International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 5, Issue 6, November- December 2023 |

DOI:10.15662/IJEETR.2023.0506021

AI-Augmented Big Data Analytics for Precision Agriculture Yield Forecasting

Shashikala Valiki

Independent Researcher, India

ABSTRACT: Significant investments in data collection through remotely-sensing and meteorological stations, the creation of soil health cards for Indian states, and the installation of crop health sensors now make it possible to apply Big Data and AI technologies in agriculture. Forecasts of crop yields made before harvesting of the respective crop, rather than estimates that follow the harvest, can guide import-export strategies, thereby enhancing food security. An architecture is proposed to perform AI-augmented Big Data analytics for predicting yields before harvest, using Big Data from various sources as well as statistical, machine learning, and deep learning methodologies. Once the required analytics are built and validated for a given region, the architecture supports seamless generalization to other regions—both within India and across the globe. Cross-regional generalization requires harnessing the large volume of analytics built across multiple combinations of yield-target crops, provinces, stations, and years in India.

Support for cross-regional generalization come from predictions being fed into virtually equipped ensembles, which subsequently report the winning model for predicting agronomically pertinent features of the region, namely Temperature, Evapotranspiration, Soil Moisture, Normalized Different Vegetation Index—either in combination or all together—and Product of Precipitation in Dry Spells—also either in combination or all together. The need for extensive and costly ground-truth measurements is then overcome through the widespread establishment of remotely-sensed data and other data sources for parts or for most regions of the world.

KEYWORDS: Key phrases of relevance: precision agriculture; yield forecasting; decomposition analysis; statistical modelling; machine learning; artificial intelligence; big data; remote sensing; crop appearance; AI-augmented solutions.

I. INTRODUCTION

AI-augmented big-data analytics promises to transform Precision Agriculture by generating accurate yield predictions over multiple leaders using remotely-sensed heterogeneous weather and environmental datasets. Accurate yield predictions are essential for supply-chain management. However, models need to be trained on historical data to be generalizable. By combining ground-level yield data for 2011–2020 available at the China/US border region with a comprehensive set of meteorological, remote-sensing, and ground-survey datasets, and employing permutational feature importance analysis, early-season yield predictions for maize, soybeans, and wheat across the entire continental USA were made. Benchmarking against independent ground-truth data from the rest of the USA demonstrated that the strongest models could predict corn yield at low levels of mean absolute error (5.6 billion kg) and mean absolute percentage error (3.98%)—both of which were inferior to the outputs from widely-used regression-based methods.

Predicting agricultural yield using machine learning on historical data is challenging due to the limited availability of ground-truth information, for example, at continental scales or for specific crops or regions. Cross-regional generalization using transfer-learning techniques represents an alternative but often faces obstacles such as reduced availability of training data, region-specific non-parametric behavior, over-fitting issues, and suboptimal architectures. Cross-validation across a set of independent future years is therefore essential for assessing the reliability and utility of a learning model for end-users and policymakers.

1.1. Background and Significance

Precision agriculture (PA) is a data-driven approach to farming, enabling increased yields while ensuring sustainability. Large volumes of data generated by multiple sensors, deployed in sensor networks, satellites, and weather stations, provide better insights into plant behavior and crop growth. However, the increased complexity of these sensor-generated big-data in agriculture requires appropriate data-analytics technologies to mine meaningful information in real time. Forecasting agricultural yield at the crop-cycle level is a critical data-mining task, as crop yield is directly related to food-security imperatives for a burgeoning human population.

International Journal of Engineering & Extended Technologies Research (IJEETR)
|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |
| Volume 5, Issue 6, November- December 2023 |
DOI:10.15662/IJEETR.2023.0506021

Yield-forecast innovation hinges on accurately exploiting heterogeneous and multimodal data streams, preferably in a scalable way. In real-world environments, sophisticated methods for integrating heterogeneous data streams for professional decision-making have been scarce. A combination of different data-science-based approaches is applied here for yield forecasting: classical time-series-based and regression techniques; machine learning (ML); deep learning (DL); and the novel AI-Augmented Analytics ecosystem, which exploits the big-data explosion in agriculture for enhanced decision support.

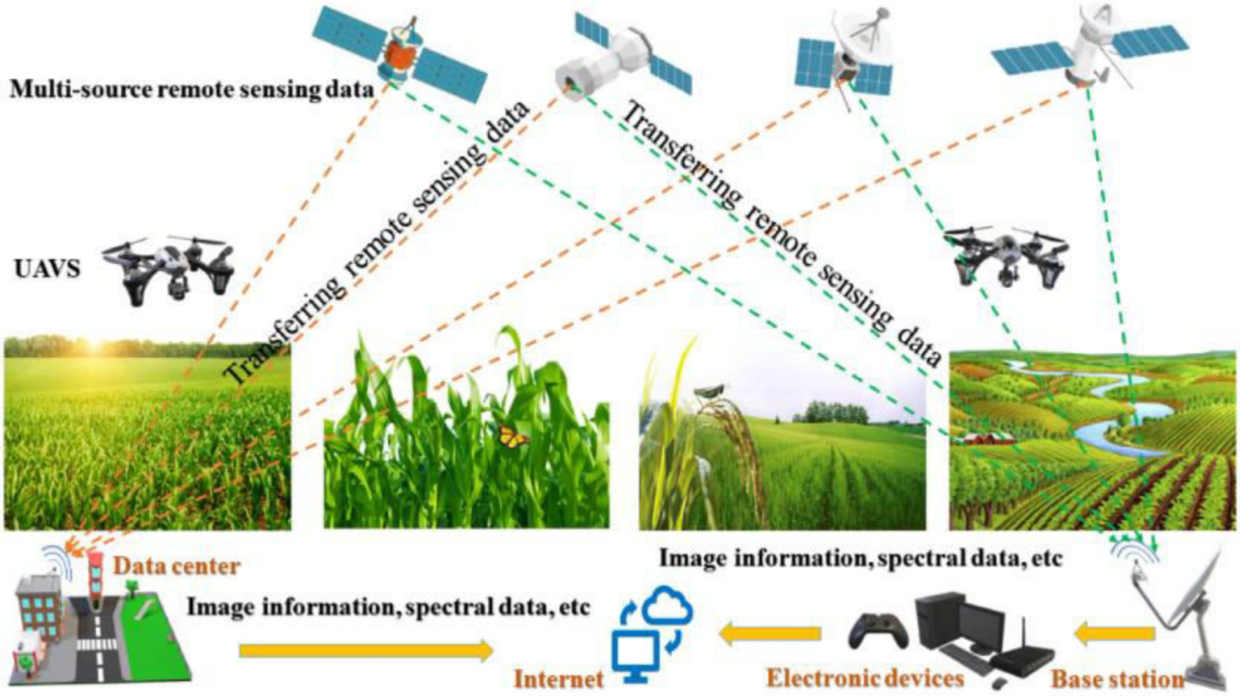


Fig 1: Integration of Remote Sensing for Precision Agriculture

1.2. Research design

The yield of a given crop in a geographical location is directly related to the spatiotemporal dynamics of salient soil, meteorological, and other environmental conditions. In the past, expert agronomists would study historical data to arrive at a forecast that would direct farming and supply-chain decisions. However, using statistical analyses on the time series of many spatiotemporal variables, Deep Learning offers a way to automate this forecasting. Moreover, the emergence of data-intensive AI has brought an even larger and richer volume of data into play.

The AI-augmented analysis facilitates the ingestion of the constantly evolving Big Data from an array of disparate sources—remote sensing satellites, ground stations, soil sensors, crop health sensors, and so on—for predicting the crop yield in a given location at multiple time resolutions, including daily, weekly, fortnightly, and in fact any user-specified period—using any combination of lateral support and historical data. In this modelling exercise, the publicly available ICAR data for soil nutrients for water-soluble N, P, K, S, Zn, Fe, Mn, and Cu in combination with NDVI as the lateral support are used to generate three different yield predictors: one using only the latest NDVI as predictor, another using last three years of NDVI as predictors, and the third combining the latest soil nutrient distribution with last three years of NDVI time series as predictors. These kernels are benchmarked against the yield obtained from on-ground horticultural experts and against the latest season of yield data for cross-region generalisation.

