AI IN MODERN FARMING · LESSON 4 OF 6

AI for Crop Prediction & Decision Making

How machine learning turns data into farming decisions.

TODAY

Predict · Detect · Decide

The 6-Lesson Journey

1. **Smart Sensors & IoT** — FOUNDATION
2. **Computer Vision for Crops** — SEEING
3. **Robotics & Automation** — DOING
4. ◀ YOU ARE HERE **Crop Prediction & Decision Making**
5. **Supply Chain AI** — MOVING
6. **Policy & Ethics** — GOVERNING

TODAY'S FOCUS

Lessons 1–3 give us the raw data. Lesson 4 turns that data into decisions — the beating heart of every smart farm.

03 · WHAT YOU'LL LEARN

Learning Objectives

By the end of this lesson, you will be able to move confidently between **raw data**, **trained models**, and **real farm decisions**.

a

EXPLAIN

Predictive analytics in farming

Define the 5-step pipeline: collect → clean → train → predict → act.

b

READ

Real agricultural datasets

Inspect columns, describe distributions, spot missing values.

c

BUILD

A simple crop-prediction model

Train a Decision Tree / Random Forest on soil & weather features.

d

DEBATE

Sustainability trade-offs

Weigh water & input savings against data-privacy & farmer-trust costs.

What is Predictive Analytics?

DEFINITION

Using **historical data** and **machine-learning models** to make quantitative forecasts about future events on the farm, from tomorrow's rainfall to next season's yield.

THE 5-STEP PIPELINE

01 Collect

Soil sensors, weather stations, satellites, farmer logs.

`→ raw CSVs, images, rows`

→

02 Clean

Fix missing values, remove outliers, align units.

`→ tidy DataFrame`

→

03 Train

Fit a model (tree, forest, CNN) on labelled examples.

`→ trained model.pkl`

→

04 Predict

Feed new field data and get a numeric or class output.

`→ yield: 4.7 t/ha`

→

05 Act

Farmer irrigates, sprays, plants — closing the loop.

`→ decision on field`

CITE

Mehta et al., *Frontiers in Plant Science* (2025) — AI & XAI methods for interpretable crop-yield prediction.

From Field Sensor to Farmer's Action

INPUT
Sensors
Soil moisture · pH · humidity · temperature · leaf images

→

STORE
Data Lake
Cloud repository — millions of rows & images

→

LEARN
ML Model
Regression · CNN · Random Forest · gradient boosting

→

ADVISE
Recommendation
Irrigate 12mm · spray only zone-3 · plant rice

→

ACT
Farmer Action
Loop closes on the field

FEEDBACK — new data retrains the model

DATASETS WE'LL TOUCH TODAY

- Kaggle Crop Recommendation (2,200 rows)
- PlantVillage (54,305 leaf images)
- India district crop production (575,210 rows)

Crop Yield Prediction

THE TASK

Given **weather + soil + historical yield**, predict a **continuous** yield value (tonnes / hectare) for the coming season. This is a classic *supervised regression* problem.

