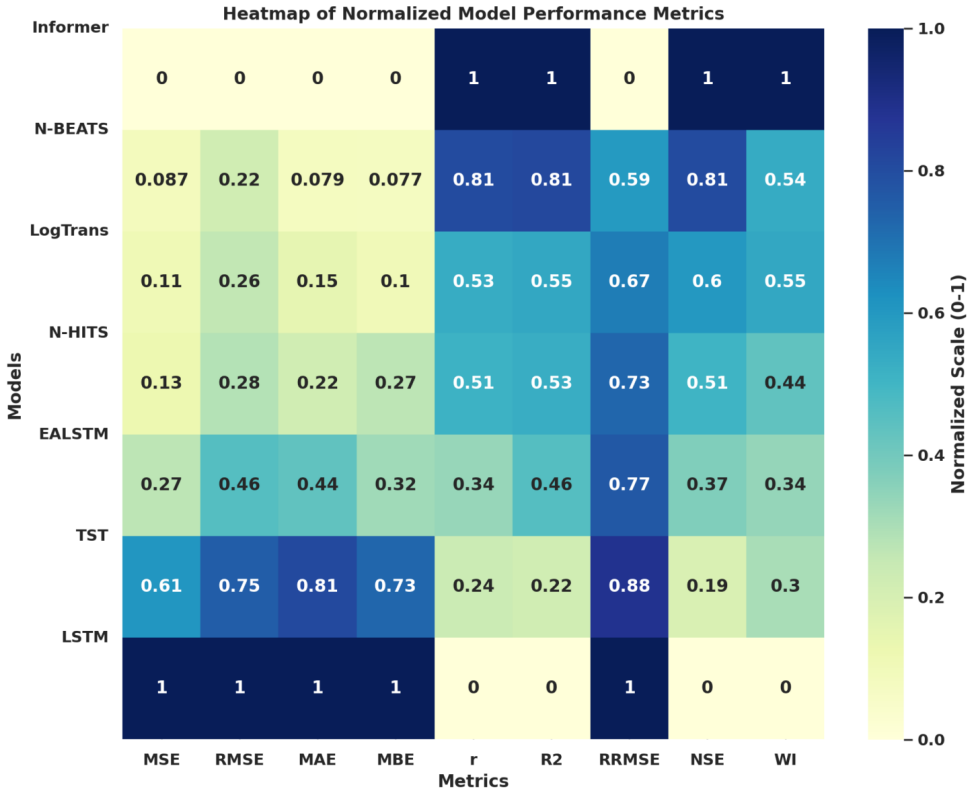

Figure 7. Heatmap of normalized performance metrics comparing Informer with baseline deep learning models.

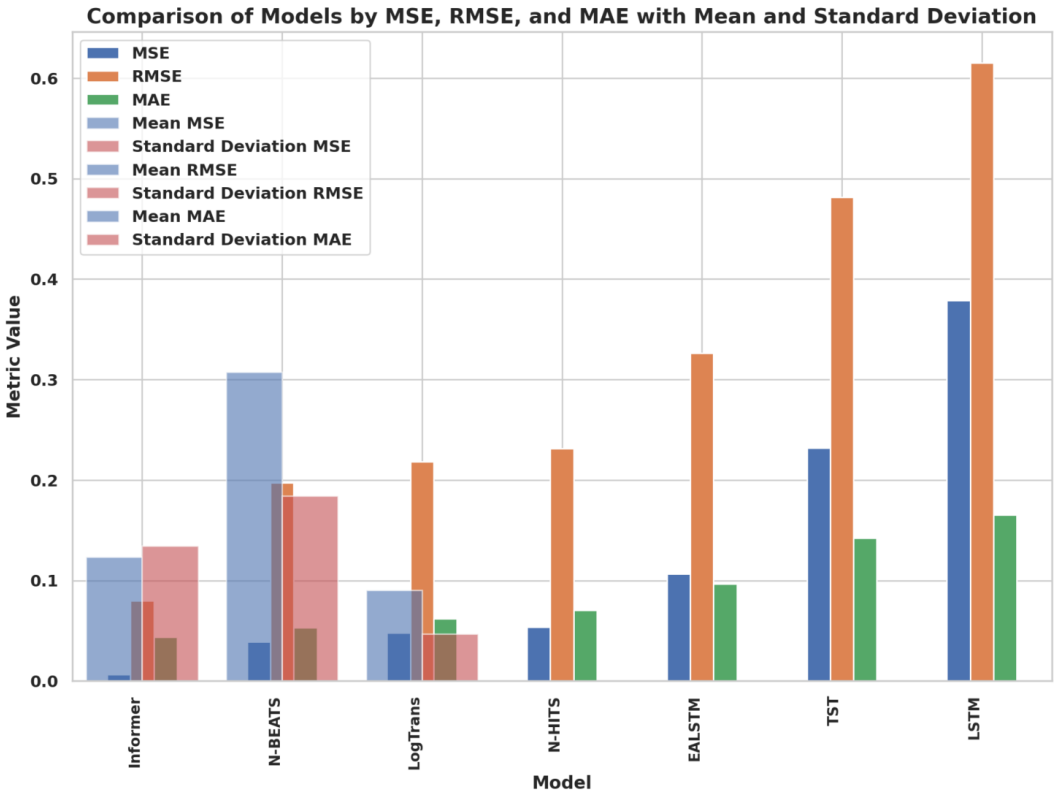

Figure 8. Comparison of models by MSE, RMSE, and MAE with mean and standard deviation.

performing model, namely the Informer architecture. This optimization process was executed using the suite of algorithms described in Section 2.5, including the Football Optimization Algorithm (FbOA), Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Whale Optimization Algorithm (WOA), Biogeography-Based Optimizer (BBO), Multiverse Optimization (MVO), Bat Algorithm (BA), Stochastic Fractal Search (SFS), Differential Evolution (DE), and Artificial Protozoa Optimizer (APO).