Equation 1: NDVI (Normalized Difference Vegetation Index)

Let:

- R_{NIR}= surface reflectance in Near-Infrared band
- R_{RED}= surface reflectance in Red band

Step-by-step derivation

1. Vegetation reflects **more NIR** and absorbs **more Red**, so a basic contrast is:

$$D = R_{NIR} - R_{RED}$$

2. But Ddepends on brightness/illumination (scale). Normalize it by total energy:

$$S = R_{NIR} + R_{RED}$$

3. Define a scale-free ratio:

$$NDVI = \frac{D}{S} = \frac{R_{NIR} - R_{RED}}{R_{NIR} + R_{RED}}$$

4. Range intuition:

○ If vegetation is strong: $R_{NIR} \gg R_{RED} \Rightarrow NDVI \to 1$

If barren/water/cloud: values tend toward 0 or negative

II. BACKGROUND

2.1. Precision Agriculture and Yield Forecasting

Precision agriculture uses real-time estimates of crop yield forecasts to support precision farming decision making. Accurate crop yield forecasts support grain trade, investment decisions, national security, and food price prediction. Precise and accurate crop yield forecasts maximize the productivity and profitability of farmers and contribute to improved economic growth of the country.

Supported by recent advances in big data and sensor technology, precision agriculture monitors spatial variation at low costs, allowing a detailed real-time understanding of crop growth dynamics. Big data research exploits low-cost remote sensing, high-resolution weather monitoring, measurement of agronomic parameters through sensors, and appropriate utilization of statistical and machine learning models to provide accurate crop yield forecasts at different resolutions.

2.2. Big Data in Agriculture

Recent advances in low-cost remote sensing, monitoring of high-resolution weather data, the use of soil and crop health sensors, and the development of machine learning and statistical methods that solely require high-dimensional data of many variables with low samples have led to breakthroughs in the area of big data research in agriculture. Low-cost remote sensing provides spatial data of multiple spectral bands and vegetation index at high temporal resolution every 8 days, thus creating an opportunity to explore the temporal growth dynamic of crops during the cropping season.

The availability of real-time and high-resolution weather data has facilitated the estimation of phenological growth stages. In the past, crop yield predictions were based mainly on statistical models such as the Agri-Met SAM, United States Department of Agriculture (USDA), and the Crop Environment Resource Synthesis model; now, machine learning techniques have recently gained attention among researchers and decision makers. Additionally, models that make use of many agronomic variables and low samples (artificial neural networks and deep learning) have gained popularity within the community.

2.1. Precision Agriculture and Yield Forecasting

The cornerstone objective of precision agriculture is to increase the yield and economic returns while reducing the environmental impact of agrarian practices. Yield forecasting constitutes an integral part of this strategy and serves several applications, including monitoring imminent yields for food security, guiding resource allocation and ecosystem services, and supporting producers in decision making, financing, and insurance activities. However, traditional yield forecasting models based on statistical regression of ground-observed crop yield statistics and ancillary variables have frequently either overestimated or underestimated crop yields, mostly because the crop yield statistics do not provide current or future yielding status.

To address the severe limitations of crop statistics, several approaches have employed remote-sensing data and statistical models. Major advances have been achieved with normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), and precipitation. There are also examples of large-sample studies based on near-real-time satellite information over extensive fields. More recently, machine learning techniques have been explored, generally exploiting NDVI and other spectral indices. Nevertheless, such studies risk continuity loss due to the presence of cloud cover or sensor failure.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 5, Issue 6, November- December 2023 |

DOI:10.15662/IJEETR.2023.0506021

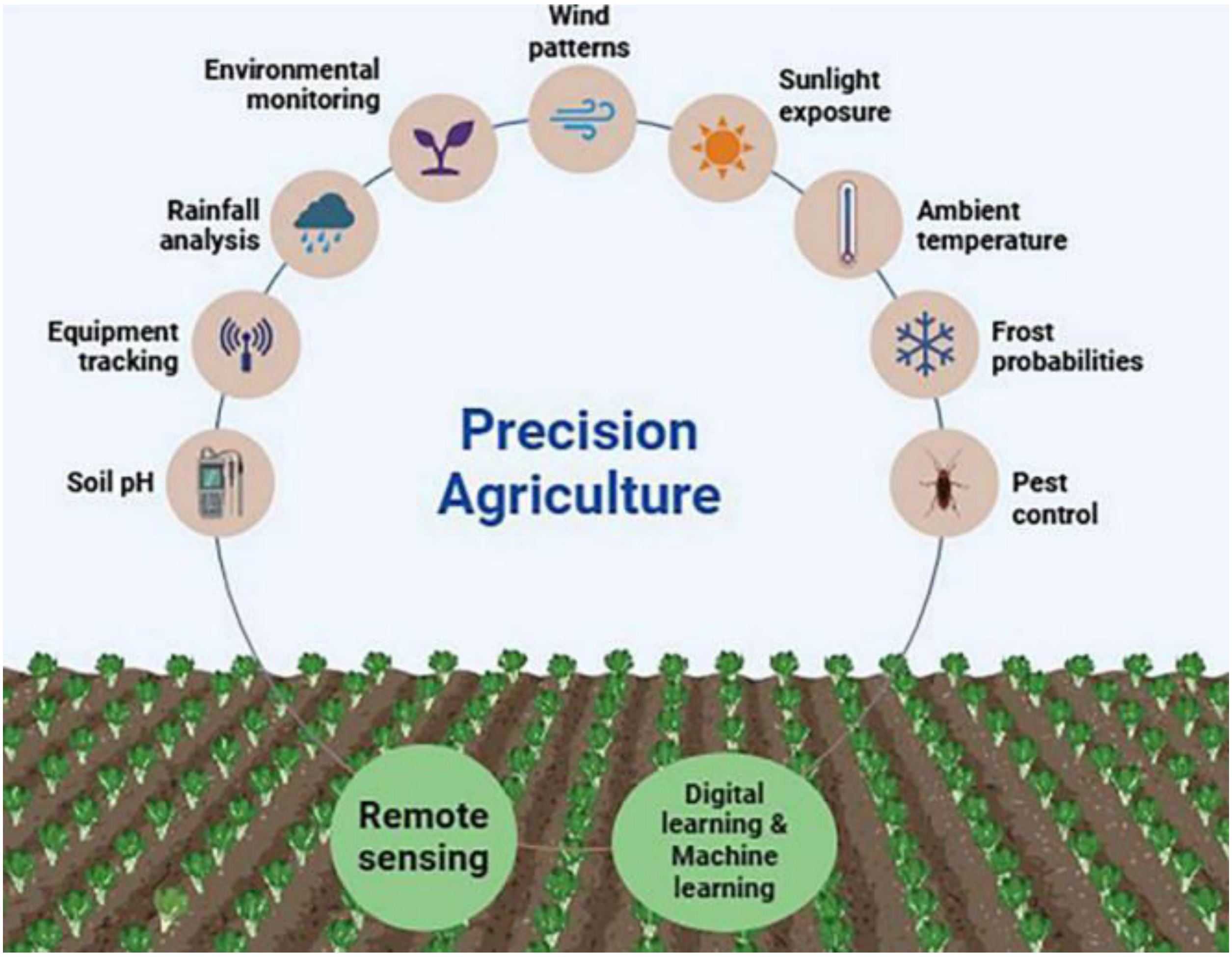


Fig 2: Precision agriculture for improving crop yield predictions

2.2. Big Data in Agriculture

Natural and anthropogenic processes govern global weather events that can affect agriculture from local to global scales and impact on food security. Climate change has raised the frequency of drought, flood, and cyclones, with regions like Bangladesh facing devastating events. Governments often declare calamities and financial aid helps farmers recover and retain food output. Tremendous hassles emerge, including loss of human lives and properties, crop failure, food price increase, and inflation. Political unrest, ethnic violence, and refugee issues can also arise. Big Data Analytics Technology in Cloud will intelligently predict climate change event occurrences using data mining technique.

The development of IoT has led to the decrease of deployment cost of remote sensor devices. Multiple low-cost remote sensors can be deployed in the field to capture the field environment at different developmental stages. It can be then integrated with meteorological data for different predictive analytical methods for real-time monitoring of disease and pest outbreaks, site-specific safety alerts for the farming community, site-specific quick response to control disease and pest, and assessing yield loss prospects of crops grown over a large area. The research effort can be extended in tandem with ground truth observations during disease, and pest attack conditions to build machine learning models and deep learning models that take embedded temperature, rainfall, and humidity level in too long to train convolutional neural networks (CNNs).