TYPICAL INPUT FEATURES

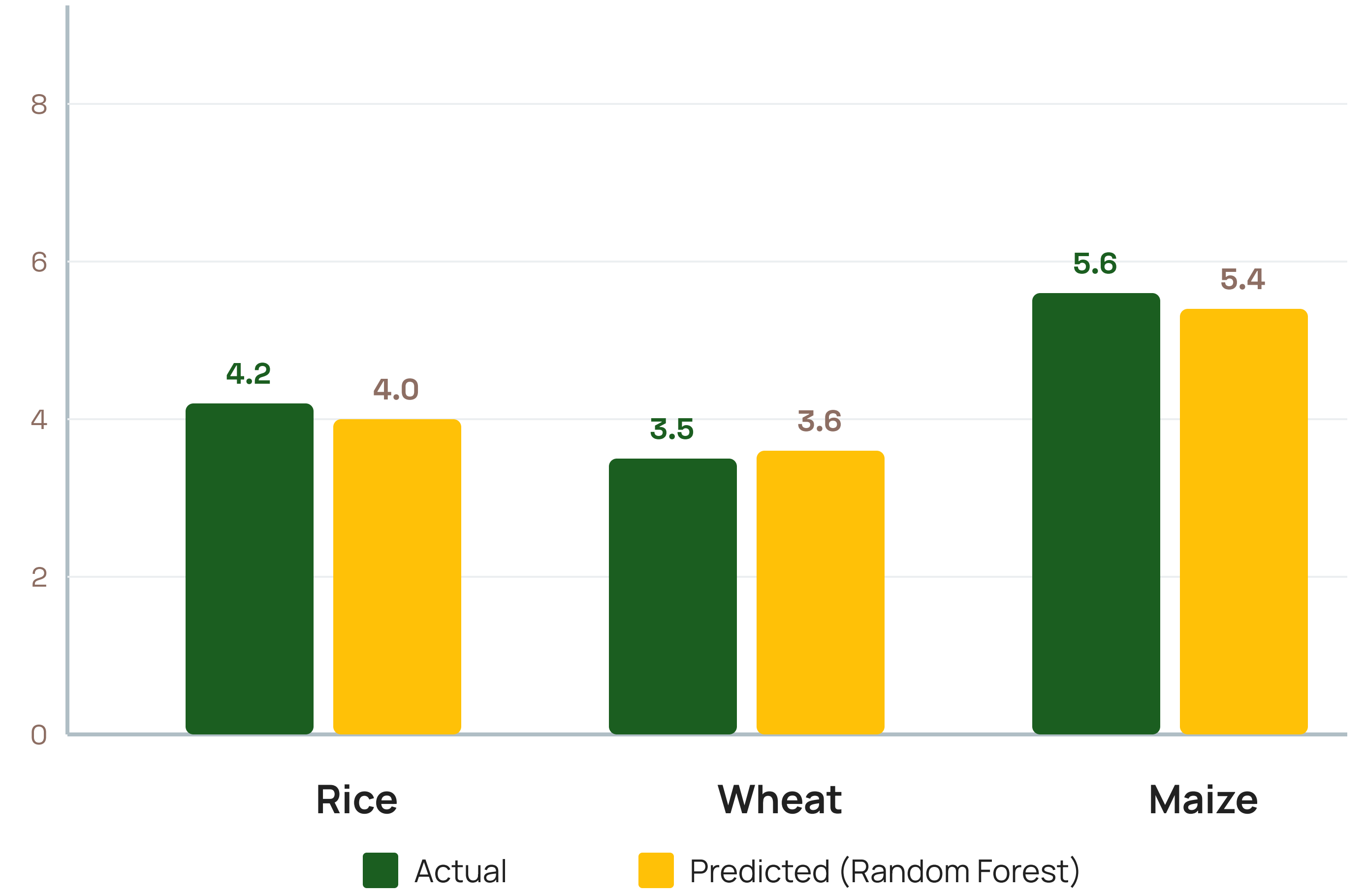
- Rainfall (mm)
- Avg temperature (°C)
- Soil N-P-K
- pH
- Prior-season yield
- NDVI
- Fertilizer applied
- Irrigation type

95%
yield-prediction accuracy achievable with modern ML

-25%
input costs saved (fertilizer, water, seed)