Each optimizer was applied independently under a standardized experimental configuration. The optimization objectives were defined by the same set of evaluation metrics described in Section 3, with the composite loss used as the fitness function guiding the search process. To ensure comparability, all optimizers were run using identical iteration limits, population sizes, and random seed initializations.

Table 4 presents the forecasting performance of the Informer model after optimization by each metaheuristic algorithm. The metrics include error-based, bias-based, and fit-based measures, providing a detailed view of how each strategy influenced model accuracy and predictive stability.

To comprehensively evaluate the predictive performance of the models, Figure 11 presents a radar plot summarizing the mean and standard deviation of key evaluation metrics across all tested models. This visualization provides a holistic comparison by simultaneously mapping multiple performance indicators such as MAE, RMSE, RRMSE, R^2, and NSE. The blue line represents the mean performance values, while the shaded region indicates the variability (mean plus standard deviation), thus highlighting the degree of consistency or uncertainty associated with each metric. By capturing both central tendency and dispersion, the radar plot offers valuable insights into the robustness and stability of model predictions.

To assess the distributional properties of the evaluation metrics and verify their adherence to normality assumptions, Figure 12 presents the Q–Q plots for all considered performance indicators, including MSE, RMSE, MAE, MBE, r, R^2, RRMSE, NSE, and WI. Each subplot compares the ordered sample values of a metric against the corresponding theoretical quantiles of a normal distribution. The alignment of points along the red reference line indicates the degree of normality, while deviations reveal potential departures from Gaussian behavior. As observed, most metrics demonstrate a reasonably close fit, suggesting that statistical analyses and model comparisons based on these metrics remain valid and reliable.

To investigate the joint behavior of error-based evaluation metrics, Figure 13 illustrates the contour density distribution between the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), with a scatter overlay of the observed data points. This visualization highlights the correlation structure and concentration of model performance outcomes. The contour levels indicate regions of higher density, suggesting where most model results cluster, while the scatter points provide the actual distribution of observations. The figure reveals that most models fall within a compact region characterized by relatively low MAE and RMSE values, emphasizing the consistency of predictive accuracy across different runs.

To facilitate a comparative evaluation of different optimization-based hybrid models, Figure 14 displays a heatmap of normalized performance metrics. This visualization condenses multiple evaluation indicators—such as MSE, RMSE, MAE, MBE, r, R^2, RRMSE, NSE, and WI—into a unified scale ranging from 0 to 1, thereby

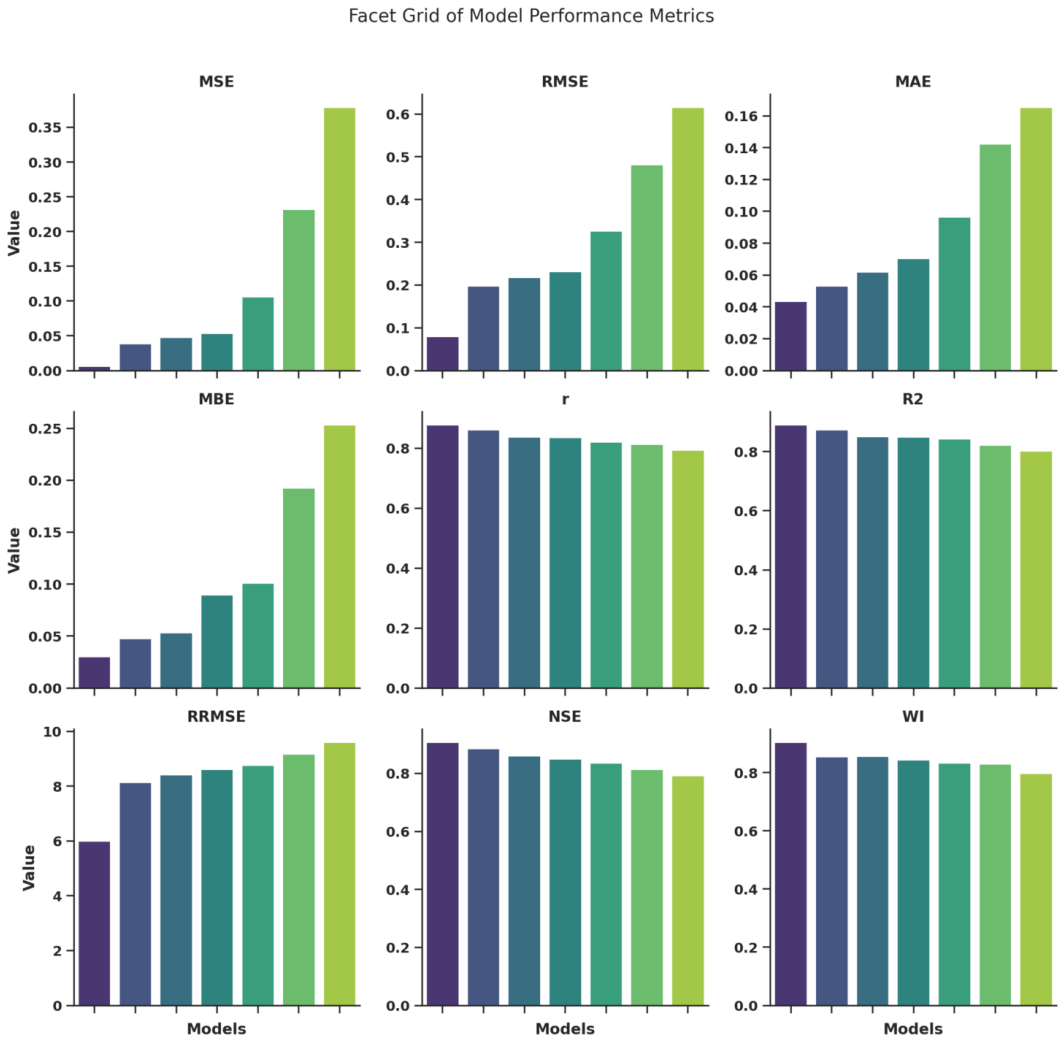

Figure 9. Facet grid visualization of model performance across multiple evaluation metrics.