Equation 2: EVI (Enhanced Vegetation Index)

Let:

- R_{BLUE}= reflectance in Blue band
- Typical constants: G(gain), C_1, C_2(aerosol resistance), L(canopy background)

Step-by-step construction

1. Start with the same vegetation contrast numerator:
$$R_{NIR} - R_{RED}$$

2. Correct the denominator for aerosol influence using Blue:
$$R_{NIR} + C_1 R_{RED} - C_2 R_{BLUE} + L$$

3. Apply gain G:
$$EVI = G \cdot \frac{R_{NIR} - R_{RED}}{R_{NIR} + C_1 R_{RED} - C_2 R_{BLUE} + L}$$

III. DATA SOURCES AND INTEGRATION

The increasing availability of Big Data in agriculture from multiple sources opens new opportunities for near-term yield forecasting at substantial geographic scales. Remote sensing provides coverage of large areas at high spatial and temporal resolutions using multiscale and multisource data. Optical sensors offer valuable spatial information of land-use and land-cover patterns and relationships across the land surface. Radar instruments capture surface and vegetation structural variability and status. The assimilation of optical and active microwave data also enables the retrieval of soil moisture. Synthetic Aperture Radar (SAR) and interferometric SAR systems allow the mapping of elevation and inundation dynamics, supporting the modelling of hydrological processes. Numerous meteorological datasets support crop growth models with near-real-time climate information.

Furthermore, a growing network of in-situ soil and crop health sensors provides critical data on the nutrient status of the soil and plants at commensurate spatial scales. These sensor data, together with the remote-sensing products, can also help to identify more-sparsely distributed agricultural activities. Recently, a robust and generalised regression modelling framework was proposed to predict crop yields using multiple sources of Big Data. By creating a Big Data layer for the current and forecasted crop state, this approach aims at predicting yields up to three months in advance. Yet, due to the lack of training datasets covering different types of crops across large regions, yield models based on remote-sensing and meteorology data cannot simply extrapolate across space. Therefore, prior to cross-regional predictions, model predictions need to be validated and possibly calibrated with independent observations. NDVI and LST fusion products from this advanced geostationary–polar-orbit combined satellite system successfully predict spatiotemporal crop yield variation based on the regular growth stages of major crops and the weather events of inclement floods, droughts, dry-hot winds, and typhoons. Meteorological data from Chinese national meteorological observatories have been collated, archived, and published in recent years, providing effective ground-truth supporting datasets for yield prediction work at various scales across China.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 5, Issue 6, November- December 2023 |

DOI:10.15662/IJEETR.2023.0506021

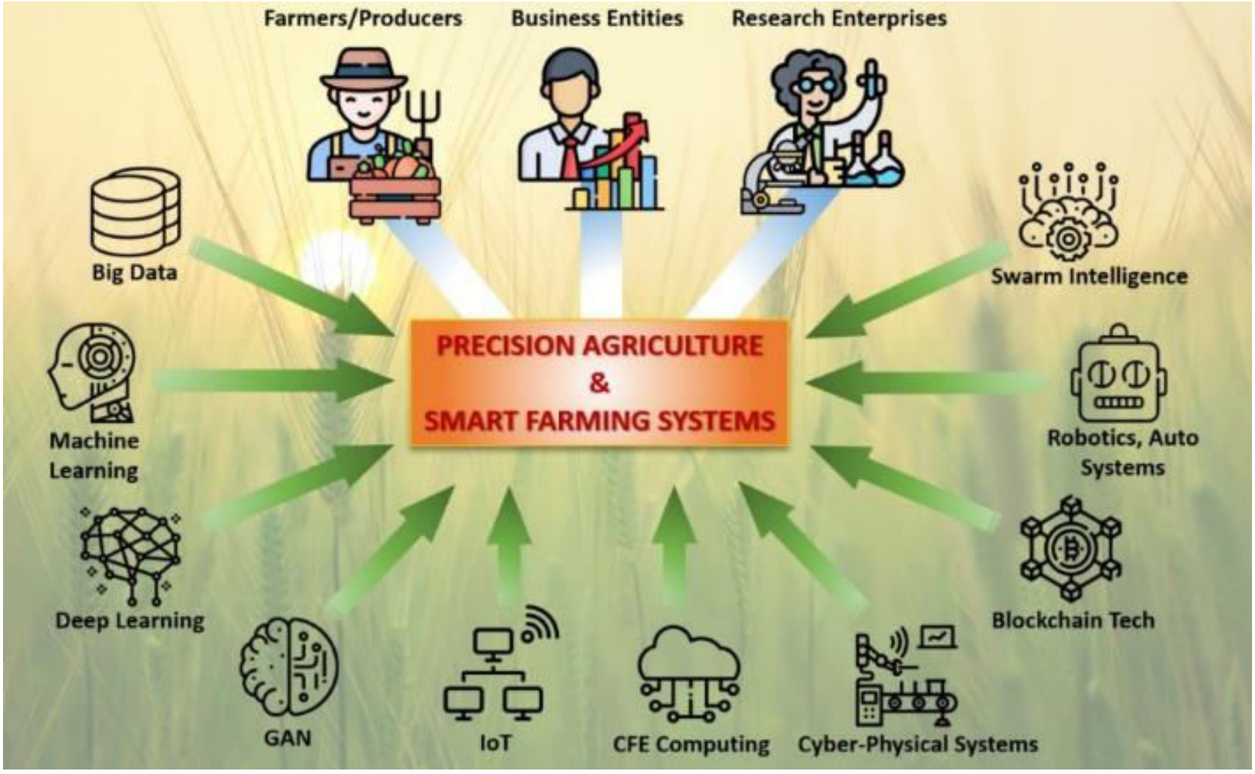


Fig 3: AI and Data Analytics are Shaping the Future of Precision Agriculture

3.1. Remote Sensing and Meteorological Data

Continuous time series of meteorological data, including temperature, precipitation, and solar irradiance during the growing season, are known to substantially influence the yield of agricultural crops. Remote sensing data, such as the normalized difference vegetation index (NDVI) or other vegetation-growth indices from various sensors, have been widely used for yield forecasting and disease detection. Vegetation indices from single satellite observations can improve the accuracy of crop yield estimations by about 5%, while multi-sensor, multi-temporal, and cloud-free NDVI data obtained from the fusing of optical sensors with higher spatial resolution and radar sensors with all-weather capability can permit a significant improvement of about 15%. In order to capture multi-scale crop status and provide an important auxiliary factor for yield forecast, the fusion of spatially continuous and temporally high-resolution remote-sensing products is essential.

Prior studies have demonstrated that the yield dynamics of major crop species can be delineated with better accuracy using remotely sensed data, including phenology, end-of-season biomass, and NDVI robustness. Satellite NDVI at a coarse spatial scale can be employed as an overarching combined NDVI but will have limited explanatory power and predictive capability if used at a finer scale. However, the fusing of land surface temperature (LST) and NDVI from coupled geostationary–polar-orbit satellite data combines their complimentary strengths and features, and yields much superior results.

3.2. Soil and Crop Health Sensors

Ground-based sensors can capture information beyond what is possible from space or air. Soil moisture sensors facilitate irrigation decision, nutrient sensors measure aspect and concentration levels for control as intervention in fertilizer and other plant growth regulators application both in time and quantity can avert loss of yield as a consequence due to toxicity or surplus application of fertilizers. The visible near infrared spectrometry augments nutrients analysis and detection of stress conceived with chemical imbalance in crops due to plant disease, deficiency of nutrients, insect/pest attack or herbicide injury. The ultra-sonic sensor for plant health monitors the overall health status related to plant biology, leaf age and leaf area and is helpful in growing plants stress-free, decaying-free and hurtea-free, which ultimately helps in producing good yield by minimizing the harvest loss. Temperature measuring sensors cool down the plant to check wilting, dry leaves and to generate fruiting. Humidity measuring sensors check the humidity effect produced inside of the crop to avoid wilting or leaf-falling. pH sensor controls the acidity and alkalinity

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 5, Issue 6, November- December 2023 |

DOI:10.15662/IJEETR.2023.0506021

of soil to overcome nutrient imbalance in the soil. NDVI sensors control crop health and promote crop growth rate to enhance the production of crops by staying away from the incursion and impact of the fungus.

