Data Driven Innovation

Data Driven Innovation

Agenda – Session 1

- Historical Perspective and Importance of Data
- Importance and Sources of Data
- Nature Inspired Computing
- Data and AI-ML
- Design Thinking Aspects in Data Driven Innovation
- Data Driven Smart Applications

Historical Perspective and Importance of Data

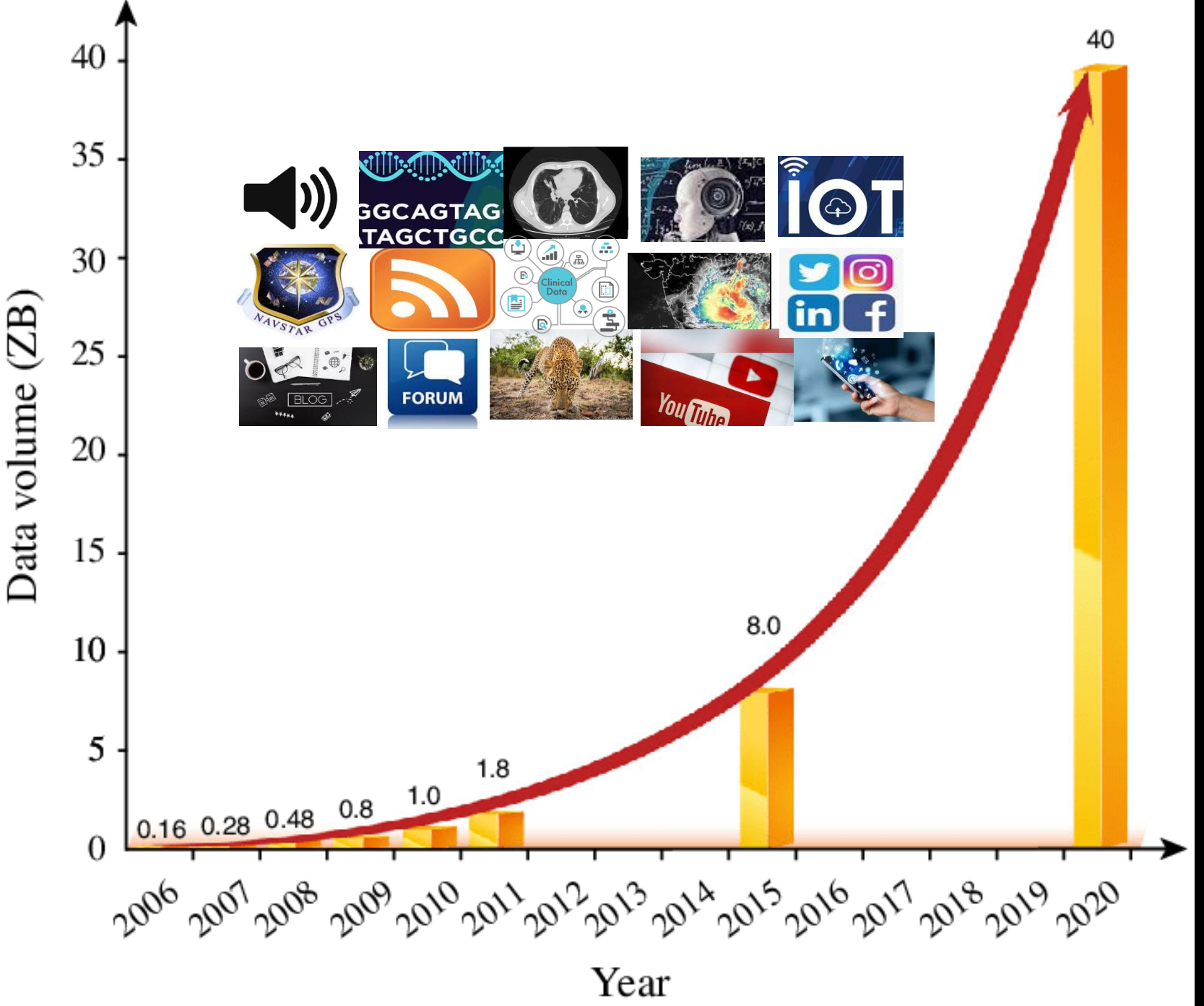
There were days when Data was viewed just as an output written by Database applications. Stored in disks and archived onto dragon sized Robotic Tape libraries

Data is still stored in Disk and archived onto Tapes

But …

Data is no longer just a secondary product

Growth of Unstructured Data


Data volume (ZB)
40
35
30
25
20
15
10
5
0
0.16
0.28
0.48
0.8
1.0
1.8
8.0
40
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Year
GGCAGTAG
TAGCTGCC
IOT
NAVSTAR GPS
Clinical Data
BLOG
FORUM
YouTube

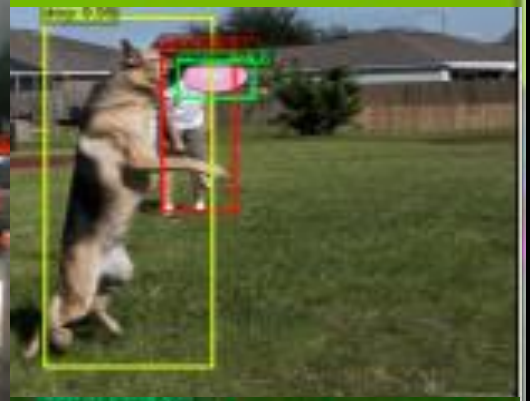
Killer Applications are now designed around data