Table 4. Performance of the optimized Informer model using different metaheuristic algorithms.

Optimizer	MSE	RMSE	MAE	MBE	r	R^2	RRMSE	NSE	WI
FbOA + Informer	1.19E-06	1.09E-03	8.62E-05	1.01E-04	0.9550	0.9600	1.39	0.9260	0.9290
PSO + Informer	1.49E-05	3.85E-03	1.76E-03	6.70E-04	0.9280	0.9380	1.94	0.9150	0.9170
GA + Informer	3.62E-05	6.02E-03	1.77E-03	7.75E-04	0.9260	0.9320	2.21	0.9110	0.9100
WOA + Informer	4.26E-05	6.53E-03	1.78E-03	8.81E-04	0.9250	0.9300	2.46	0.9090	0.9070
BBO + Informer	6.47E-05	8.04E-03	1.81E-03	1.04E-03	0.9120	0.9290	2.66	0.9030	0.9050
MVO + Informer	6.86E-05	8.28E-03	1.84E-03	1.12E-03	0.9110	0.9270	2.81	0.9000	0.9020
BA + Informer	7.01E-05	8.37E-03	1.86E-03	1.18E-03	0.9110	0.9230	2.97	0.8900	0.9040
SFS + Informer	9.72E-05	9.86E-03	1.89E-03	1.43E-03	0.9080	0.9190	3.39	0.8870	0.8980
DE + Informer	1.11E-04	1.06E-02	1.89E-03	1.61E-03	0.9070	0.9180	3.57	0.8850	0.8970
APO + Informer	1.53E-04	1.24E-02	1.91E-03	1.63E-03	0.9030	0.9160	3.65	0.8820	0.8960

enabling direct comparisons across models. The heatmap highlights how each model performs relative to others, with darker shades indicating stronger performance and lighter shades representing weaker outcomes. For instance, models such as `APO + Informer` and `DE + Informer` achieve consistently high values across several metrics, whereas others demonstrate more uneven results. This comparative framework is particularly useful in identifying models that strike a balance between accuracy, consistency, and robustness.

To examine the progression of model errors in a continuous manner, Figure 15 presents the cubic spline interpolation of Mean Squared Error (MSE) values across the different hybrid optimization–Informer models. This approach smooths the discrete error values (red points) into a continuous curve (blue line), making it easier to detect underlying patterns and transitions in model performance. The visualization highlights how error magnitudes vary across models, with a general upward trend observed from `FbOA + Informer` to `APO + Informer`. Such interpolation not only clarifies the compar-

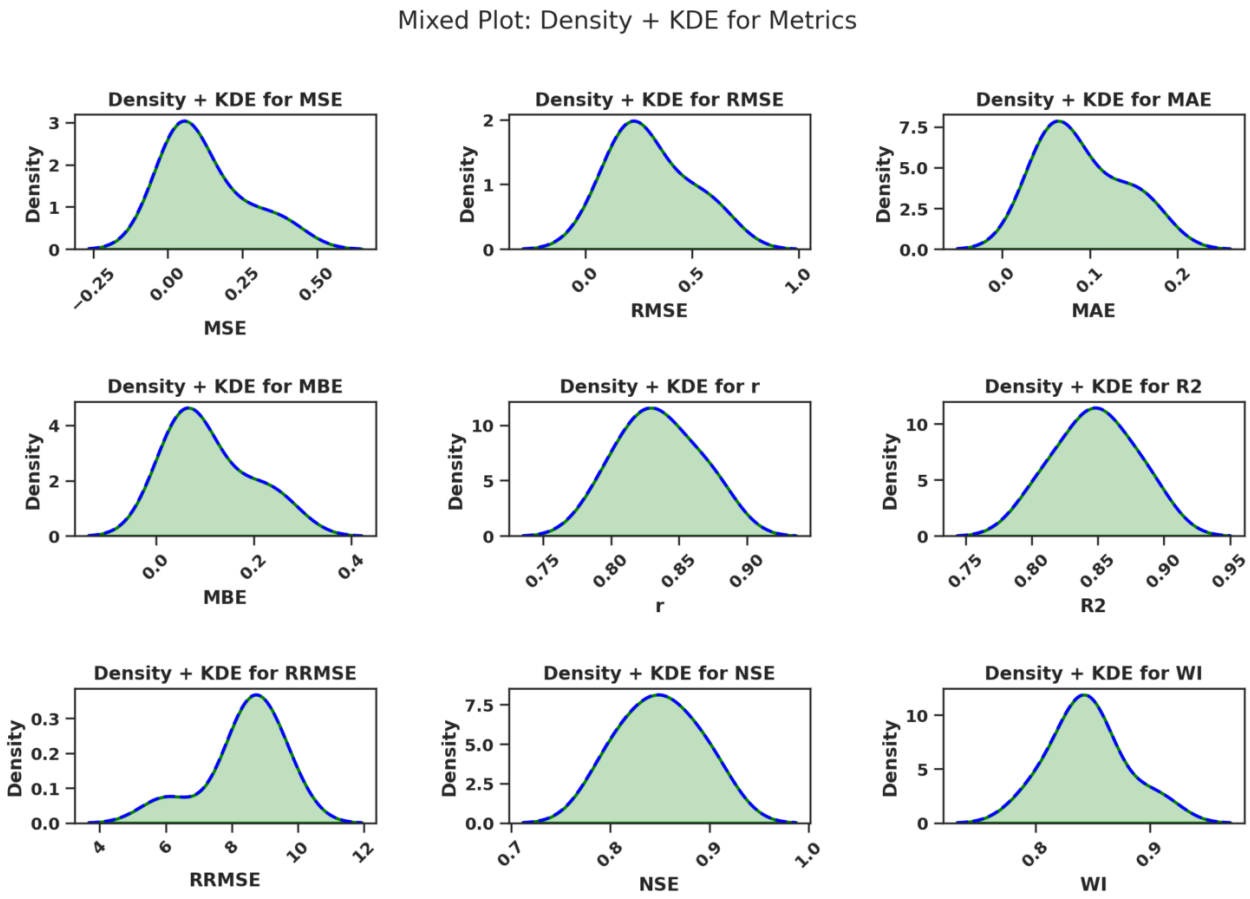

Figure 10. Density and KDE plots of model evaluation metrics.