Agricultural yield data can be collected and assimilated from various public sources and relevant information is provided for crop yield prediction. Weather parameters and cultivated area are accumulated from India Meteorological Department (IMD) and National Remote Sensing Centre (NRSC) and India's National Academy of Agricultural Sciences. Data from IMD has been supported with Weather Research Forecast (WRF) and Automatic Weather Station (AWS)-based predictions for few districts/crops. Soil Health Card data and crop yield have been downloaded from the India Government's Department of Agriculture and Farmers Welfare and Ministry of Agriculture and Farmers Welfare (MoAFW) website. NDVI data were procured from MODIS, retrieved from Terra and Aqua satellites of National Aeronautics and Space Administration (NASA). GOES imager rainfall content as an additional input for certain crops in the northeastern region was used as a prime effect factor. The India Meteorological Department (IMD) provides meteorological information for climate-based statistical prediction.

Equation 3: Growing Degree Days (GDD) and cumulative GDD
Let:

- Daily maximum/minimum temperatures: $T_{\max}(d), T_{\min}(d)$
- Base temperature for crop development: T_{base}

Step-by-step

1. Daily mean temperature:

$$T_{mean}(d) = \frac{T_{\max}(d) + T_{\min}(d)}{2}$$

2. "Effective heat" above base:

$$g(d) = T_{mean}(d) - T_{base}$$

3. Crop growth doesn't go negative; clamp at 0:

$$GDD(d) = \max(0, g(d))$$

4. Cumulative over a season (days $d = 1 \dots D$):

$$CGDD = \sum_{d=1}^{D} GDD(d)$$

IV. METHODOLOGIES FOR YIELD FORECASTING

Forecasting yield in precision agriculture is a complex task traditionally addressed by statistical models of univariate or bivariate historical relationships between climate and harvest. More recent approaches exploit big data analytics and AI methodologies, especially with machine learning and deep learning. Machine learning models usually leverage exploratory data analysis (EDA) and expert knowledge to reduce the feature space size to a configurationally manageable level and focus on the most relevant inputs for yield prediction, sometimes also using other censuses over the same geographical area, such as soil-quality information or crop-health multispectral sensing data, helping to shed light on the yield response to climate variations and disturbances.

Deep learning, in contrast, can use the entire set of features—indeed, the more noise the better—but tends to require a large amount of data, as well as the FT for the geographical area. Evidence shows, however, that deep learning approaches can yield low error predictions—with suitable generalization power even though they are trained only on agronomic CL data at relatively low density and without the support of other censuses or noise sources, provided that they are properly regularized and validated. The different methodologies used for yield forecasting clustering the algorithms into statistical, machine learning, and deep learning models are outlined.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 5, Issue 6, November- December 2023 |

DOI:10.15662/IJEETR.2023.0506021

Fig 4: Crop yield data in agriculture

4.1. Statistical Approaches

Classical yield forecasting models are based on regression analysis of meteorological variables, trend functions, and Huet's index of climatic seasonality. Cumulative growing degree days are also being used in regression-based yield forecasting studies for nearly all crops. M. S. Srinivasan uses Long Range Suggestion Model (LRSM) to forecast paddy yield on a year-to-year basis. Multiple regression analysis is used with secondary data for paddy crop-and-distance-wise estimation of yield and crop area under rice transplanting in order to suggest a forward-looking paddy development program for the state of Kerala. A regression model based on time series data for 1975-76 to 1994-95 has been developed to explain and forecast the area and production of groundnut in Andhra Pradesh.

Regrettably, conventional regression-based models fail to explain food grain yields in some regions of India; Cross-Validation Difference in Difference (CVDID) analysis and the widely accepted Nested Logit model confirm their ineffectiveness in some areas. Therefore, there is a need to develop reliable forecasting methods to alleviate dependence on inappropriate techniques. In this regard, time series data for 1980-81 to 2010-11 were statistically analysed to forecast the area and production of maize in Karnataka, and forecasted estimates were compared with those obtained by other forecasting groups. The SATC model indicated the strong influence of government efforts on the achievement of maize production in Punjab during 2000-01. A Logit-type probabilistic econometric model, with time series data, has been developed for forecasting food production thanks to its roots in economic theory and subsequent generalization, yielding forecasts faithful to the underlying theory.

4.2. Machine Learning and Deep Learning Models

Despite the appeal of using machine learning and deep learning algorithms for near real-time precision agriculture yield forecasting, limited large and regionally comprehensive datasets have constrained progress. Crowdsourced social media data provide a possible way to forecast crop yield through unstructured sources, but such predictive models are computationally expensive and less interpretable. Successful framework demonstrations with hybrid models trained on large datasets across multiple areas have only emerged recently, along with forecast generation for different regions. Yet the nested deep neural network precision agriculture yield model remains domain-knowledge limited, as does the variety of active-feature-rank-based hybrid model focused on a single region of China.

Nonetheless, regional evaluation of the L-BFGS-optimized deep-learning model for hybrid urban-vegetation scheme yield modeling indicates general suitability for other regions in the southern third of China that share suitable climate conditions, especially abundant soil water transfer capability. Recent models coupling deep-learning with Tacotron-2 and global crowd-sourcing networks also leverage social media with fine-tuned datasets to balance forecasting precision and execution speed within a full-cycle framework. Here, the Loop–King model constructs accurate near real-time sweet-potato yield forecasting through dynamic pairing of social-media LULC and global crown-scale remotely sensed dataset. Machine learning and deep-learning meta-algorithm models boast better season efficiency and validation performance than single models, as illustrated in cross-regional forecasting. A machine-learning-based ensemble modeling approach combined with self-organizing maps and PCA analysis further supports reliable and interpretable crop-yield forecasting.

Equation 4: Evapotranspiration (ET) as a feature
Energy balance + aerodynamic transport
1. Evaporation demand rises with **net radiation** and dries with **wind/vapor deficit**.
2. Penman–Monteith combines:
   - energy term (radiation driven)
   - aerodynamic term (wind + humidity driven)

Canonical FAO-56 form:

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T + 273} u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)}$$

V. AI-AUGMENTED ANALYTICS ARCHITECTURE

The architecture for implementing AI-augmented big data analytics is discussed, focusing on subcomponents that enhance user productivity in making harvest predictions. Two areas that capitalize on AI augmentation are featured: 1) a distributed framework to facilitate the collection, storage, and processing of data from multiple sources at a scalable rate; and 2) an automatic feature engineering capability to generate agronomically meaningful predictors that span space, time, and variables beyond the yield data.

Earlier research demonstrated how yield data sourced from United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) crop reports could be leveraged as a ground reference for validating analytical approaches used in creating remote-sensing-based NDVI time series for multiple crops through multiple growing seasons. For the Iowa study, the NASS data provided an inherent benchmark; however, due to the aggregation of yield data at the state level, the utilization of other USDA sources—NASS Crop Scanning Survey data—provided the missing link to implement an actual forecast at a higher spatial resolution.

The NASS Crop Scanning Survey dataset comprises groundtruth yield data collected by USDA–NASS field enumerators. Essentially an aerial vantage point survey, yield estimates of specific crop types are generated for areas comprising between 3,500 and 7,500 acres (1,400 and 3,000 ha) using ground-level observations and a visual assessment of potential yields for each sampled field. Positioning within a single county or statistical area is not mandatory, although most observations are, with large feature fields favoring higher accuracy.

The precision-yield-forecasting line of research combines ground-tested econometric models with deep learning (DL) technologies to forecast crop yield at enhanced spatial resolution by making use of remote-sensing and meteorological data readily available in the cloud for the entire globe.

5.1. Data Ingestion and Storage

The proposed analytical architecture leverages a cloud data lake to ingest and store Big Data from multiple sources that support yield forecasting at multiple scales. A new scalable architecture built upon readily available cloud platforms serves as the backbone for data storage and analytics. The architecture uses an enterprise-class data lake and sets up three specific layers: an ingestion layer, a storage layer, and an analytics layer. The framework draws data from heterogeneous sources into the cloud-based data lake. Data ingestion uses multiple cloud-native services—such as AWS Data Pipeline, AWS Glue—to transform raw data and usable features in the cloud. The cloud data lake enables (a) efficient local, regional or national-scale data analytics across the available data and applied trained ML/DL algorithms; and (b) easy ML/DL model updates as new data become available. Region-specific yield forecasting and the use of ground truth data from nearby regions with similar crop-growing patterns yield capable ML/DL models that can extrapolate and produce increased spatial yield information density with a coarser spatial resolution.