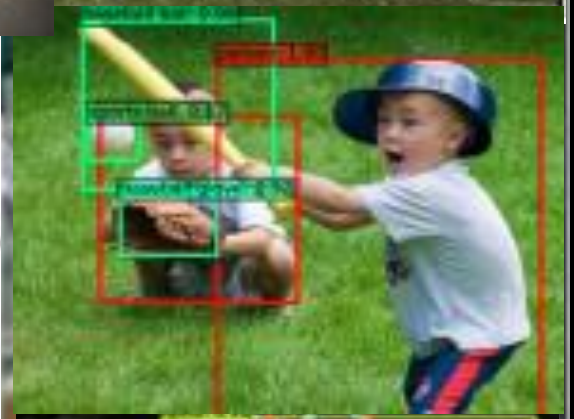
to

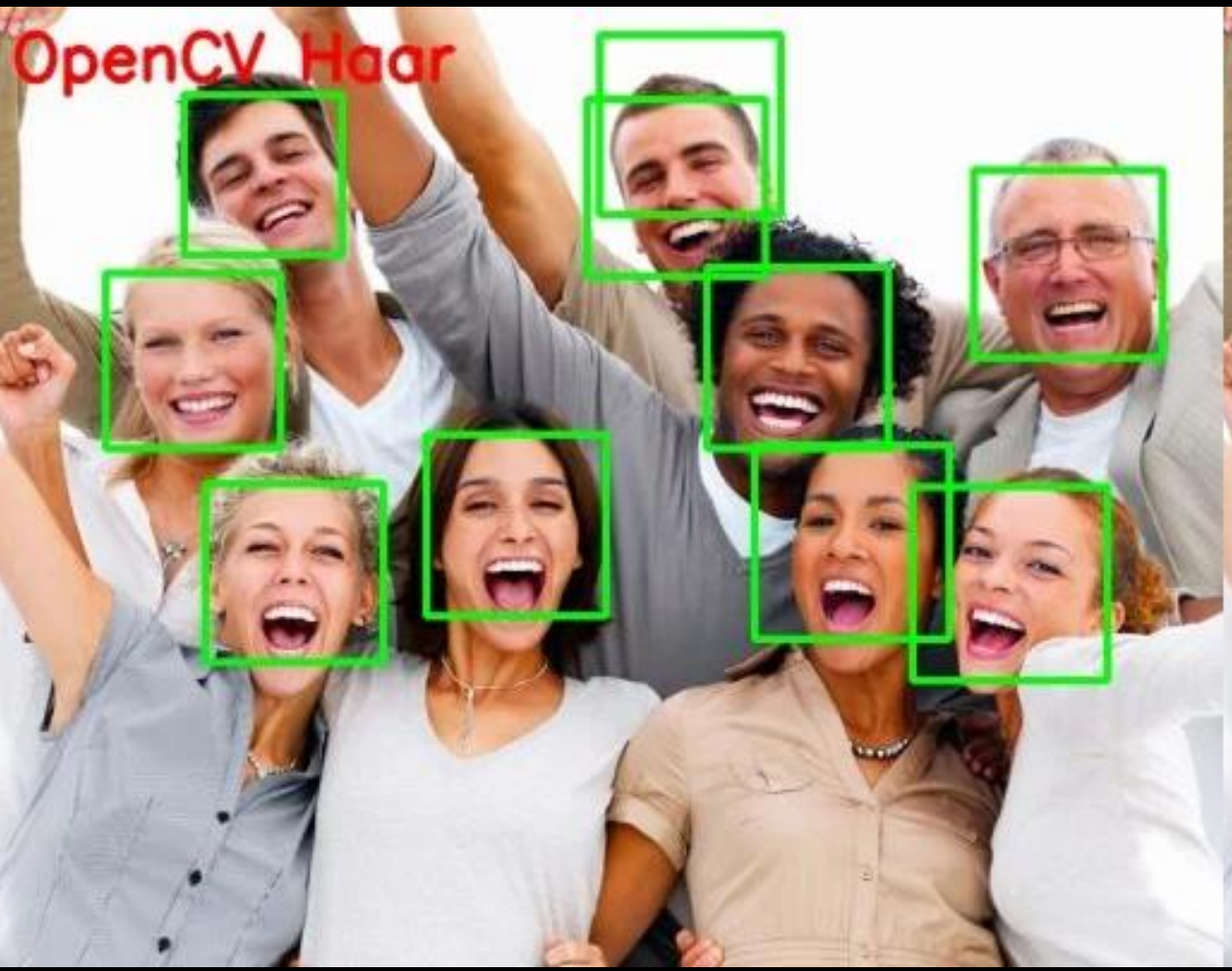
Make Decisions

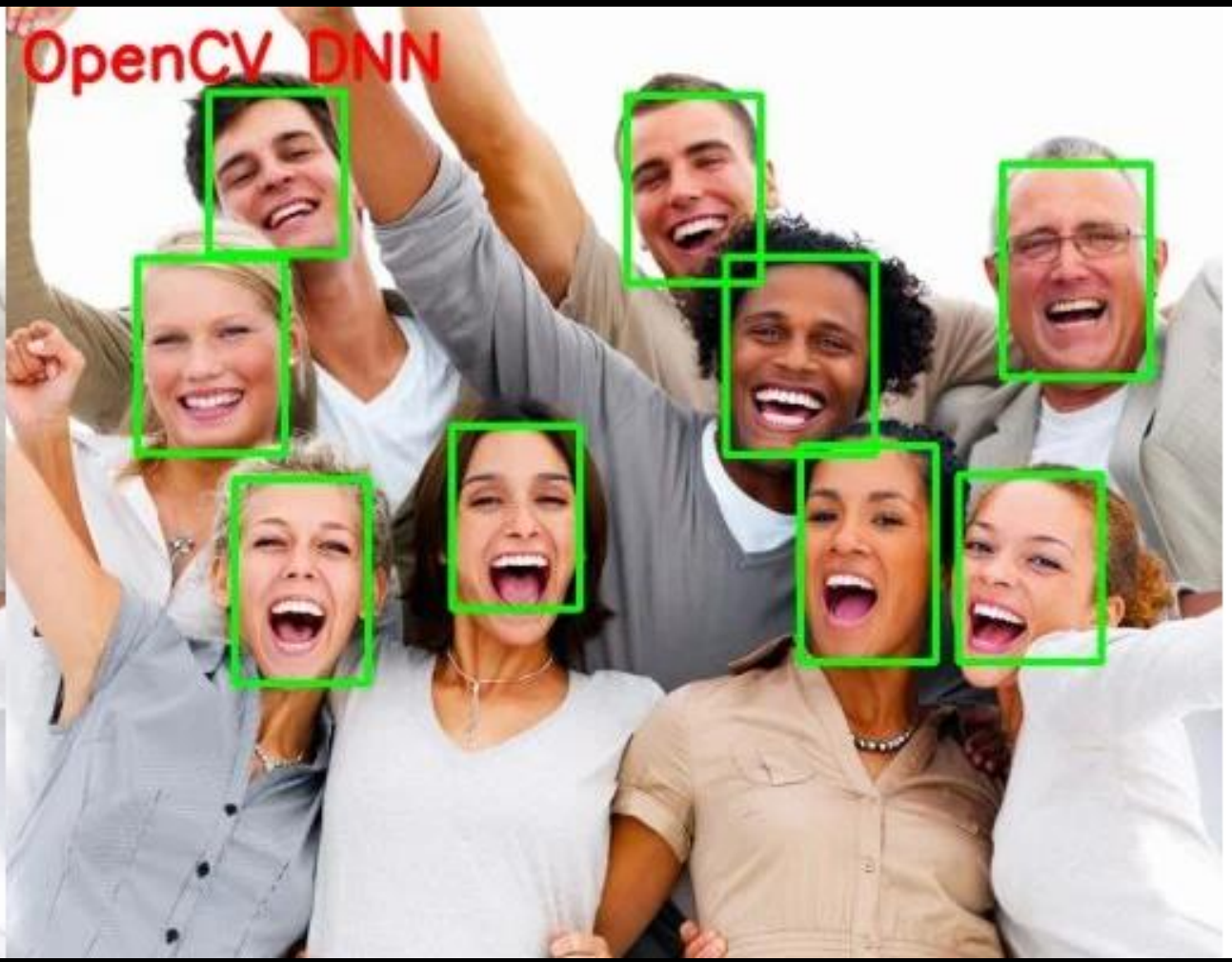
Image Classification Object Detection

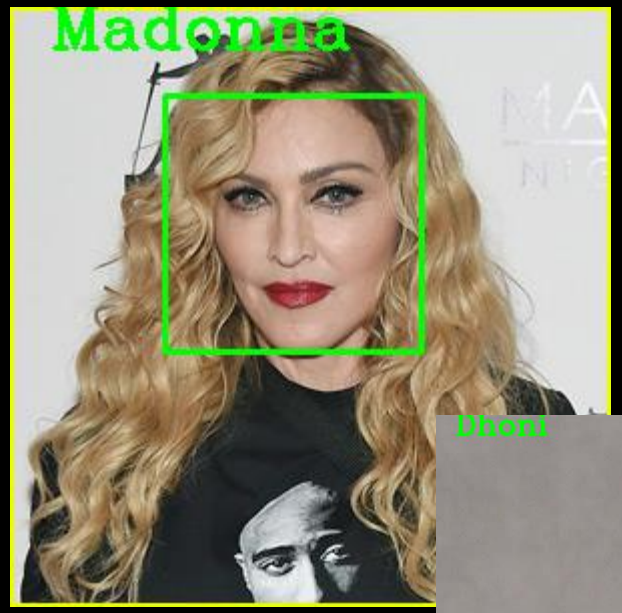
COMPUTER VISION

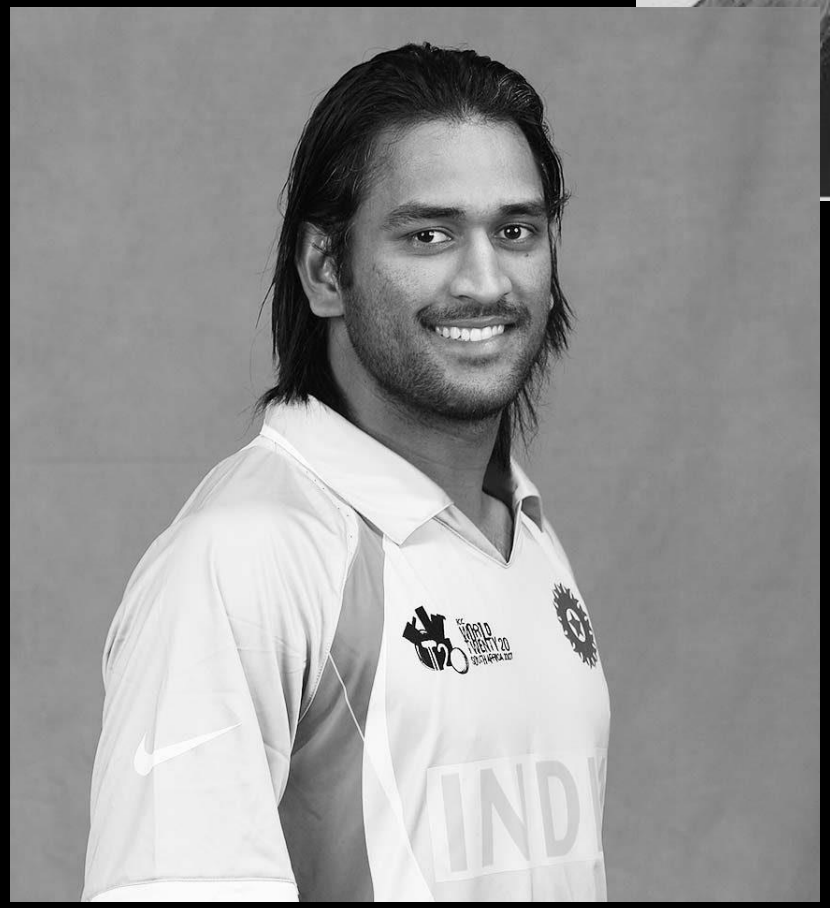


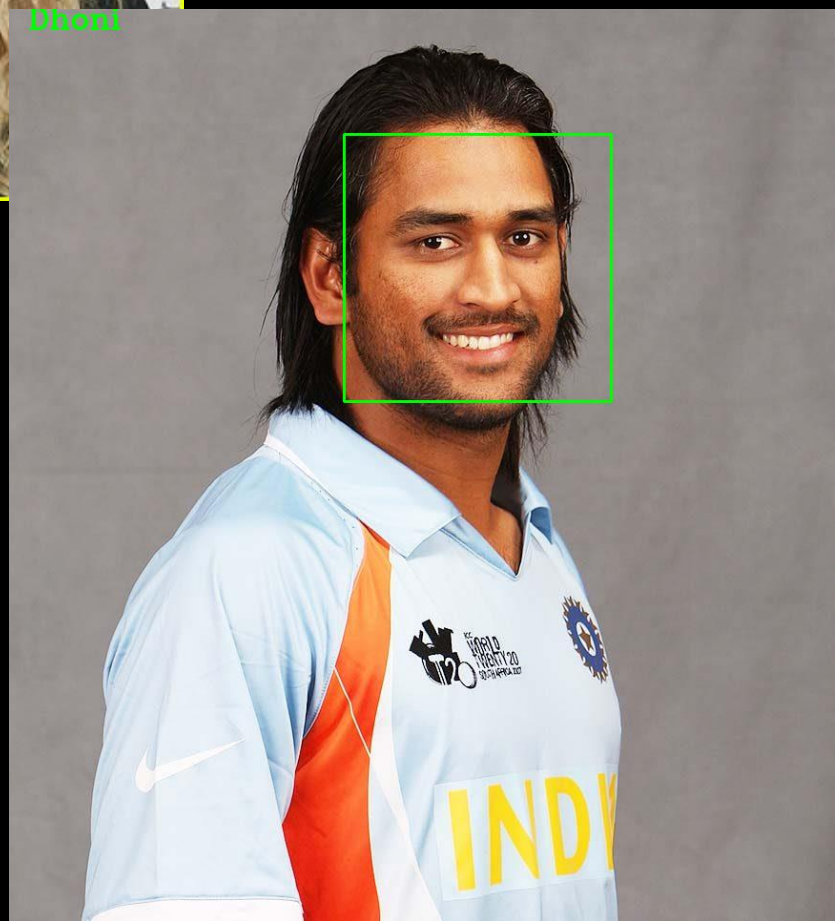


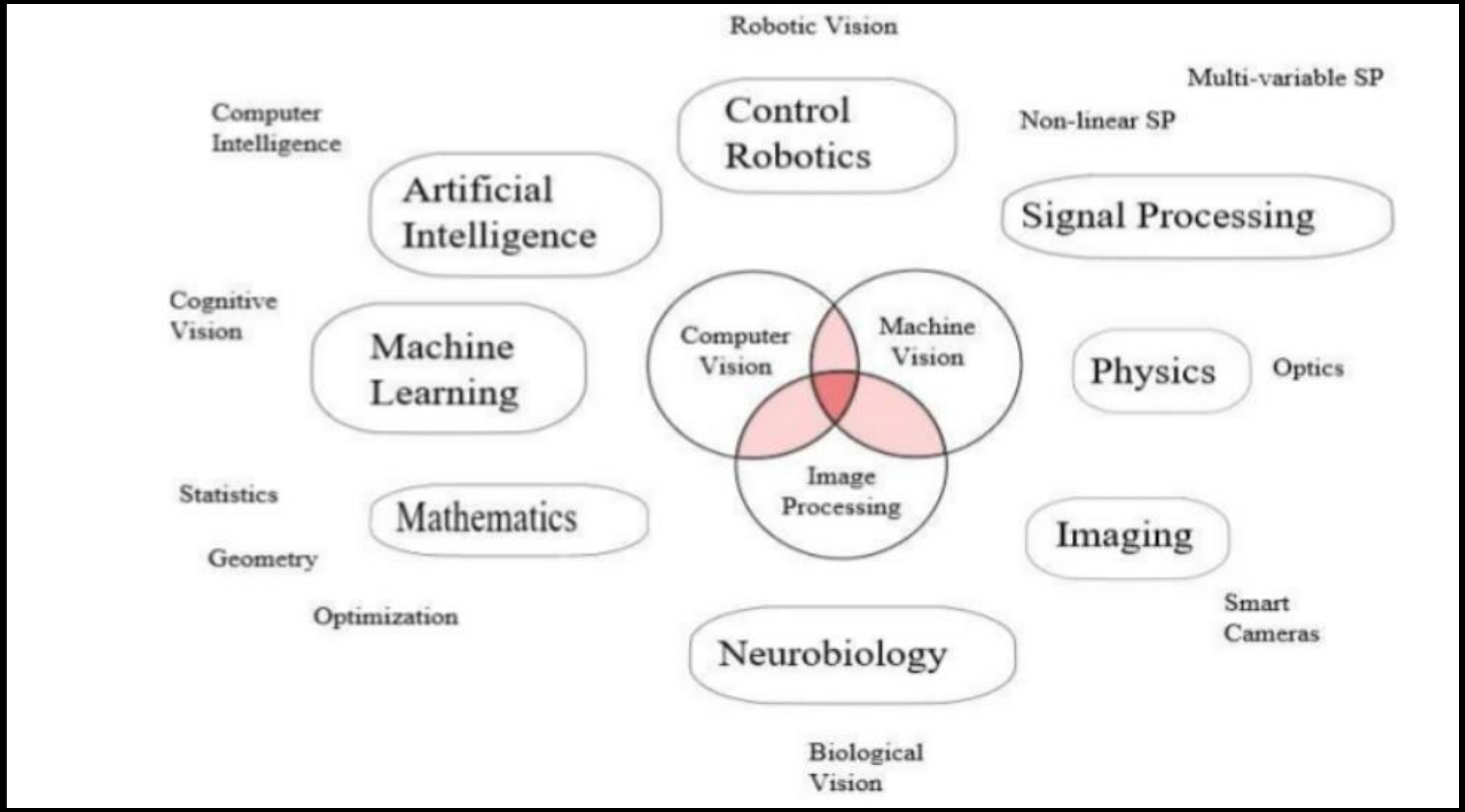

OpenCV_Haar