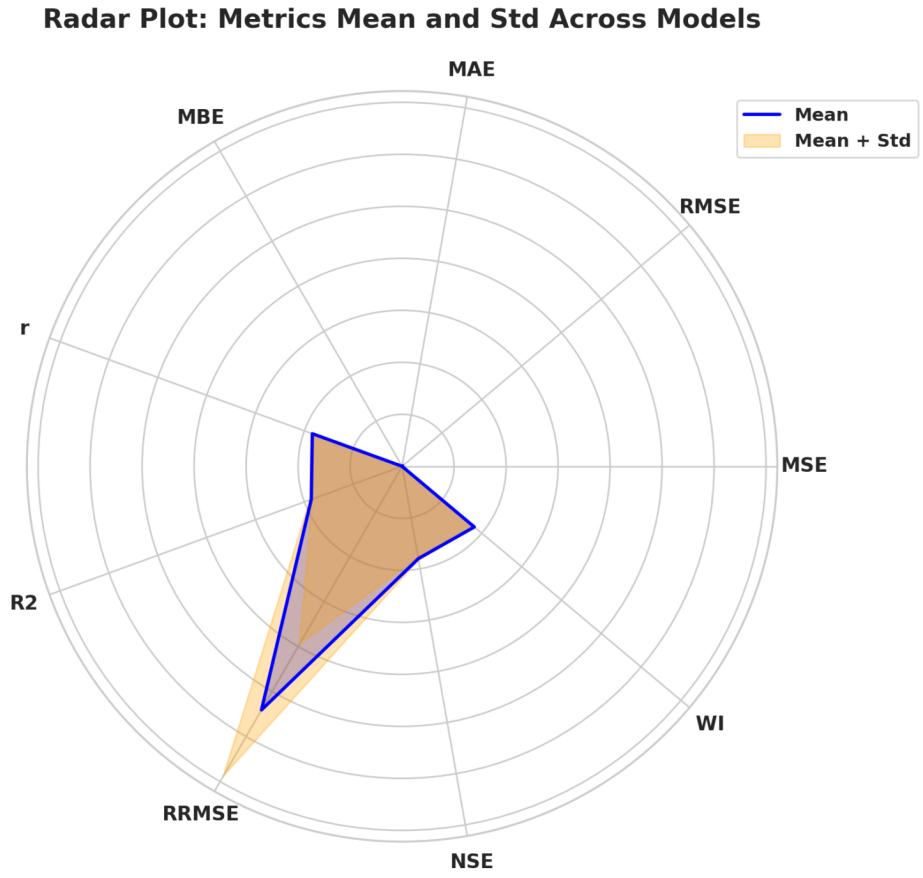

Figure 11. Radar plot of mean and standard deviation of evaluation metrics across models.