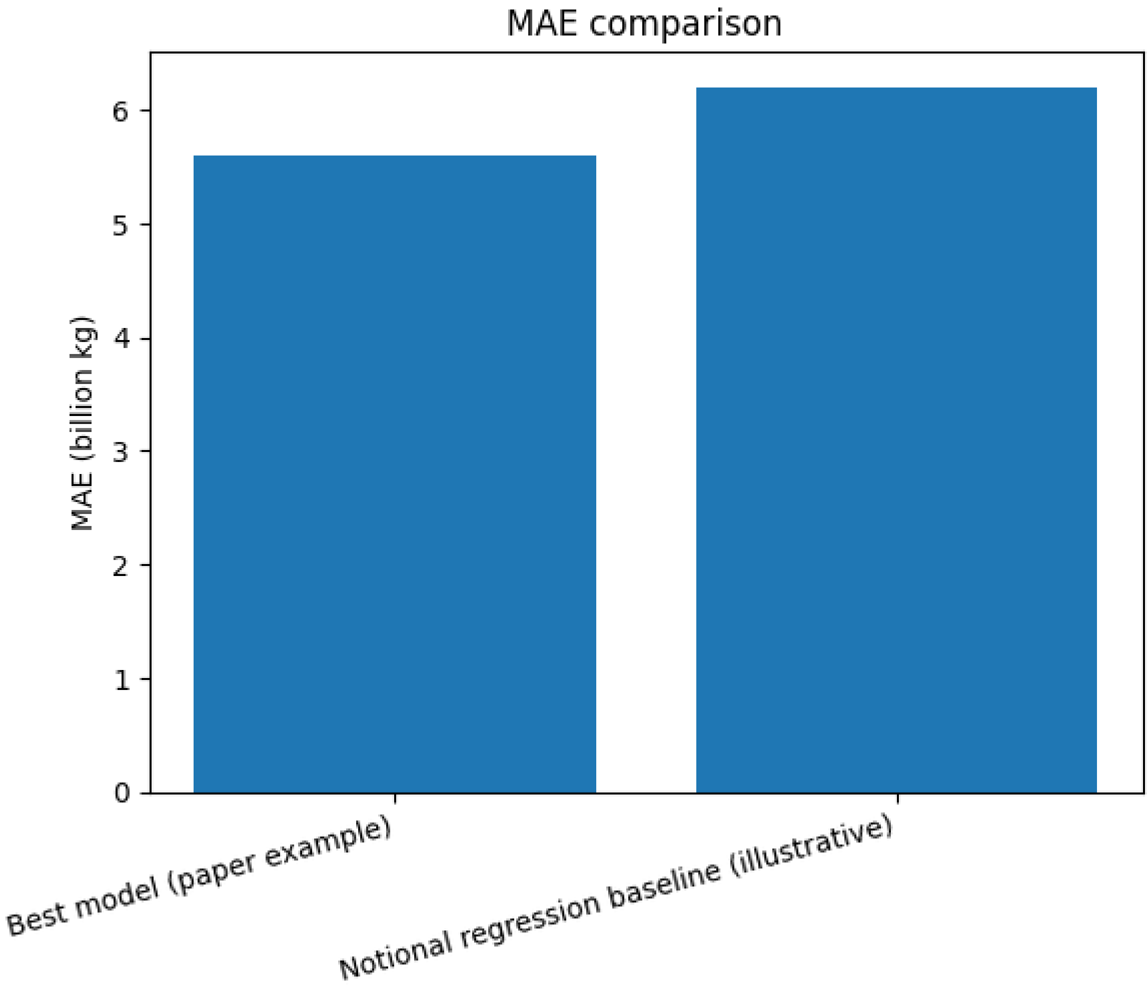
The proposed YOLO-PAT framework advances precision agriculture and supports yield prediction, which is essential to address food, water and energy security in the 21st century and beyond. AI-augmented yield forecasting augments

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 5, Issue 6, November- December 2023 |

DOI:10.15662/IJEETR.2023.0506021

traditional data, information and knowledge-based precision-agriculture analytics by leveraging the Big Data available in the cloud. Emerging trends toward a provision-based application model, such as Software-as-a-Service (SaaS) or Data-as-a-Service (DaaS), facilitate the provision of cloud-based analytics.



5.2. Feature Engineering for Agronomic Relevance

Error-prone feature selection is a significant issue when learning yield forecast models with big data. A pipeline that creates several models for each crop's yield using various learning algorithms, then combines the best-performing models is one way to overcome that problem. Feature engineering can apply additional domain-centric ideas. Yields, influences, and corresponding features that were found meaningful in work not using big data can be synthesized for faster development of big-data-based yield forecasts. Coefficients of influence can help guide investigation into the data-driven identification of interactions, transformations, and additional features, particularly from soil and crop health sensors. These extra sensor-generated yield-affecting properties may also be incorporated into explaining agronomic changes in crop yields between regions.

Additional logical, statistical, and machine learning models designed to capture relevant underlying relationships can be created. Beyond combination models leveraging many candidate models provided from shared yield features augmented with sensors, temporal, cross-regional, and cross-application transfer learning can test generality of learnt patterns. Indeed, detection of newly introduced sensor-measured processes for a crop in a region, combining previously learnt data with borders newly entering data, or transferring learnt relationships to a similar crop in a different region use quantity of data to aid development of predictive power.

Equation 5: Soil moisture (simple water-balance feature)

Let:

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 5, Issue 6, November- December 2023 |

DOI:10.15662/IJEETR.2023.0506021

- SM_t=soil moisture at day t
- P_t=precipitation
- I_t=irrigation
- ET_t=evapotranspiration
- R_t=runoff
- D_t=deep drainage

Step-by-step

1. Start from conservation of water in root zone:

$$SM_{t+1} = SM_t + \text{inflows} - \text{outflows}$$

2. Substitute inflows/outflows:

$$SM_{t+1} = SM_t + (P_t + I_t) - (ET_t + R_t + D_t)$$

3. In practice, clip to physical bounds (wilting point to field capacity).

VI. VALIDATION AND EVALUATION

Evaluation of predictive accuracy is a crucial step in the development of yield forecasting models, particularly deep learning models, which are generally considered black boxes with little mathematical and statistical justification. Addressing this challenge requires bridging the gap between satellite- or sensor-based predictions and in situ measurements or ground-truth data. Two approaches for benchmarking predictive accuracy stand out. The first involves direct comparison of pixel-wise predictions against ground-truth harvest yield data, available for thousands of individual fields across multiple years. The second evaluates the generalization performance of models across space and time, predicting yield for relatively large and spatially coherent regions (e.g., states) using limited historical training data.