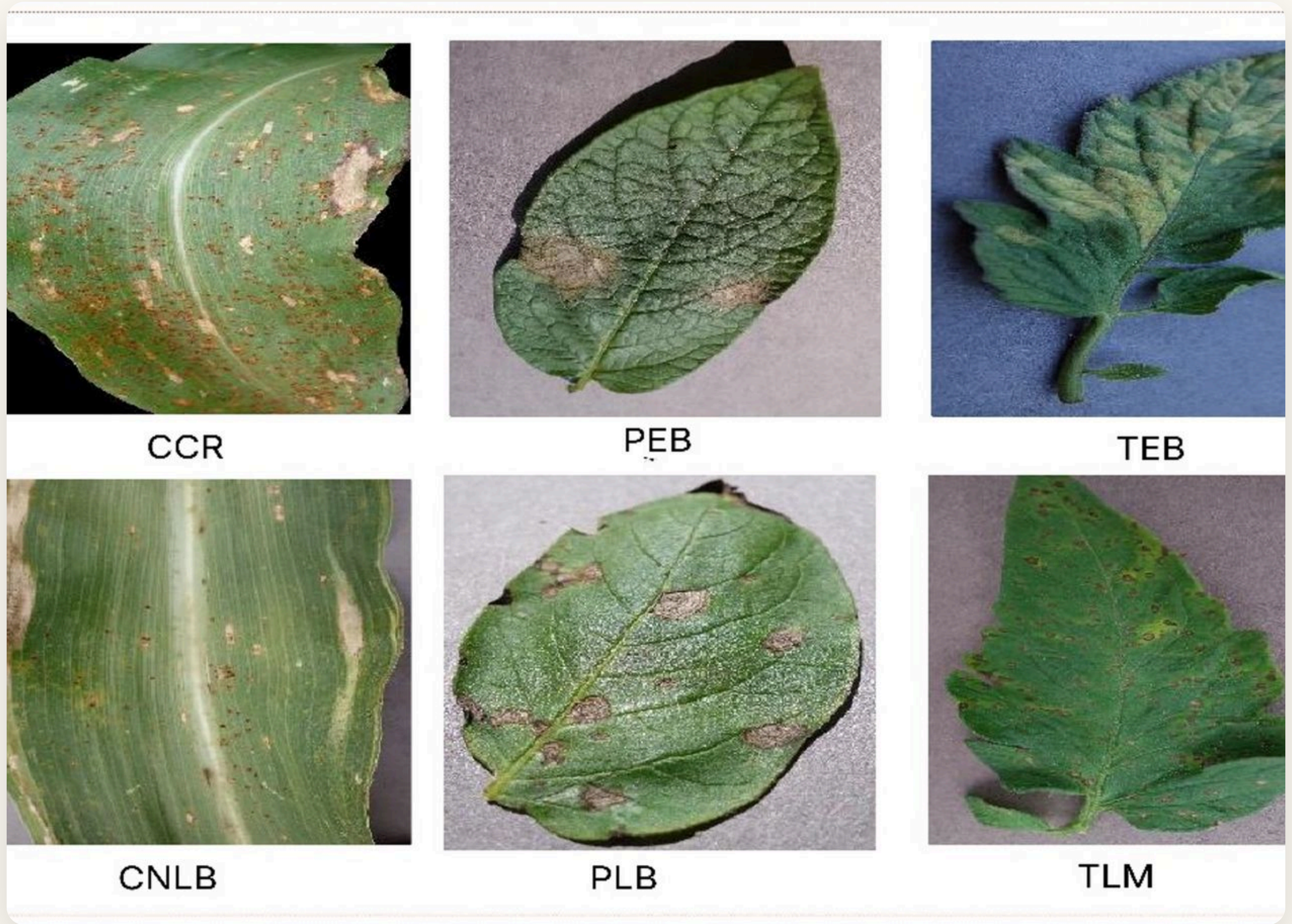
CHART · ILLUSTRATIVE

Actual vs. Predicted yield (t/ha)



CITE · Shawon et al. 2025 (*Smart Agric. Tech.*) — ML crop-yield prediction review. Stat: Omdena (2025), AI for Crop Yield.

Disease Detection Systems



PlantVillage sample — 38 disease classes across 14 crops

HOW IT WORKS

A **Convolutional Neural Network** (CNN) sees a leaf photo pixel-by-pixel, learns which textures signal each disease, and outputs a class label + confidence score.

THE DATASET

54,305

labelled leaf images

38 disease classes · **14** crops

PlantVillage · open dataset

STATE-OF-THE-ART ARCHITECTURES

- **VGG16** — classic depth baseline
- **ResNet-50** — skip connections
- **EfficientNet** — best acc/params
- **YOLOv8** — bbox + class in one pass

?

KNOWLEDGE CHECK

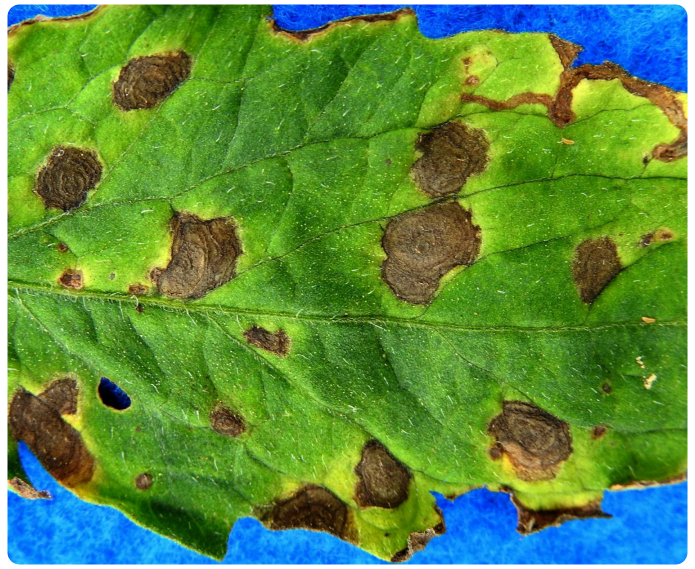
A CNN classifies leaf images as diseased. This is a task of:

A) supervised classification · B) unsupervised clustering · C) regression · D) reinforcement learning

CITE · Upadhyay & Singh 2025 (*Artif. Intell. Review*) — deep-learning approaches for plant-disease detection.