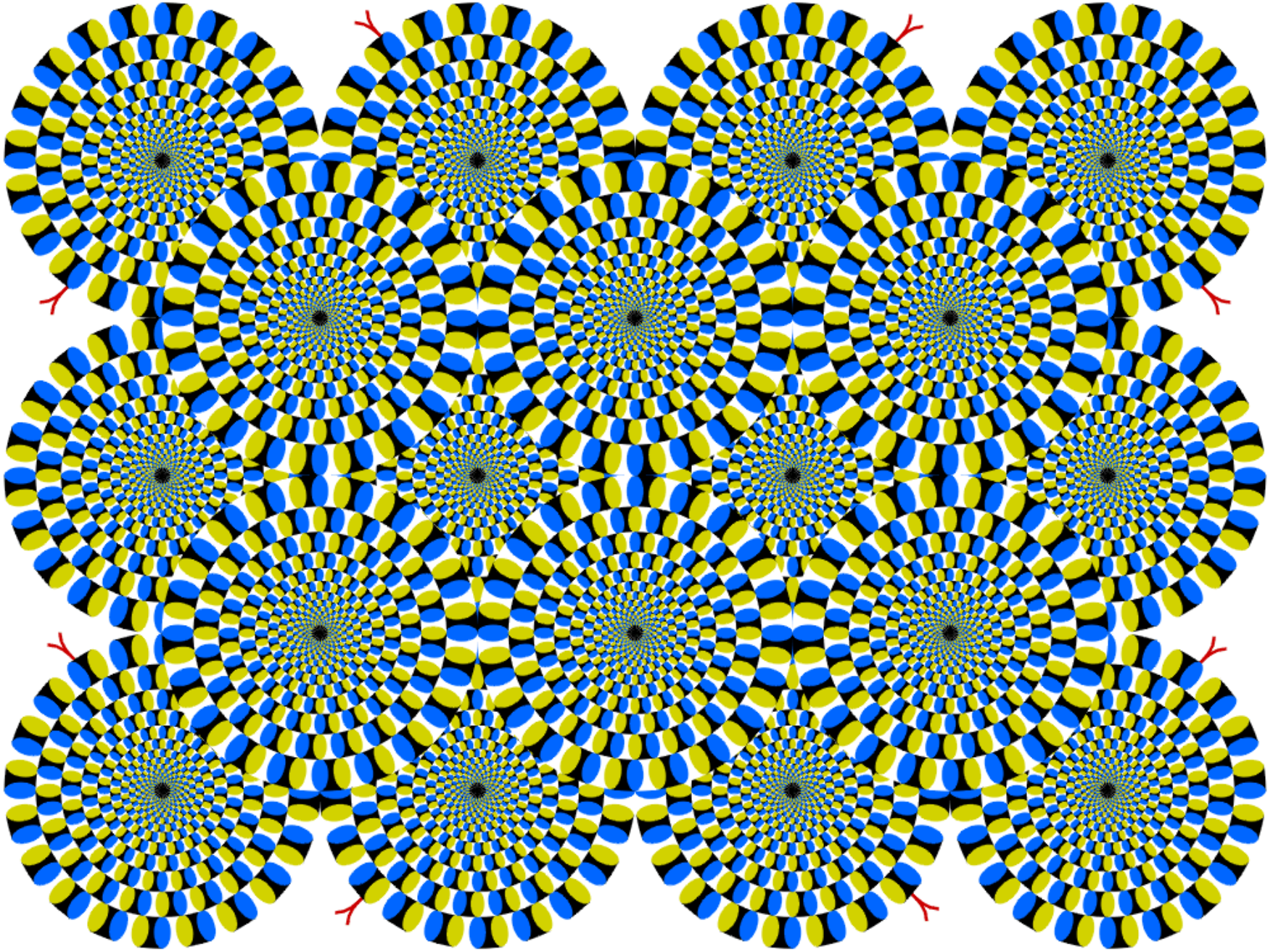

OpenCV_DNN


Madonna

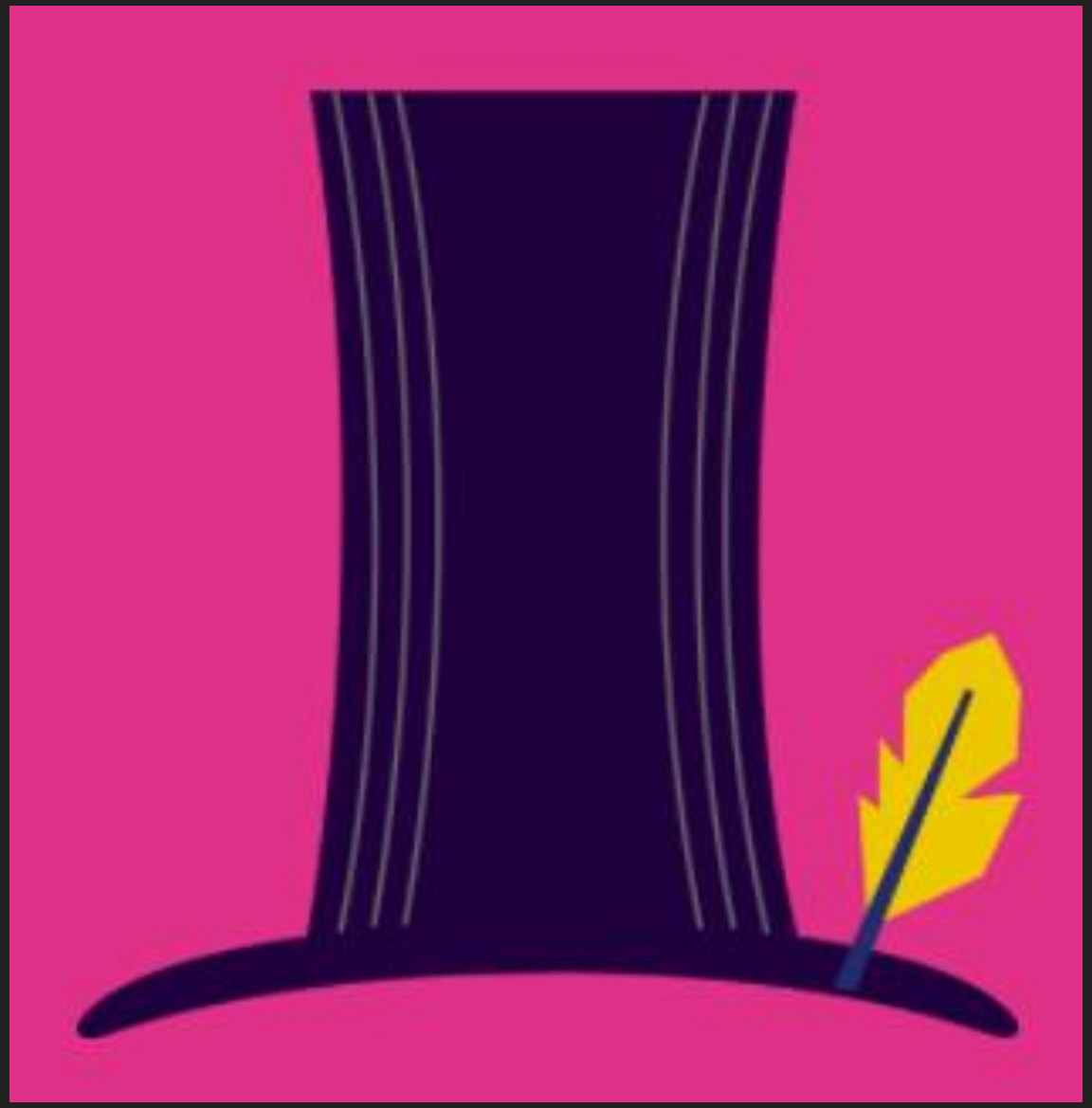


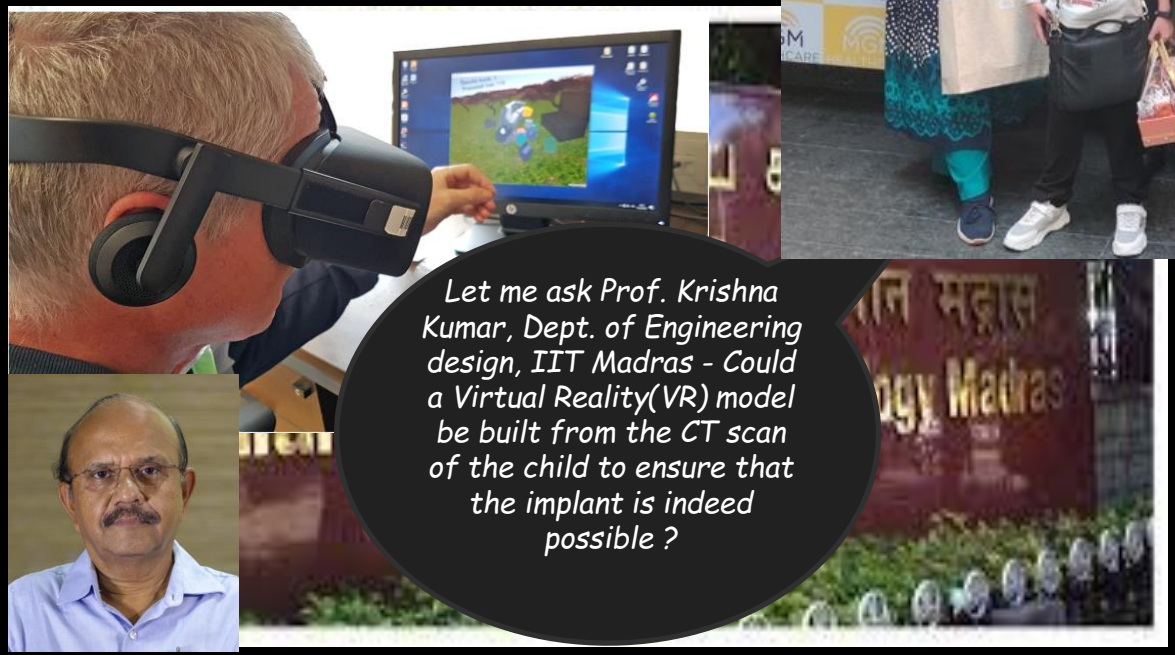

Dhoni


Robotic Vision
Multi-variable SP
Computer Intelligence
Control Robotics
Non-linear SP
Artificial Intelligence
Signal Processing
Cognitive Vision
Machine Learning
Computer Vision
Machine Vision
Physics
Optics
Image Processing
Statistics
Mathematics
Imaging
Geometry
Smart Cameras
Optimization
Neurobiology
Biological Vision



Circular snakes appear to rotate 'spontaneously'
Copyright A.Kitaoka 2003





Presented at the Annual conference of The American Society of Artificial Internal Organs in Chicago, 2020

Nature Inspired Computing (NIC)

What is NIC ?

- **Nature Inspired Computing**, or NIC, a new area of computing science - strives to develop new computing techniques – by observing how naturally occurring phenomena behave - to solve complex problems in various environmental situations

- NIC has produced groundbreaking research that has created new branches, like neural networks, swarm intelligence, evolutionary computation and artificial immune systems