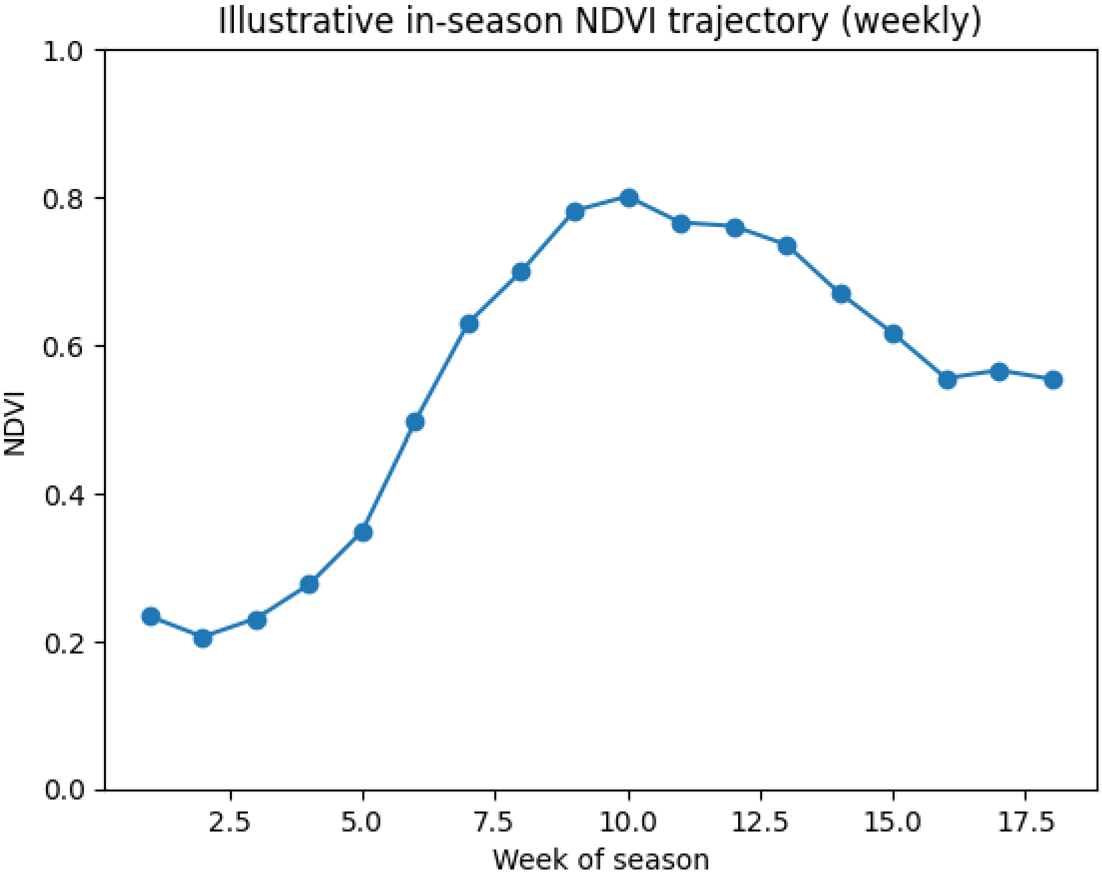
Real-world applications of machine learning and deep learning for yield forecasting can be found in numerous studies, confirming that predictive accuracy often exceeds that of conventional statistical methods. However, the lack of cross-validation and absence of ground-truth comparisons serve as major caveats to such claims. AI-augmented analytics strengthens the reliability of model predictions by providing designed sensor networks and crophealth monitoring, which reduce training input uncertainty and improve accuracy, thereby widening the applicability of machine-learning and deep-learning-based forecasting. Model predictions can also be evaluated against sensor observations for field-scale validation.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 5, Issue 6, November- December 2023 |

DOI:10.15662/IJEETR.2023.0506021



6.1. Benchmarking Against Ground Truth

Either direct comparison against ground truth measures or absolute error thresholds are ser-viceable methods against which specific models can be tested. Ideal comparisons use held-out ground truth values from the target region which have not been employed for training any of the component models in the broader yield forecasting pipelines. Selected datasets offer such opportunities: the publicly available, gridded wheat yields produced by the Indian Remote Sensing Agency for 2013–2014 and 2014–2015; the yield data maintained at the crop reporting district level by the Ministry of Agriculture and Farmers Welfare for the state of Punjab in northwest India from 1985 to 2018; and the yield estimates published in the "District Level Estimates of Major Crops for 2017–18" by the Ministry of Agriculture and Farmers Welfare.

Leveraging the cross-regional applicability of agronomic knowledge, models can be run against resampled predictors from different regions. Generaliza-tion capability is assessed by comparing yields across the Indian states of Punjab and Haryana. In both states, winter wheat is the most important crop; high-resolution yield data from the Indo-Gangetic Plain are also available. In addition, high resolution gridded rainfall data are available publically. These datasets permit an indirect assessment of model performance in Haryana using yields from Punjab and vice versa.

6.2. Cross-Regional Generalization

Cross-regional yield predictions were tested for generalization capacity by using a ML approach to predict maize yields in South America. Modelling without data from Argentina or Brazil confirmed their predictive quality when forecasting maize yields in both countries individually and together. Prediction on a pan-tropical content scale (i.e., including three African countries) further underlined generalization capability, although differences in rainfall seasonality and soil-coarse fractions between African and South American countries complicated simulations for the eastern and western African sites. Successful cross-regional yield forecasts over large geographical extents suggest a potential to predict regional crops not represented in the training dataset.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 5, Issue 6, November- December 2023 |

DOI:10.15662/IJEETR.2023.0506021

The quality assessment of the ML models provided an important validation of yield dynamics across South America when benchmarking simulated against reference yields. While the evaluation based on regional yields was performed here for maize, the strategy could be applied to other crops in the future. Such an approach may become particularly useful when validating projected yields in the context of climate impacts.

VII. CONCLUSION

The research highlights the increasing significance of AI-augmented big-data analytics infrastructures, enabling performance improvements across statistical, machine-learning, and deep-learning approaches. Yield prediction serves as a prime indicator to assess the suitability of the various methods within the AI-augmented big-data analytics architecture. In analysing yield prediction, the resources available are assimilated into a coherent framework to evaluate all three classes of prediction techniques simultaneously. Prediction performance for all three classes of model is examined across diverse corn and soybean-growing regions and through both temporal and spatial cross-validation to address both the persistence and signature of large-scale climatic events.

The overarching aim of this investigation is a proposal for new analysis methods: emerging resources are integrated into established predictive frameworks, and the general potential of these methods as a predictive facilitator of high-resolution precipitation forecasts is explored. The following research questions emerge: RQ1: What are the possible advantages of using AI-augmented infrastructure in analysis methods based on statistics, machine learning, and deep learning? RQ2: When deep learning resources are used to predict large-scale features of the climate system associated with precipitation over the central United States, how can the resulting predictions aid in improving the signature of droughts and pluvial phases? RQ3: How are the signatures of pronounced climate events amenable to prediction? A specific aspect of the investigation validates the cross-regional applicability of the discussed methods by examining yield prediction in locations outside the regions used to train the prediction models.

Field	AUC_NDVI	Yield_t_per_ha
1	9.464942909603668	4.1659846644248875
2	9.643589514925594	4.396275914133352
3	9.476835483684697	3.9288770163064823
4	9.661302807344654	4.086308031374159
5	9.617759514520372	3.77021524737332
6	9.595247669211663	4.746752379618273
7	9.40292180741474	4.19936465651745

Table : Illustrative dataset: NDVI season summary vs yield

7.1. Emerging Trends

Input data from complementary sources is essential to enhancing the accuracy and robustness of yield models. Future studies should focus on the intelligent synthesis of multimodal ground truth data from multiple sources to realize cross-farming generalization capability – that is, validation and application to crop production regions outside the domain of the training dataset. Furthermore, the continuous flow of AI-augmented analysis from Big Data is paving the way for precision agriculture 2.0, covering more delineated and subtle aspects of the decision-making process and outperforming state-of-the-art solutions.

Recent developments talk of agricultural management practices resembling telemetry, with automated warning signals in case of abnormal behavior. The horizon of AI-augmented Big Data pushes toward automatic intelligent translations of such event detections into management actions to increase or maintain yields sustainably at lesser costs both economically and in terms of environmental impacts. The closure of the value chain of AI-augmented Big Data analytics toward the farmers continues to widen the access for localized, affordable, updated solutions thanks to the ubiquity of mobile devices and instant communication.