How the Model "Sees" Disease

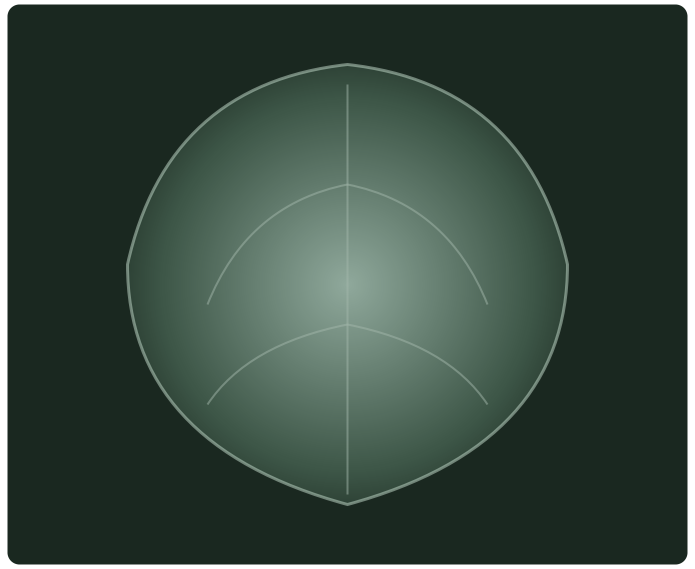
FEATURE-MAP VISUALIZATION · Tomato · Early Blight



STAGE 1 · INPUT

Raw RGB image

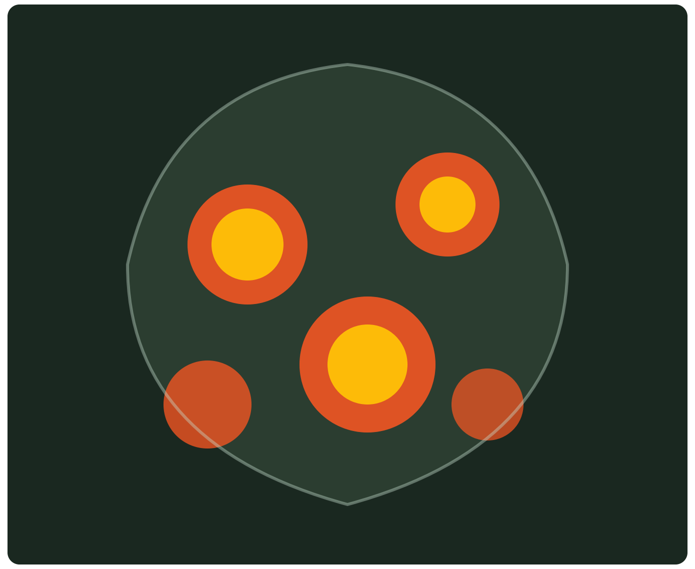
→



STAGE 2 · EARLY LAYERS

Edges & veins

→



STAGE 3 · GRAD-CAM

Lesion heatmap

DECISION

Early Blight detected

Confidence: **96.4%**

Recommended: **copper-based fungicide**, not fertilizer.

FERTILIZER vs. PESTICIDE — the model chooses

Yellowing + rings → **fungal disease** → fungicide.
Uniform pale leaves → **nutrient deficiency** → nitrogen top-up.

TIE FACT

‹ 2 seconds

to detect disease from a leaf image on a modern smartphone.

CITE · Mohan et al. 2025 (*Frontiers in Plant Science*) — feature-map interpretability for crop disease models. Heatmap stages illustrative.

Pest Monitoring from the Sky

STEP 1
UAV / drone flies grid

STEP 2
RGB + multispectral capture

STEP 3
ML flags pest hotspots

INDIA

AI drones in Maharashtra cotton

Pink bollworm devastates Indian cotton every season. Drones equipped with computer-vision models fly the crop and flag infested rows in near-real time, letting farmers spot-spray instead of blanket-spray.

→ Targeted spraying reduces pesticide load and protects pollinators.

UNITED STATES

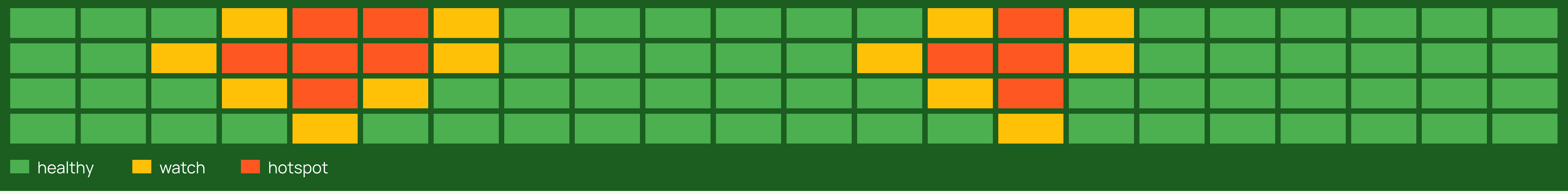
John Deere See & Spray

Cameras mounted on the sprayer boom identify weeds in real time and open only the nozzle over that plant. The tractor drives through row-crops (corn, soy, cotton) and spot-treats.

→ Reported ~60% reduction in herbicide use vs. broadcast spraying.

PEST HOTSPOT MAP

Model output: red squares = infestation risk › 70%.