Ant Colony Optimization

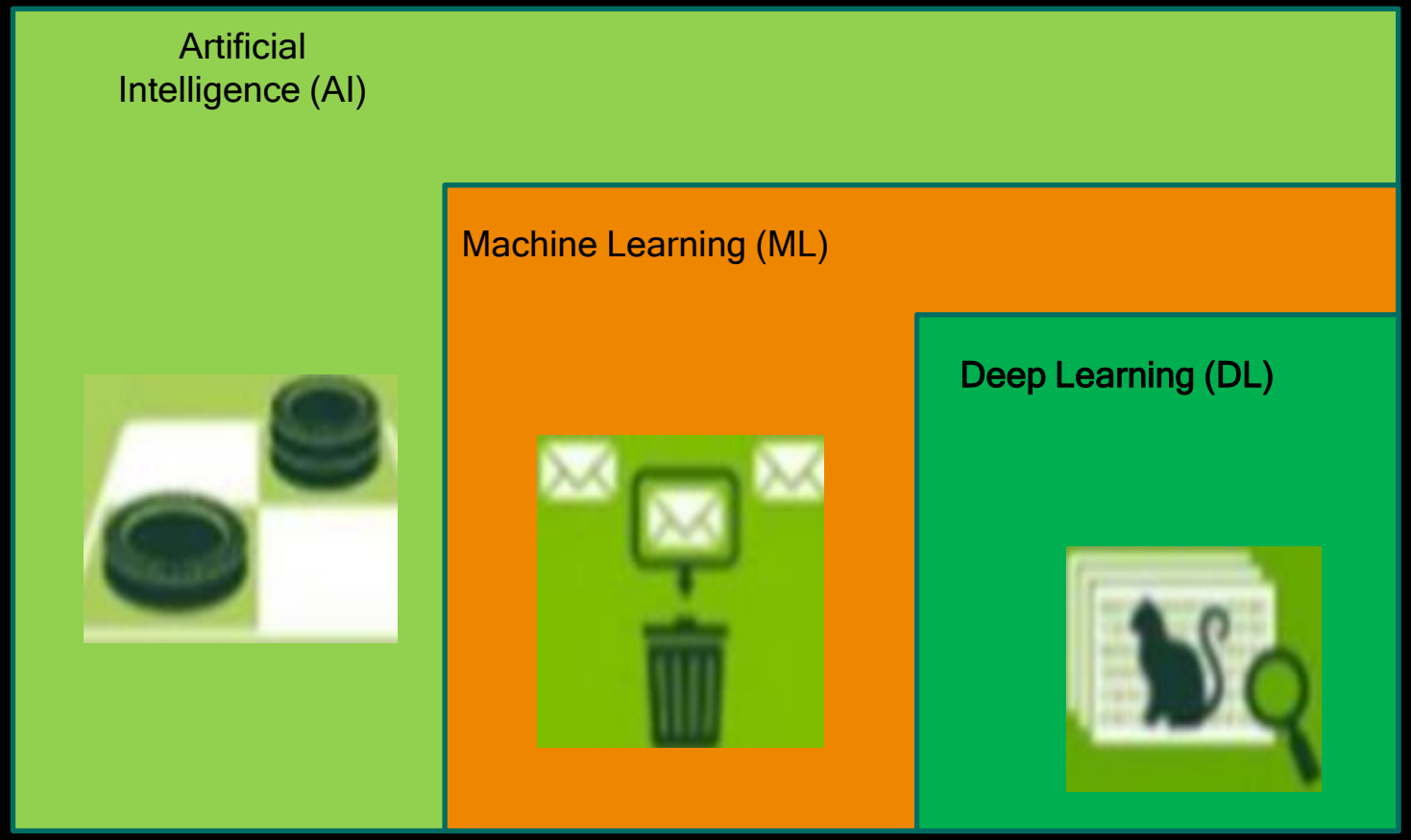
Crow Search Algorithm

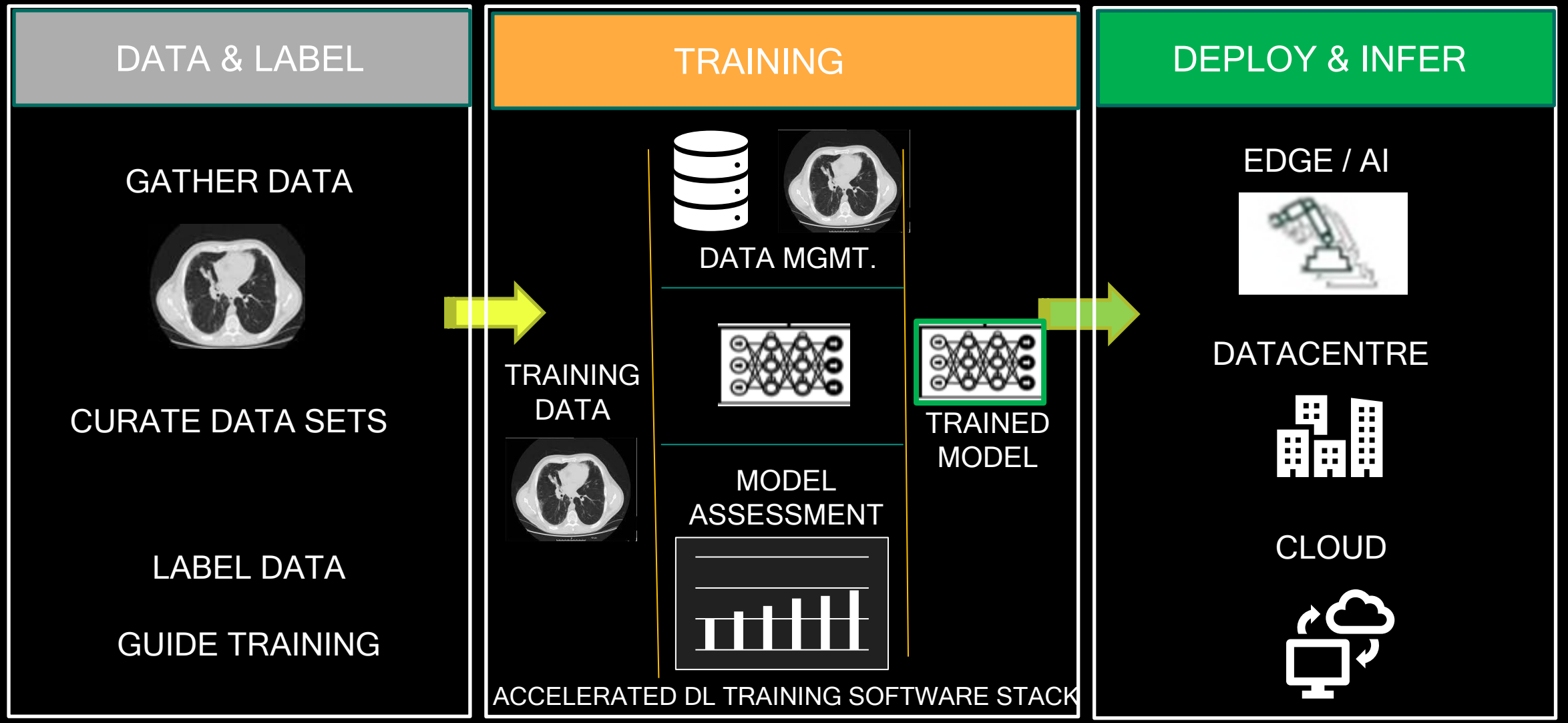
ANN
Evolutionary Computing (Genetic Algorithm)
Firefly Algorithm

Bat Algorithm ...

Grey Wolf Optimization

AI – ML - DL


Artificial
Intelligence (AI)
Machine Learning (ML)
Deep Learning (DL)


DATA & LABEL
GATHER DATA
CURATE DATA SETS
LABEL DATA
GUIDE TRAINING
TRAINING
DATA MGMT.
TRAINING DATA
MODEL ASSESSMENT