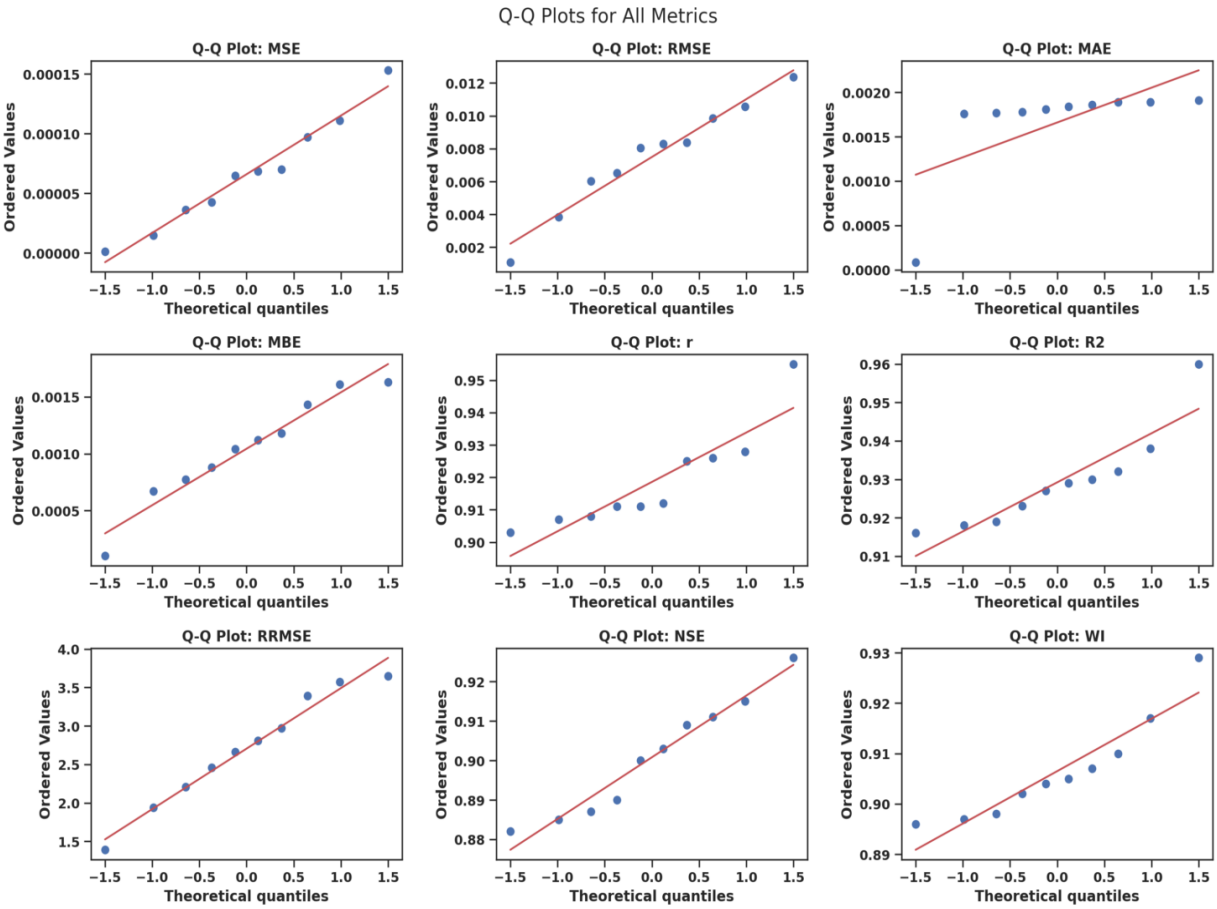

Figure 12. Q–Q plots of all evaluation metrics used in model assessment.

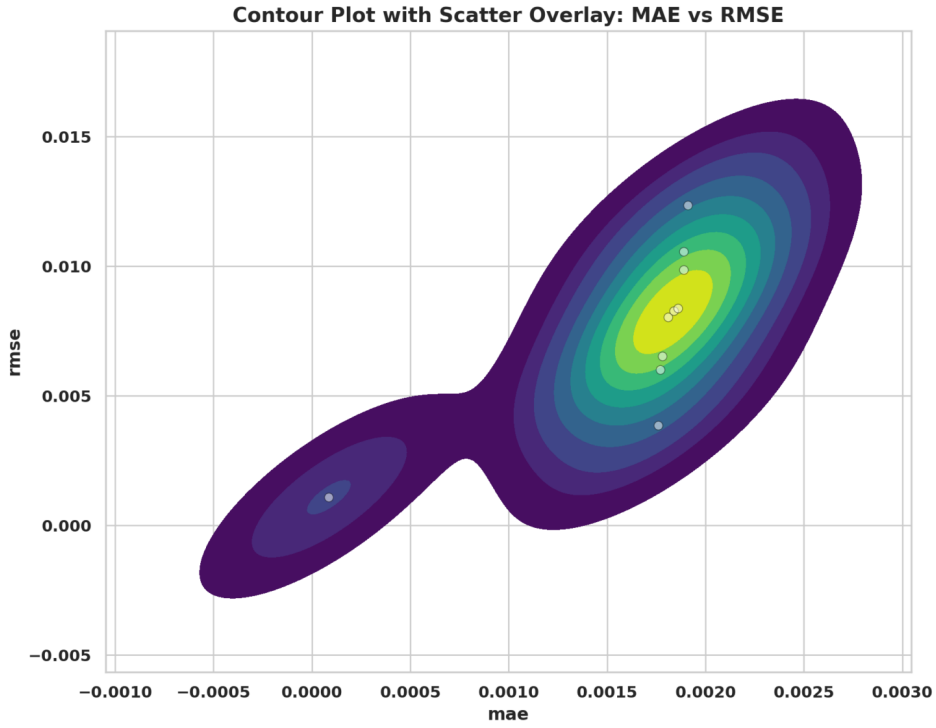

Figure 13. Contour plot with scatter overlay of MAE vs. RMSE across models.

ative performance of models but also provides insight into gradual shifts in predictive reliability.

Among the optimization strategies tested, the Football Optimization Algorithm (FbOA) achieved the strongest improvement, yielding the lowest mean squared error (MSE) and root mean squared error (RMSE), while also achieving the highest Nash–Sutcliffe Efficiency (NSE), Willmott Index (WI), and Pearson correlation coefficient. These improvements are accompanied by a significant reduction in both absolute and

Figure 14. Heatmap of normalized performance metrics across optimization-based hybrid models.