CITE · Anam et al. 2024 (*Smart Agric. Tech.*) — UAV + AI systems for pest and crop monitoring.

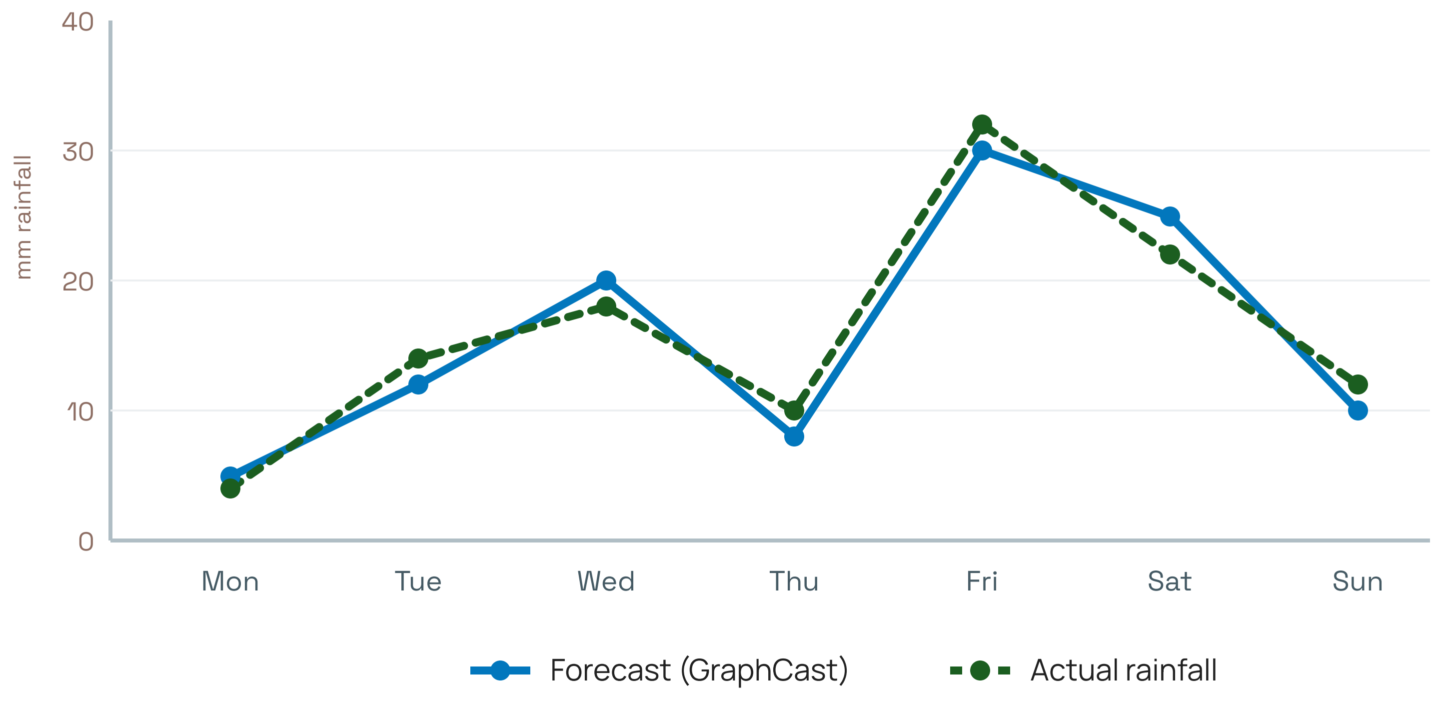
Weather Forecasting Models

HYBRID APPROACH

Classical **Numerical Weather Prediction (NWP)** solves atmospheric physics. Modern **ML models** learn corrections from historical data — 10 000× faster inference and often more accurate at the farm scale.

CHART · ILLUSTRATIVE

7-day rainfall: forecast vs. actual



FRONTIER WEATHER-AI MODELS

DEEPMIND

GraphCast

Graph neural net, 10-day global forecast in ~60 s.

HUAWEI

Pangu-Weather

3-D transformer, medium-range forecasts.

NVIDIA

FourCastNet

Fourier neural operator for extreme events.

?

KNOWLEDGE CHECK

Which forecast horizon is most useful for deciding today's irrigation?

A) 5-minute nowcast · **B) 24-72 h farm-scale forecast** · C) 30-day monthly average · D) 12-month climate outlook

CITE · Interreg Central Europe (2024) — AI-integrated weather forecasting advances field-scale decision support.