TRAINED MODEL
ACCELERATED DL TRAINING SOFTWARE STACK
DEPLOY & INFER
EDGE / AI
DATACENTRE
CLOUD

DATA:GATHER & LABEL

GATHER DATA

CURATE DATA SETS

LABEL DATA

GUIDE TRAINING

TRAINING

Digital Public Goods

PRE-TRAINED MODELS

+ Simple to Implement
+ Achieve good model performance quickly
+ ResNet

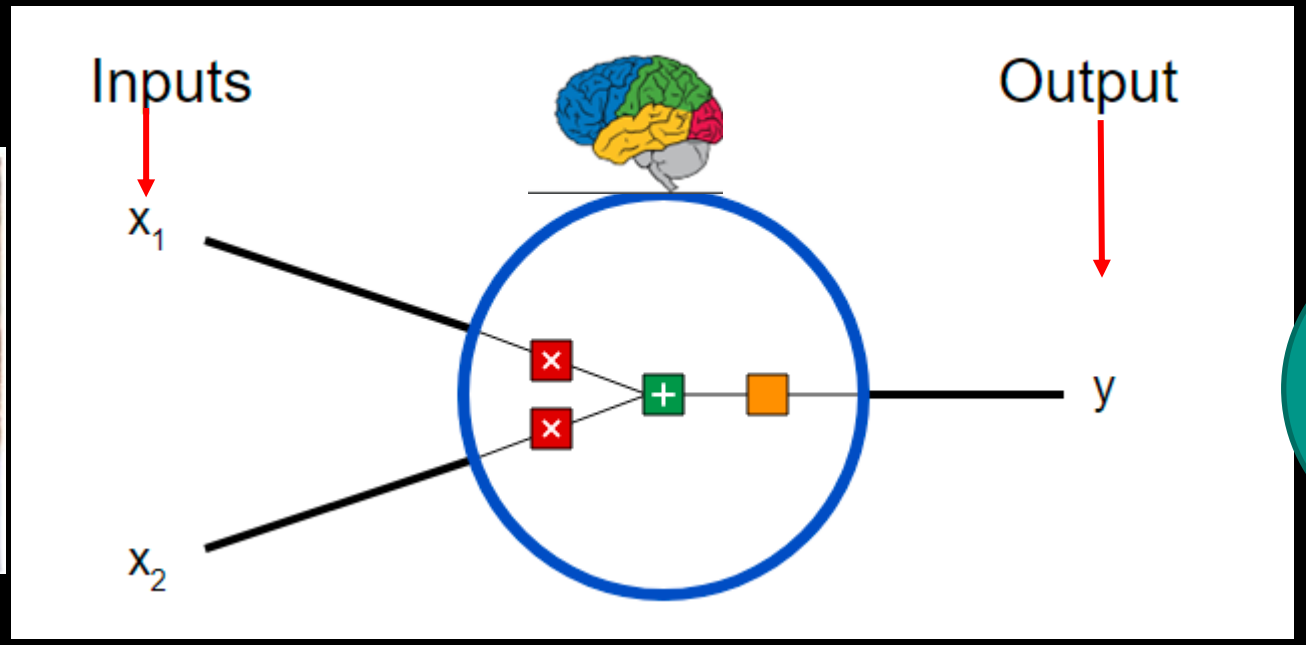
DEPLOY & INFER

EDGE / AI

DATACENTRE

CLOUD

Combine Neurons into Neural Network(NN)