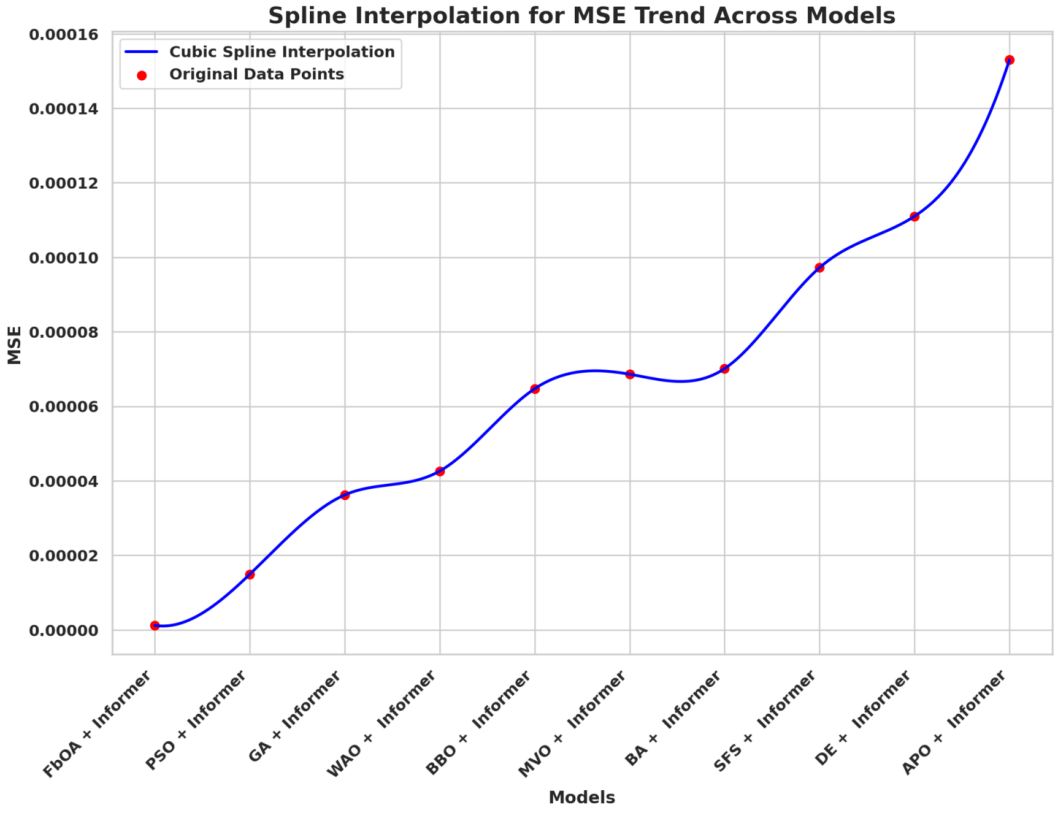

Figure 15. Cubic spline interpolation of MSE trends across hybrid optimization–Informer models.

relative prediction error metrics, suggesting that FbOA was able to effectively explore the hyperparameter search space and converge on a highly generalizable configuration for

5. DISCUSSION

The experimental results discussed in the sections above allow making some valuable observations about the functionality

of high-performance machine learning systems of long-range agricultural prognostication and the role of metaheuristic optimization in improving model effectiveness. The findings indicate that, although the architectural pick could be a determining factor in the development of the predictive accuracy at the base, the fine-tuning of hyperparameters with the help of smart search strategies could achieve significant enhancement to the forecasting accuracy, consistency, and generalization. The Informer model that came out as the best performing architecture in the baseline comparison showed strong predictive behavior even when the default parameters were set. This result is in agreement with the accumulating literature that confirms the usefulness of attention-based architectures in learning complex temporal dependencies on long horizons. Encoding long-range trends and the capacity of the model to combine multi-step flow of information seems especially appropriate to model the year-over-year changes in crop production, which usually represent structural trends, cyclical patterns, and long response time to economic or environmental changes. When the Informer model was exposed to metaheuristic-based hyperparameter optimization, the performance increase was found in all the evaluation metrics. The best advancements were obtained by the Football Optimization Algorithm (FbOA) that was always superior to the other algorithms in terms of error-based, statistical-fit, and agreement measures. This indicates that the hybrid exploration-exploitation strategy of FbOA that simulates cooperative and adaptive movement based on the football team dynamics could be an exceptionally successful mechanism of exploring the high-dimensional and non-linear search space of deep learning model configuration. It is interesting to note the significant increase in such metrics as MSE, RMSE, NSE, and the Willmott Index (WI). These benefits indicate a minimized absolute forecasting error on top of increased conceptual concord between anticipated and observed temporal patterns. The growth in values of Pearson correlation coefficient and R 2 also prove that the optimized Informer model was more faithful to underlying production trends in capturing both magnitude and direction of year-to-year changes. Other metaheuristics including PSO, GA, and WOA produced fairly varied increases in performance, in that they enhance predictive stability and minimize error, but with less ability to achieve the accuracy and reliability of FbOA. The operation of other algorithms such as DE, SFS, and APO was relatively worse and this could have been due to the fact that their search dynamics might have been less able to work with the structural property of the Informer model or the nature of hyperparameter space in this case. In addition to performance metrics, the results support the significance of assimilating the model-optimizer. Although the architecture of Informer seems to be inherently advantageous in terms of the global search of population-based optimizers, not every optimization algorithm was as effective. These better performance advances with FbOA indicate that, in certain cases, optimization efficiency can be improved with domain-inspired search heuristics, including those that simulate tactical coordination and/or adaptive positioning, than with general-purpose metaheuristics. One should also not overlook the drawbacks of this method. Although the process of optimization is effective, it is computationally expensive and susceptible to the size of a population and convergence levels, as well as stochastic starting point. In addition, the optimization of one model (Informer) could restrict the applicability of results to all forecasting structures. The framework may be furthered in future literature to evaluate optimizers-model combinations in a more systematic fashion and further to study hybrid or ensemble optimization methods that integrate more than one search mechanism. In practical terms, the gains made on the use of metaheuristic optimization and especially FbOA have significant implications to the agricultural forecasting systems. Improved predictive precision helps to make improved decisions in areas like the crop yield planning, food supply chain coordination, and policy forming. When the gains in this study are applied to real time or operation based forecasting systems, the benefits may be translated to proactively and efficiently manage the resources within the agricultural economies. To conclude, the findings show that modern forecasting architectures like Informer can be highly baseline-performing, but it is only possible to achieve their maximum predictive scope through strategic optimization. Metaheuristic algorithms, and specifically based on dynamic and cooperative search concepts such as FbOA are a potent means of tapping into this potential. The results can be added to the already accumulated evidence base on the combination of deep learning with intelligent optimization and indicated recent research directions in the field of artificial intelligence, agriculture, and decision analytics.