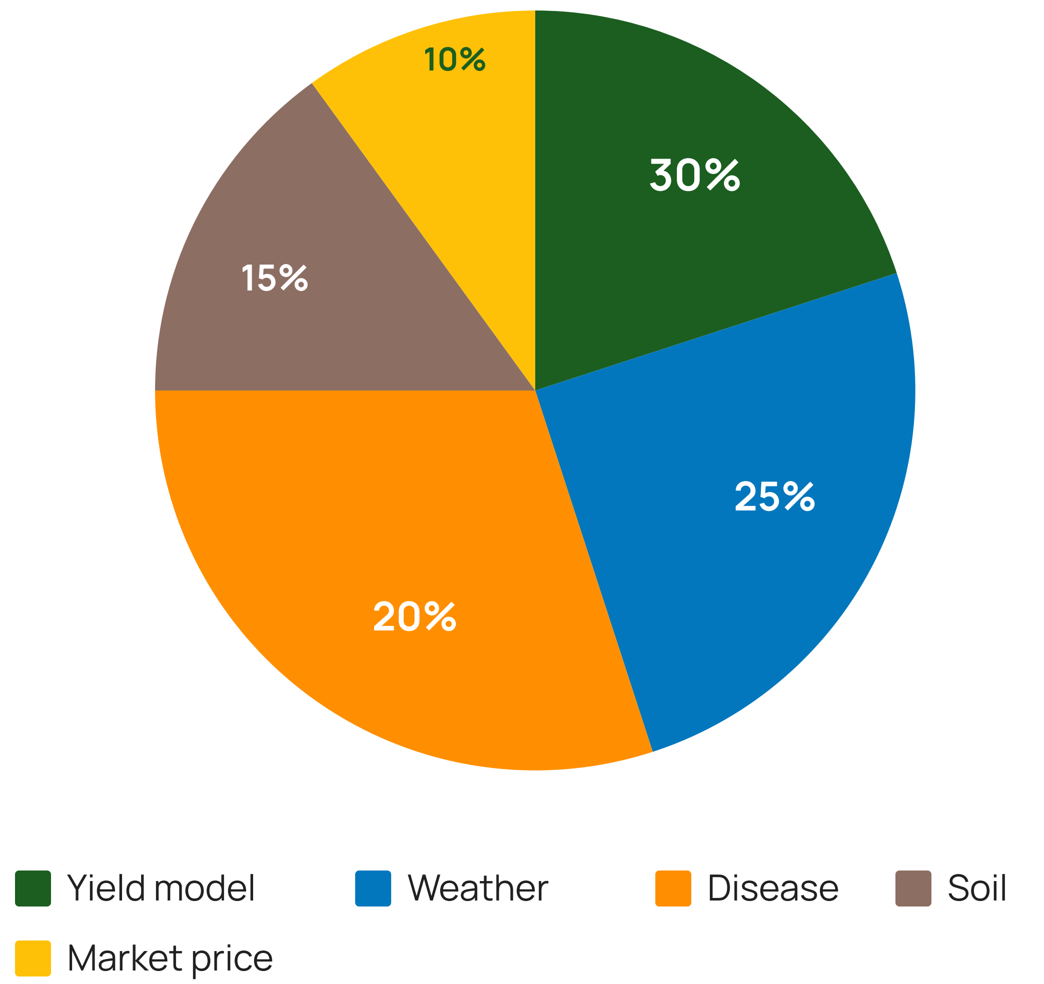
AI-based Decision Support Systems

DEFINITION

A **Decision Support System (DSS)** pulls signals from every layer we've seen — yield model, disease detector, weather forecast, market price — and returns **one clear recommendation** the farmer can act on.

CHART · ILLUSTRATIVE

Weight of decision factors



REAL-WORLD PLATFORMS

AgriTech Platform

Government-of-India digital platform connecting farmers to advisory services, mandi prices, and weather.

ICAR VCoE on AI in Agriculture

Virtual Centre of Excellence coordinating national AI-in-agriculture research and pilots across ICAR institutes.

FIVE SIGNALS FEED THE DSS

- 🌾 **Yield model** — how much to expect
- 🍃 **Disease detector** — do we need to spray
- 🌦️ **Weather forecast** — irrigate today or wait
- 💰 **Market price** — best time to sell
- 🌱 **Soil health** — fertilize which zone

AI Cuts Waste, Not Yield

BEFORE — CONVENTIONAL

Broadcast approach

- 💧 Water — uniform irrigation across entire field
- 🧪 Fertilizer — same rate everywhere regardless of soil
- 🐛 Pesticide — blanket spraying by calendar schedule
- ⏱️ Timing — based on tradition or generic almanac
- 📊 Data — pen-and-paper crop diary

→

AFTER — AI-DRIVEN

Precision approach

- 💧 Water — variable-rate by moisture sensor zone
- 🧪 Fertilizer — rate per NDVI & soil test grid
- 🐛 Pesticide — spot-spray only detected hotspots
- ⏱️ Timing — model-predicted optimum date
- 📊 Data — continuous sensor + drone stream

20–30%

less **water** used under precision-irrigation regimes

15–25%

less **fertilizer & pesticide** applied at equal yield

CITE · Mana et al. 2024 — Sustainable production agriculture through precision farming (literature averages).