REFERENCES

1. Siva Hemanth Kolla. (2023). Deep Learning–Driven Retrieval-Augmented Generation for Enterprise ITSM Automation: A Governance-Aligned Large Language Model Architecture. Journal of Computational Analysis and Applications (JoCAAA), 31(4), 2489–2502. Retrieved from https://www.eudoxuspress.com/index.php/pub/article/view/4774
2. Acharya, S., Bhattacharya, S., & Das, S.Agentic artificial intelligence for smart and sustainable precision agriculture. Frontiers in Plant Science, 16, 1706428.
3. Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19–31.
4. Gottimukkala, V. R. R. (2023). Privacy-Preserving Machine Learning Models for Transaction Monitoring in Global Banking Networks. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 633-652.
5. Aljawarneh, S., Aldwairi, M., & Yassein, M. B. (2018). Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model. Journal of Computational Science, 25, 152–160.
6. Kummari, D. N. (2023). Energy Consumption Optimization in Smart Factories Using AI-Based Analytics: Evidence from Automotive Plants. Journal for Reattach Therapy and Development Diversities. https://doi. org/10.53555/jrtdd. v6i10s (2), 3572.
7. Bates, D. W., Saria, S., Ohno-Machado, L., et al. (2014). Big data in health care. Health Affairs, 33(7), 1123–1131.
8. Keerthi Amistapuram. (2023). Privacy-Preserving Machine Learning Models for Sensitive Customer Data in Insurance Systems. Educational Administration: Theory and Practice, 29(4), 5950–5958. https://doi.org/10.53555/kuey.v29i4.10965
9. Belle, A., Thiagarajan, R., Soroushmehr, S. M. R., et al. (2015). Big data analytics in healthcare. BioMed Research International, 2015, 370194.
10. Guntupalli, R. (2023). AI-Driven Threat Detection and Mitigation in Cloud Infrastructure: Enhancing Security through Machine Learning and Anomaly Detection. Available at SSRN 5329158.
11. Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000). LOF: Identifying density-based local outliers. ACM SIGMOD Record, 29(2), 93–104.
12. Unifying Data Engineering and Machine Learning Pipelines: An Enterprise Roadmap to Automated Model Deployment. (2023). American Online Journal of Science and Engineering (AOJSE) (ISSN: 3067-1140) , 1(1). https://aojse.com/index.php/aojse/article/view/19
13. Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. Mobile Networks and Applications, 19(2), 171–209.
14. Meda, R. (2023). Developing AI-Powered Virtual Color Consultation Tools for Retail and Professional Customers. Journal for ReAttach Therapy and Developmental Diversities. https://doi. org/10.53555/jrtdd. v6i10s (2), 3577.
15. Cios, K. J., & Moore, G. W. (2002). Uniqueness of medical data mining. Artificial Intelligence in Medicine, 26(1–2), 1–24.
16. Kummari, D. N., & Burugulla, J. K. R. (2023). Decision Support Systems for Government Auditing: The Role of AI in Ensuring Transparency and Compliance. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 493-532.
17. Dasgupta, D., & Nino, F. (2009). Immunological computation. CRC Press.
18. Varri, D. B. S. (2023). Advanced Threat Intelligence Modeling for Proactive Cyber Defense Systems. Available at SSRN 5774926.
19. Dwork, C. (2008). Differential privacy. ICALP Proceedings, 1–12.
20. Bandi, V. D. V. K. (2023). Production-Grade Machine Learning Pipelines For Healthcare Predictive Analytics. South Eastern European Journal of Public Health, 189–205. Retrieved from https://www.seejph.com/index.php/seejph/article/view/7057
21. Kolla, S. K. (2021). Architectural Frameworks for Large-Scale Electronic Health Record Data Platforms. Current Research in Public Health, 1(1), 1–19. Retrieved from https://www.scipublications.com/journal/index.php/crph/article/view/1372
22. Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874.
23. Friedman, C., & Elhadad, N. (2014). Natural language processing in health care. In Biomedical Informatics. Springer.
24. Garapati, R. S. (2022). AI-Augmented Virtual Health Assistant: A Web-Based Solution for Personalized Medication Management and Patient Engagement. Available at SSRN 5639650.
25. Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms. Pattern Recognition, 64, 206–223.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 5, Issue 6, November- December 2023 |

DOI:10.15662/IJEETR.2023.0506021

26. Maguluri, K. K., Pandugula, C., Kalisetty, S., & Mallesham, G. (2022). Advancing Pain Medicine with AI and Neural Networks: Predictive Analytics and Personalized Treatment Plans for Chronic and Acute Pain Managements. Journal of Artificial Intelligence and Big Data, 2(1), 112-126.
27. Segireddy, A. R. (2021). Containerization and Microservices in Payment Systems: A Study of Kubernetes and Docker in Financial Applications. Universal Journal of Business and Management, 1(1), 1–17. Retrieved from https://www.scipublications.com/journal/index.php/ujbm/article/view/1352
28. He, J., Baxter, S. L., Xu, J., et al. (2019). The practical implementation of AI in healthcare. Nature Medicine, 25(1), 30–36.
29. Inala, R. AI-Powered Investment Decision Support Systems: Building Smart Data Products with Embedded Governance Controls.
30. Assimakopoulos, G., Mana, A., & Papageorgiou, E. Predictive analytics for maximizing crop yields with minimal resource consumption. Agricultural Systems Research, 19(1), 112-125.
31. Gottimukkala, V. R. R. (2021). Digital Signal Processing Challenges in Financial Messaging Systems: Case Studies in High-Volume SWIFT Flows.
32. Iglewicz, B., & Hoaglin, D. C. (1993). How to detect and handle outliers. ASQC.
33. Johnson, A. E. W., Pollard, T. J., Shen, L., et al. (2016). MIMIC-III database. Scientific Data, 3, 160035.
34. Yandamuri, U. S. (2022). Big Data Pipelines for Cross-Domain Decision Support: A Cloud-Centric Approach. International Journal of Scientific Research and Modern Technology, 1(12), 227–237. https://doi.org/10.38124/ijsrmt.v1i12.1111
35. Alfizar, A., & Nasution, M. Climate change amplification of pests and diseases in agricultural systems. Journal of Sustainable Agriculture, 12(2), 45-58.
36. Rongali, S. K. (2022). AI-Driven Automation in Healthcare Claims and EHR Processing Using MuleSoft and Machine Learning Pipelines. Available at SSRN 5763022.
37. Kriegel, H. P., Kröger, P., Schubert, E., & Zimek, A. (2009). Outlier detection in axis-parallel subspaces. PKDD Proceedings, 831–838.
38. Kummari, D. N. (2023). AI-Powered Demand Forecasting for Automotive Components: A Multi-Supplier Data Fusion Approach. European Advanced Journal for Emerging Technologies (EAJET)-p-ISSN 3050-9734 en e-ISSN 3050-9742, 1(1).
39. Kalisetty, S. (2023). The Role of Circular Supply Chains in Achieving Sustainability Goals: A 2023 Perspective on Recycling, Reuse, and Resource Optimization. Reuse, and Resource Optimization (June 15, 2023).
40. Li, Y., Chen, C. Y., Wasserman, W. W., & Ramani, A. K. (2016). Deep feature selection. Bioinformatics, 32(5), 743–750.
41. Varri, D. B. S. (2022). A Framework for Cloud-Integrated Database Hardening in Hybrid AWS-Azure Environments: Security Posture Automation Through Wiz-Driven Insights. International Journal of Scientific Research and Modern Technology, 1(12), 216-226.
42. Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long short-term memory networks for anomaly detection. ESANN Proceedings.
43. Mandl, K. D., & Kohane, I. S. (2015). Data sharing in healthcare. BMJ, 350, h988.
44. Garapati, R. S. (2023). Optimizing Energy Consumption in Smart Build-ings Through Web-Integrated AI and Cloud-Driven Control Systems.
45. Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare. Briefings in Bioinformatics, 19(6), 1236–1246.
46. Kushvanth Chowdary Nagabhyru. (2023). Accelerating Digital Transformation with AI Driven Data Engineering: Industry Case Studies from Cloud and IoT Domains. Educational Administration: Theory and Practice, 29(4), 5898–5910. https://doi.org/10.53555/kuey.v29i4.10932
47. Meda, R. (2023). Data Engineering Architectures for Scalable AI in Paint Manufacturing Operations. European Data Science Journal (EDSJ) p-ISSN 3050-9572 en e-ISSN 3050-9580, 1(1).
48. Guntupalli, R. (2023). Optimizing Cloud Infrastructure Performance Using AI: Intelligent Resource Allocation and Predictive Maintenance. Available at SSRN 5329154.
49. Patcha, A., & Park, J. M. (2007). An overview of anomaly detection techniques. Computer Networks, 51(12), 3448–3470.
50. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn. Journal of Machine Learning Research, 12, 2825–2830.
51. Aitha, A. R. (2023). CloudBased Microservices Architecture for Seamless Insurance Policy Administration. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 607-632.
52. Rajkomar, A., Oren, E., Chen, K., et al. (2018). Scalable deep learning with EHRs. NPJ Digital Medicine, 1, 18.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 5, Issue 6, November- December 2023 |