6. CONCLUSION AND FUTURE WORK

This paper introduced a general system architecture of predicting agricultural production that combines sophisticated deep learning models with metaheuristic hyperparameter optimization algorithms. The empirical basis of the assessment of the predictive ability of various state-of-the-art machine learning models was a multi-decade, multi-country dataset of national records on tomato and potato production. Of the tested models, the Informer architecture presented the best baseline performance, due to the capabilities of the long-range temporal dependency modeling and sequence-level attention mechanisms. Nonetheless, significant profit was obtained when this model was optimized with the help of metaheuristic algorithms. Specifically, a new optimizer, Football Optimization Algorithm (FbOA), that is based on the principles of strategic interaction in the football gameplay provided better results on all the considered metrics, which are error magnitude, correlation structure, and predictive agreement. The combination of FbOA with the training process resulted in a dramatic change of forecast accuracy, model stability, and generalization. These improvements lead to the strength of smart hyperparameter optimization, particularly in nonlinear time-series modeling tasks with high dimensionality as is found in the real-world predictive time-series scenario. Comparison and contrast to other metaheuristics provided evidence that the dynamic, role-based coordination schemes that are incorporated in the design of FbOA provide competitive benefits in terms of the efficiency of search and convergence behavior. In practical terms, it is possible to see a high potential of the proposed forecasting pipeline to be deployed in the field of agricultural monitoring, resource management, and policy development. A more precise and predictable method of crop production can allow the stakeholders, such as governments,

cooperatives and supply chain managers, to make proactive and data-driven decisions, especially within the context of climatic variability and world food security. In spite of these contributions, a number of avenues in future research exist. To start with, whereas FbOA proved to be highly performing within this environment, further testing on larger datasets and other forecasting problems is necessary to confirm its external validity. Second, it can be expanded by the optimization framework to incorporate hybrid metaheuristics or multi-objective approaches, which are able to strike a balance between accuracy, computational cost, and interpretability at the same time. Third, research that is deployment-oriented might be interested in how this framework can be incorporated into operational forecasting systems, such as real-time dashboards or built-in decision support systems of agriculture ministries or development agencies. Lastly, future researchers can consider using exogenous factors, i.e., weather conditions, economic data, or satellite-based indices to improve model explanatory capacity and improve the applicability of the models to agriculture planning in climate resilient situations. The combination of intelligent optimization methods with explainable, scalable prediction systems is one opportunity in the application of AI to agricultural systems worldwide.

REFERENCES

[1] M. Sangeetha, G. Thejaswini, A. Shoba, S. S. Gaikwad, R. T. Amretasre, and S. Nivedita, "Design and development of a crop quality monitoring and classification system using iot and blockchain," *Journal of Physics Conference Series*, vol. 1964, pp. 62011–62011, 07 2021.

[2] L. Hou, Z. Liu, J. You, Y. Liu, J. Xiang, J. Zhou, and P. Yu, "Tomato sorting system based on machine vision," *Electronics*, vol. 13, pp. 2114–2114, 05 2024.

[3] N. G. Abd El-Fattah, M. S. Abd El-baki, M. Maher, and S. Elsayed, "Integrating climate and plant variables with machine learning models to forecast tomato yield at different soil moisture levels," *Egyptian Journal of Soil Science*, 2024.

[4] U. S, V. Muralibhaskaran, M. G, and V. G, "Forecasting of tomato prices and yield based on season and weather using lstm and arima algorithms," in *2024 International Conference on Emerging Research in Computational Science (ICERCS)*, pp. 1–6, 2024.

[5] S. M. H. Khatibi and J. Ali, "Harnessing the power of machine learning for crop improvement and sustainable production," 08 2024.

[6] X. Qin, B. Wu, H. Zeng, M. Zhang, and F. Tian, "Global gridded crop production dataset at 10 km resolution from 2010 to 2020," *Scientific Data*, vol. 11, p. 1377, 2024.

[7] B. Bischl, M. Binder, M. Lang, T. Pielok, J. Richter, S. Coors, J. Thomas, T. Ullmann, M. Becker, A. Boulesteix, D. Deng, and M. Lindauer, "Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges," *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery*, vol. 13, 01 2023.

[8] H. T. Nguyen, H.-Y. Jeong, K.-D. Kim, N. Xuan-Mung, N. Dao, H.-N. Tran, V.-K. Hoang, N. N. Anh, and M. T. Vu, "A new hyperparameter tuning framework for regression tasks in deep neural network: Combined-sampling algorithm to search the optimized hyperparameters," *Mathematics*, vol. 12, pp. 3892–3892, 12 2024.

[9] C. Aviles Toledo, M. M. Crawford, and M. R. Tuinstra, "Integrating multi-modal remote sensing, deep learning, and attention mechanisms for yield prediction in plant breeding experiments," *Frontiers in Plant Science*, vol. 15, p. 1408047, 2024.

[10] L. Franceschi, M. Donini, V. Perrone, A. Klein, C. Archambeau, M. Seeger, M. Pontil, and P. Frasconi, "Hyperparameter optimization in machine learning," *arXiv preprint arXiv:2410.22854*, 2024.

[11] S. Barke, C. Poelitz, C. Negreanu, B. Zorn, J. Cambronero, A. Gordon, V. Le, E. Nouri, N. Polikarpova, A. Sarkar, B. Slininger, N. Toronto, and J. Williams, "Solving data-centric tasks using large language models," in *Findings of the Association for Computational Linguistics: NAACL 2024*, pp. 626–638, Association for Computational Linguistics, 2024.

[12] H. Zhang, Y. Dong, C. Xiao, and M. Oyamada, "Jellyfish: Instruction-tuning local large language models for data preprocessing," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 8754–8782, Association for Computational Linguistics, 2024.

[13] L. R. Khot, S. Sankaran, A. H. Carter, D. A. Johnson, and T. F. Cummings, "Uas imaging-based decision tools for arid winter wheat and irrigated potato production management," *International Journal of Remote Sensing*, vol. 37, no. 1, pp. 125–137, 2016.

[14] C. Singha and K. C. Swain, "Rice and potato yield prediction using artificial intelligence techniques," in *Internet of Things and Analytics for Agriculture, Volume 3* (P. K. Pattnaik, R. Kumar, and S. Pal, eds.), pp. 185–199, Springer, 2022.