🇮🇳 INDIA SPOTLIGHT

AI Meets Bharat's Fields

A public-sector-led model — ICAR, IFPRI, WEF partnerships, and open dataset infrastructure through AgriStack.

FEATURED DATASET

575,210

rows of district-level crop production & yield

`Kaggle · Indian Agriculture Crop Production & Yield · 1998–2021`

The India Stack for Ag

ICAR AgriStack

Digital public infrastructure — farmer IDs, geo-tagged land parcels, unified crop registry.

IFPRI × ICAR partnership

Policy-research collaboration on data-driven productivity for smallholders.

Major crops

Rice, wheat, sugarcane dominate acreage — priority targets for prediction pilots.

?

KNOWLEDGE CHECK

What is the primary role of AgriStack for an AI system?

A) Sell fertilizer · **B) Provide the unified data layer** · C) Train drones · D) Insure crops

CITE · WEF (2025), Future Farming in India Report.

🇺🇸 UNITED STATES SPOTLIGHT

The Corn Belt Goes AI-Native

USDA AI Strategy · FY 2025-2026

Federal roadmap for AI adoption across USDA agencies, from research to operations.

NASS Quick Stats

County-level open data — planted acres, yield, price. Public API for anyone.

FARM AI Act (2025) · NIFA Data Science

Legislative and grant funding for AI research at land-grant universities.

FLAGSHIP MARKET

Corn/soy belt runs on variable-rate equipment

Every planter, sprayer & combine now streams telemetry to a farm-scale ML platform.

ON THE GROUND

Center-pivot irrigation + drone monitoring + variable-rate planters

Compared to India's public-infrastructure model, the US pipeline is private-sector-led: John Deere, Climate FieldView, Granular. USDA funds the research and open data; farmers buy the tools.

CITE · USDA (FY 2025-2026), Artificial Intelligence Strategy.

15 · CROP PREDICTION CHALLENGE

Predict the Best Crop for a Field

🎯 GOAL

Given a field's soil chemistry & climate, predict which of **22 crops** will thrive there.

DATASET

Kaggle Crop Recommendation

by Atharva Ingle · ~2,200 rows · CC0 license

kaggle.com/datasets/atharvaingle/crop-recommendation-dataset

22 POSSIBLE CROPS (LABELS)

rice · maize · jute · cotton · coconut · papaya · orange · apple · muskmelon · watermelon · grapes · mango · banana · pomegranate · lentil · blackgram · mungbean · mothbeans · pigeonpeas · kidneybeans · chickpea · coffee

FIRST 5 ROWS

df.head()

N	P	K	temp °C	hum %	pH	rain mm	label
90	42	43	20.87	82.00	6.50	202.9	rice
85	58	41	21.77	80.31	7.03	226.6	rice
60	55	44	23.00	82.32	7.84	263.9	rice
74	35	40	26.49	80.16	6.98	242.9	rice
78	42	42	20.13	81.60	7.63	262.7	rice

Columns: N, P, K (kg/ha) · temperature (°C) · humidity (%) · pH · rainfall (mm) · **label** = crop

16 · STEP-BY-STEP · 6 STEPS

Build the Crop Recommender

1 Load CSV

```
import pandas as pd
df = pd.read_csv('Crop_recommendation.csv')
```

2 Explore

```
df.describe()
df['label'].value_counts()
```

3 Visualize

```
df[['N','P','K','rainfall']]
  .hist(bins=30)
```

4 Split 80 / 20

```
from sklearn.model_selection import
train_test_split
X_train, X_test, y_train, y_test =
train_test_split(X, y, test_size=0.2)
```

5 Train the model

```
from sklearn.ensemble import
RandomForestClassifier
clf = RandomForestClassifier().fit(X_train,
y_train)
```

6 Predict new farm

```
new = [[90, 42, 43,
  21, 82, 6.5, 203]]
clf.predict(new)
```

🎯 EXPECTED OUTPUT

Input → N=90, P=42, K=43,
temp=21°C, humidity=82%, pH=6.5, rainfall=203mm

MODEL ANSWER

→ rice 🌾

INSTRUCTOR NOTE · accept any reasonable tree-based model (Decision Tree, Random Forest, XGBoost) — expected top-1 accuracy ≥ 95 %.

17 · DISEASE DETECTIVE

Classify a Leaf Photo in 60 Seconds

🎯 GOAL

Upload a leaf photo. A pre-trained **MobileNetV2** returns the disease class and confidence.