DOI:10.15662/IJEETR.2023.0506021

53. Avinash Reddy Segireddy. (2022). Terraform and Ansible in Building Resilient Cloud-Native Payment Architectures. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 444–455. Retrieved from https://www.ijisae.org/index.php/IJISAE/article/view/7905.
54. Ringberg, H., Soule, A., Rexford, J., & Diot, C. (2007). Sensitivity of PCA for anomaly detection. SIGMETRICS Proceedings.
55. Koppolu, H. K. R., Sheelam, G. K., & Komaragiri, V. B. (2023). Autonomous Telecommunication Networks: The Convergence of Agentic AI and AI-Optimized Hardware. International Journal of Science and Research (IJSR), 12(12), 2253-2270.
56. Ruff, L., Vandermeulen, R. A., Görnitz, N., et al. (2018). Deep one-class classification. ICML Proceedings.
57. Rongali, S. K. (2023). Explainable Artificial Intelligence (XAI) Framework for Transparent Clinical Decision Support Systems. International Journal of Medical Toxicology and Legal Medicine, 26(3), 22-31.
58. Salfner, F., Lenk, M., & Malek, M. (2010). Survey of failure prediction methods. ACM Computing Surveys, 42(3), 1–42.
59. Nagubandi, A. R. (2023). Advanced Multi-Agent AI Systems for Autonomous Reconciliation Across Enterprise Multi-Counterparty Derivatives, Collateral, and Accounting Platforms. International Journal of Finance (IJFIN)-ABDC Journal Quality List, 36(6), 653-674.
60. Schölkopf, B., Platt, J. C., Shawe-Taylor, J., et al. (2001). Estimating the support of a high-dimensional distribution. Neural Computation, 13(7), 1443–1471.
61. Uday Surendra Yandamuri. (2023). An Intelligent Analytics Framework Combining Big Data and Machine Learning for Business Forecasting. International Journal Of Finance, 36(6), 682-706. https://doi.org/10.5281/zenodo.18095256
62. Sipos, R., Fradkin, D., Moerchen, F., & Wang, Z. (2014). Log-based predictive maintenance. KDD Proceedings.
63. Meda, R. (2023). Intelligent Infrastructure for Real-Time Inventory and Logistics in Retail Supply Chains. Educational Administration: Theory and Practice.
64. Kolla, S. K. (2021). Designing Scalable Healthcare Data Pipelines for Multi-Hospital Networks. World Journal of Clinical Medicine Research, 1(1), 1–14. Retrieved from https://www.scipublications.com/journal/index.php/wjcmr/article/view/1376
65. Bandi, V. D. V. K. (2023). Cloud-Native Model Lifecycle Management for Enterprise AI Systems. International Journal of Scientific Research and Modern Technology, 2(12), 78–90. https://doi.org/10.38124/ijsrmt.v2i12.1236
66. Inala, R. Revolutionizing Customer Master Data in Insurance Technology Platforms: An AI and MDM Architecture Perspective.
67. Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society B, 58(1), 267–288.
68. Garapati, R. S. (2022). Web-Centric Cloud Framework for Real-Time Monitoring and Risk Prediction in Clinical Trials Using Machine Learning. Current Research in Public Health, 2, 1346.
69. Keerthi Amistapuram. (2023). Privacy-Preserving Machine Learning Models for Sensitive Customer Data in Insurance Systems. Educational Administration: Theory and Practice, 29(4), 5950–5958. https://doi.org/10.53555/kuey.v29i4.10965
70. AI Powered Fraud Detection Systems: Enhancing Risk Assessment in the Insurance Sector. (2023). American Journal of Analytics and Artificial Intelligence (ajaai) With ISSN 3067-283X, 1(1). https://ajaai.com/index.php/ajaai/article/view/14
71. Weber, G. M., Mandl, K. D., & Kohane, I. S. (2014). Finding the missing link for big biomedical data. JAMIA, 21(1), 1–3.
72. Kolla, S. H. (2021). Rule-Based Automation for IT Service Management Workflows. Online Journal of Engineering Sciences, 1(1), 1–14. Retrieved from https://www.scipublications.com/journal/index.php/ojes/article/view/1360
73. Kalisetty, S., Vankayalapati, R. K., Reddy, L., Sondinti, K., & Valiki, S. (2022). AI-Native Cloud Platforms: Redefining Scalability and Flexibility in Artificial Intelligence Workflows. Linguistic and Philosophical Investigations, 21(1), 1-15.
74. Zhang, Y., & Yang, Q. (2021). A survey on multi-task learning. IEEE Transactions on Knowledge and Data Engineering, 34(12), 5586–5609.
75. Gottimukkala, V. R. R. (2022). Licensing Innovation in the Financial Messaging Ecosystem: Business Models and Global Compliance Impact. International Journal of Scientific Research and Modern Technology, 1(12), 177-186.
76. Jabed, M. A., & Azmi Murad, M. A. Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. Heliyon, 10(e40836), 1-18.
77. Vijayakumar, H. (2023). Business value impact of AI-powered service operations (AIServiceOps). SSRN Electronic Journal.
78. Little, R. J. A., & Rubin, D. B. (2002). Statistical analysis with missing data. Wiley.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 5, Issue 6, November- December 2023 |

DOI:10.15662/IJEETR.2023.0506021

79. Siva Hemanth Kolla. (2022). Knowledge Retrieval Systems for Enterprise Service Environments. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 495–506. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/8037
80. Bishop, C. M. (1994). Novelty detection and neural network validation. IEE Proceedings, 141(4), 217–222.
81. Goutham Kumar Sheelam, Hara Krishna Reddy Koppolu. (2022). Data Engineering And Analytics For 5G-Driven Customer Experience In Telecom, Media, And Healthcare. Migration Letters, 19(S2), 1920–1944. Retrieved from https://migrationletters.com/index.php/ml/article/view/11938
82. Becker, M., Khan, A., & Schmidt, J. (2023). Heat stress impacts on crop yields in Punjab: A data-driven analysis. Climate Risk Management, 41, 100522.
83. Cai, L., Zhang, Y., & Wang, H. (2023). Digitalization and technology adoption in smallholder rice farming. Journal of Development Economics, 162, 103055.
84. Amistapuram, K. (2022). Fraud Detection and Risk Modeling in Insurance: Early Adoption of Machine Learning in Claims Processing. Available at SSRN 5741982.
85. Kalisetty, S., & Singireddy, J. (2023). Optimizing Tax Preparation and Filing Services: A Comparative Study of Traditional Methods and AI Augmented Tax Compliance Frameworks. Available at SSRN 5206185.
86. Ramesh Inala. (2023). Big Data Architectures for Modernizing Customer Master Systems in Group Insurance and Retirement Planning. Educational Administration: Theory and Practice, 29(4), 5493–5505. https://doi.org/10.53555/kuey.v29i4.10424
87. Aggarwal, C. C. (2017). Outlier analysis (2nd ed.). Springer.
88. Davuluri, P. N. AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems.
89. Bifet, A., & Gavalda, R. (2007). Learning from time-changing data with adaptive windowing. SDM Proceedings.
90. Nagabhyru, K. C. (2023). From Data Silos to Knowledge Graphs: Architecting CrossEnterprise AI Solutions for Scalability and Trust. Available at SSRN 5697663.
91. Zaharia, M., Chowdhury, M., Franklin, M. J., et al. (2010). Spark: Cluster computing. HotCloud Proceedings.
92. Avinash Reddy Aitha. (2022). Deep Neural Networks for Property Risk Prediction Leveraging Aerial and Satellite Imaging. International Journal of Communication Networks and Information Security (IJCNIS), 14(3), 1308–1318. Retrieved from https://www.ijcnis.org/index.php/ijcnis/article/view/8609
93. Kalisetty, S., & Ganti, V. K. A. T. (2019). Transforming the Retail Landscape: Srinivas's Vision for Integrating Advanced Technologies in Supply Chain Efficiency and Customer Experience. Online Journal of Materials Science, 1, 1254.
94. Alenezi, M., & Akour, M. AI-driven innovations in software engineering: A review of current practices and future directions. Applied Sciences, 15(3), 1344. https://doi.org/10.3390/app15031344 Cited by: 149
95. Davuluri, P. N. Integrating Artificial Intelligence into Event-Driven Financial Crime Compliance Platforms.