[15] K. A. Al-Gaadi, A. A. Hassaballa, E. Tola, A. G. Kayad, R. Madugundu, B. Alblewi, and F. Assiri, "Prediction of potato crop yield using precision agriculture techniques," *PLOS ONE*, vol. 11, no. 9, p. e0162219, 2016.

[16] F. Ferracina, B. Krishnamoorthy, M. Halappanavar, S. Hu, and V. Sathuvalli, "Predictive analytics of varieties of potatoes," *arXiv preprint arXiv:2404.03701*, 2024.

[17] D. Gómez, P. Salvador, J. Sanz, and J. L. Casanova, "Potato yield prediction using machine learning techniques and sentinel 2 data," *Remote Sensing*, vol. 11, no. 15, p. 1745, 2019.

[18] D. Gómez, P. Salvador, J. Sanz, and J. L. Casanova, "New spectral indicator potato productivity index based on sentinel-2 data to improve potato yield prediction: a machine learning approach," *International Journal of Remote Sensing*, vol. 42, no. 9, pp. 3426–3444, 2021.

[19] D. Li, Y. Miao, S. K. Gupta, C. J. Rosen, F. Yuan, C. Wang, L. Wang, and Y. Huang, "Improving potato yield prediction by combining cultivar information and uav remote sensing data using machine learning," *Remote Sensing*, vol. 13, no. 16, p. 3322, 2021.

[20] S. K. Towfek and A. A. Alhussan, "Potato production forecasting based on balance dynamic biruni earth radius algorithm for long short-term memory models," *Potato Research*, vol. 67, no. 4, pp. 1927–1963, 2024.

[21] A. A. Alhussan, D. S. Khafaga, M. Abotaleb, P. Mishra, and E.-S. M. El-Kenawy, "Global potato production forecasting based on time series analysis and advanced waterwheel plant optimization algorithm," *Potato Research*, vol. 67, no. 4, pp. 1965–2000, 2024.

[22] P. Mishra, A. A. Alhussan, D. S. Khafaga, P. Lal, S. Ray, M. Abotaleb, K. Alakkari, M. M. Eid, and E.-S. M. Elkenawy, "Forecasting production of potato for a sustainable future: Global market analysis," *Potato Research*, vol. 67, no. 4, pp. 1671–1690, 2024.

[23] K. Alibabaei, P. D. Gaspar, and T. M. Lima, "Crop yield estimation using deep learning based on climate big data and irrigation scheduling," *Energies*, vol. 14, no. 11, p. Article 11, 2021.

[24] R. Raymundo, S. Asseng, R. Robertson, A. Petsakos, G. Hoogenboom, R. Quiroz, G. Hareau, and J. Wolf, "Climate change impact on global potato production," *European Journal of Agronomy*, vol. 100, pp. 87–98, 2018.

[25] N. Mishra, S. Mishra, and H. K. Tripathy, "Rice yield estimation using deep learning," in *Innovations in Intelligent Computing and Communication* (M. Panda, S. Dehuri, M. R. Patra, P. K. Behera, G. A. Tsihrintzis, S.-B. Cho, and C. A. C. Coello, eds.), pp. 379–388, Springer International Publishing, 2022.

[26] DeepSeek-AI, "Deepseek-v3 technical report," *arXiv preprint arXiv:2412.19437*, 2024.

[27] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," *arXiv preprint arXiv:2012.07436*, 2021.

[28] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, "N-beats: Neural basis expansion analysis for interpretable time series forecasting," *arXiv preprint arXiv:1905.10437*, 2020.

[29] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y. Wang, and X. Yan, "Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting," *arXiv preprint arXiv:1907.00235*, 2020.

[30] C. Challu, K. G. Olivares, B. N. Oreshkin, F. Garza, M. Mergenthaler-Canseco, and A. Dubrawski, "N-hits: Neural hierarchical interpolation for time series forecasting," *arXiv preprint arXiv:2201.12886*, 2022.

[31] F. Kratzert, D. Klotz, G. Shalev, G. Klambauer, S. Hochreiter, and G. Nearing, "Towards learning universal, regional, and local hydrological behaviors via machine-learning applied to large-sample datasets," *arXiv preprint arXiv:1907.08456*, 2019.

[32] G. Zerveas, S. Jayaraman, D. Patel, A. Bhamidipaty, and C. Eickhoff, "A transformer-based framework for multivariate time series representation learning," in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, (New York, NY, USA), p. 2114–2124, Association for Computing Machinery, 2021.

[33] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[34] E.-S. M. El-Kenawy, F. H. Rizk, A. M. Zaki, M. E. Mohamed, A. Ibrahim, A. A. Abdelhamid, N. Khodadadi, E. M. Almetwally, and M. M. Eid, "Football optimization algorithm (fboa): A novel metaheuristic inspired by team strategy dynamics," *Journal of Artificial Intelligence and Metaheuristics*, vol. 8, no. 1, pp. 21–38, 2024.

[35] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 4, pp. 1942–1948, 1995.

[36] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016.

[37] D. Simon, "Biogeography-based optimization," *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 6, pp. 702–713, 2008.

[38] S. Mirjalili, S. Mirjalili, and A. Hatamlou, "Multi-verse optimizer: a nature-inspired algorithm for global optimization," *Neural Computing and Applications*, vol. 27, 2015.

[39] X.-S. Yang, "A new metaheuristic bat-inspired algorithm," in *Studies in Computational Intelligence*, vol. 284, 2010.

[40] H. Salimi, "Stochastic fractal search: A powerful metaheuristic algorithm," *Knowledge-Based Systems*, vol. 75, 2014.

[41] R. Storn and K. Price, "Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, pp. 341–359, 1997.

[42] X. Wang, V. Snášel, S. Mirjalili, J.-S. Pan, L. Kong, and H. A. Shehadeh, "Artificial protozoa optimizer (apo): A novel bio-inspired metaheuristic algorithm for engineering optimization," *Knowledge-Based Systems*, vol. 295, p. 111737, 2024.