DATASET

PlantVillage (subset of 38 classes)

kaggle.com/datasets/abdallahalidev/plantvillage-dataset

54,305 leaf images · 14 crops · standardized 256×256 px

CLASSES TO USE TODAY

- 🍅 Tomato_Early_blight
- 🍅 Tomato_Late_blight
- 🍅 Tomato_Healthy
- 🥔 Potato_Early_blight
- 🥔 Potato_Late_blight
- 🥔 Potato_Healthy

4-STEP PIPELINE

1. **Load leaf image**
```
img = tf.io.read_file('leaf.jpg')
```
2. **Resize to 224×224**
```
img = tf.image.resize(img, [224, 224])
```
3. **Predict with MobileNetV2**
```
probs = model.predict(img[None,...])
```
4. **Show confidence bars**
```
plt.barh(class_names, probs[0])
```

💡 NO-CODE ALTERNATIVE

Teachable Machine by Google

Drag & drop 20–30 leaves per class in the browser; get a working classifier with no Python at all.

18 · SMALL-DATASET RESULTS

What the Model Got Right — and Wrong

#	LEAF SAMPLE	GROUND TRUTH	MODEL PREDICTION	CONFIDENCE	RESULT
1	 leaf_001.jpg	Tomato · Early_blight	Tomato · Early_blight	96.4 %	✓ CORRECT
2	 leaf_002.jpg	Tomato · Late_blight	Tomato · Late_blight	92.1 %	✓ CORRECT
3	 leaf_003.jpg	Tomato · Healthy	Tomato · Healthy	99.2 %	✓ CORRECT
4	 leaf_004.jpg	Potato · Late_blight	Tomato · Late_blight	58.3 %	✗ WRONG
5	 leaf_005.jpg	Potato · Healthy	Potato · Healthy	94.7 %	✓ CORRECT

💬 DISCUSSION

Row 4 confuses Potato Late_blight with Tomato Late_blight — **same pathogen (*Phytophthora infestans*), different host**. Which crop-disease pairs are most likely to be confused, and why?

Five things to walk away with

01 **Predictive analytics = data + ML + decisions**
If any of the three is missing, you're not doing AI-driven farming — you're just collecting numbers.

02 **Datasets are the real fuel — never skip EDA**
A great model on a bad dataset is still a bad model. df.describe() and simple histograms save weeks of debugging.

03 **AI predicts yield, detects disease, monitors pests, forecasts weather & advises farmers**
Five capabilities, one integrated Decision Support System.

04 **India & the US are scaling AI in very different ways**
Public digital infrastructure (India) vs. private-sector platforms with federal research funding (US).

05 **Sustainability gains are real — but require farmer trust**
A model that saves 25% of water is only useful if the farmer chooses to follow it.

Where to Read More

📚 ACADEMIC LITERATURE

1. **Shawon et al.** (2025). ML crop-yield prediction review. *Smart Agric. Tech.*
   sciencedirect.com/science/article/pii/S2772375524003228
2. **Upadhyay & Singh** (2025). Deep learning for plant disease. *Artif. Intell. Review.*
   link.springer.com/article/10.1007/s10462-024-11100-x
3. **Mohan et al.** (2025). AI + XAI for crop yield. *Frontiers in Plant Science.*
   frontiersin.org/journals/plant-science/articles/10.3389/fpls.2024.1451607/full
4. **Anam et al.** (2024). UAV + AI for pest monitoring. *Smart Agric. Tech.*
   sciencedirect.com/science/article/pii/S2772375524002521
5. **Mana et al.** (2024). Sustainable precision agriculture.
   sciencedirect.com/science/article/pii/S2772375524000212
6. **Interreg Central EU** (2024). AI-integrated weather forecasting.
   interreg-central.eu/news/ai-integrated-weather-forecasting-advances-field-scale-decision-support-in-european-agriculture

🏛 POLICY REPORTS

7. **WEF (2025).** Future Farming in India Report.
   reports.weforum.org/docs/WEF_Future_Farming_in_India_2025.pdf
8. **USDA (FY 2025-2026).** USDA AI Strategy.
   usda.gov/sites/default/files/documents/fy-2025-2026-usda-ai-strategy.pdf
9. **Omdena (2025).** AI for Crop Yield Prediction.
   omdena.com/blog/ai-for-crop-yield-prediction-future-agriculture-2025

💾 DATASETS TO DOWNLOAD

10. **Kaggle Crop Recommendation** (Activity 1)
    kaggle.com/datasets/atharvaingle/crop-recommendation-dataset
11. **Kaggle India Crop Production & Yield**
    kaggle.com/datasets/arjunyadav99/indian-agriculture-crop-production-and-yield
12. **PlantVillage** (Activity 2)
    kaggle.com/datasets/abdallahalidev/plantvillage-dataset