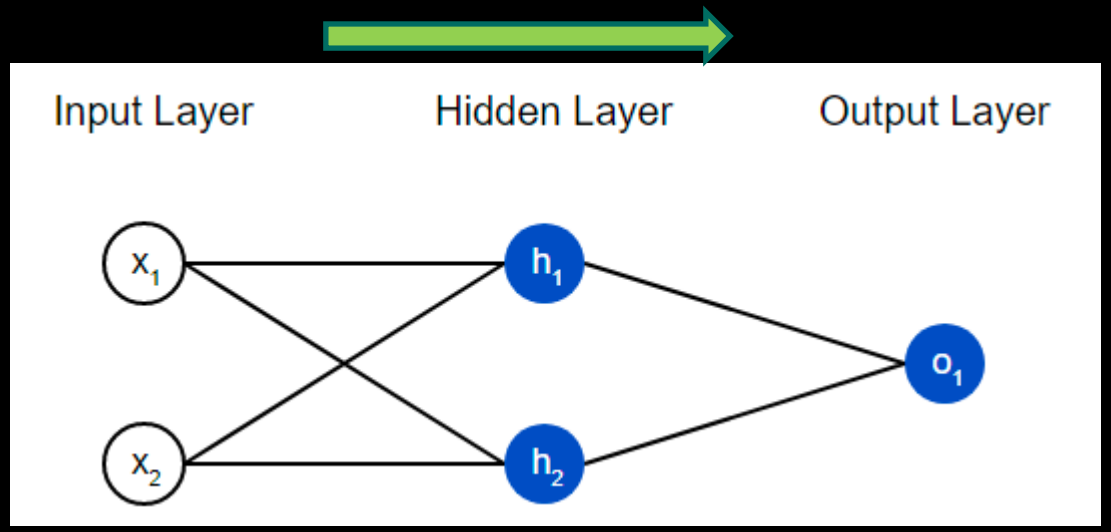
CAT

Input, processing and output together is also known as

Perceptron

Combining Neurons into Neural Network(NN)

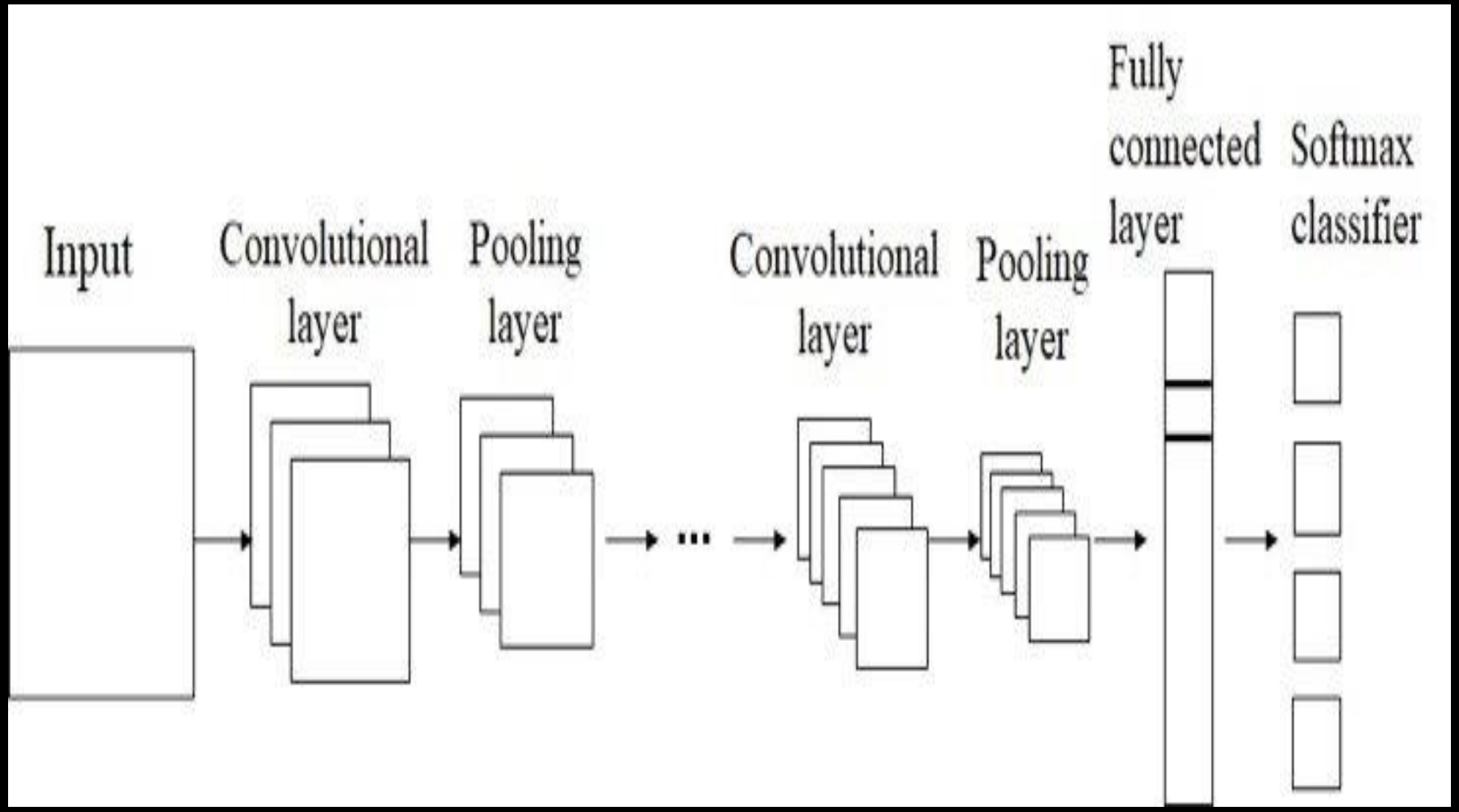
Forward propagation



Back propagation

This network has 2 inputs, a hidden layer with 2 neurons ($h1$ and $h2$), and an output layer with 1 neuron ($o1$). Notice that the inputs for $o1$ are the outputs from $h1$ and $h2$ - that's what makes this a network.

Convolutional Neural Network (CNN)


Input
Convolutional layer
Pooling layer
Convolutional layer
Pooling layer
Fully connected layer
Softmax classifier

IS THIS A
CAT or DOG?

CAT DOG

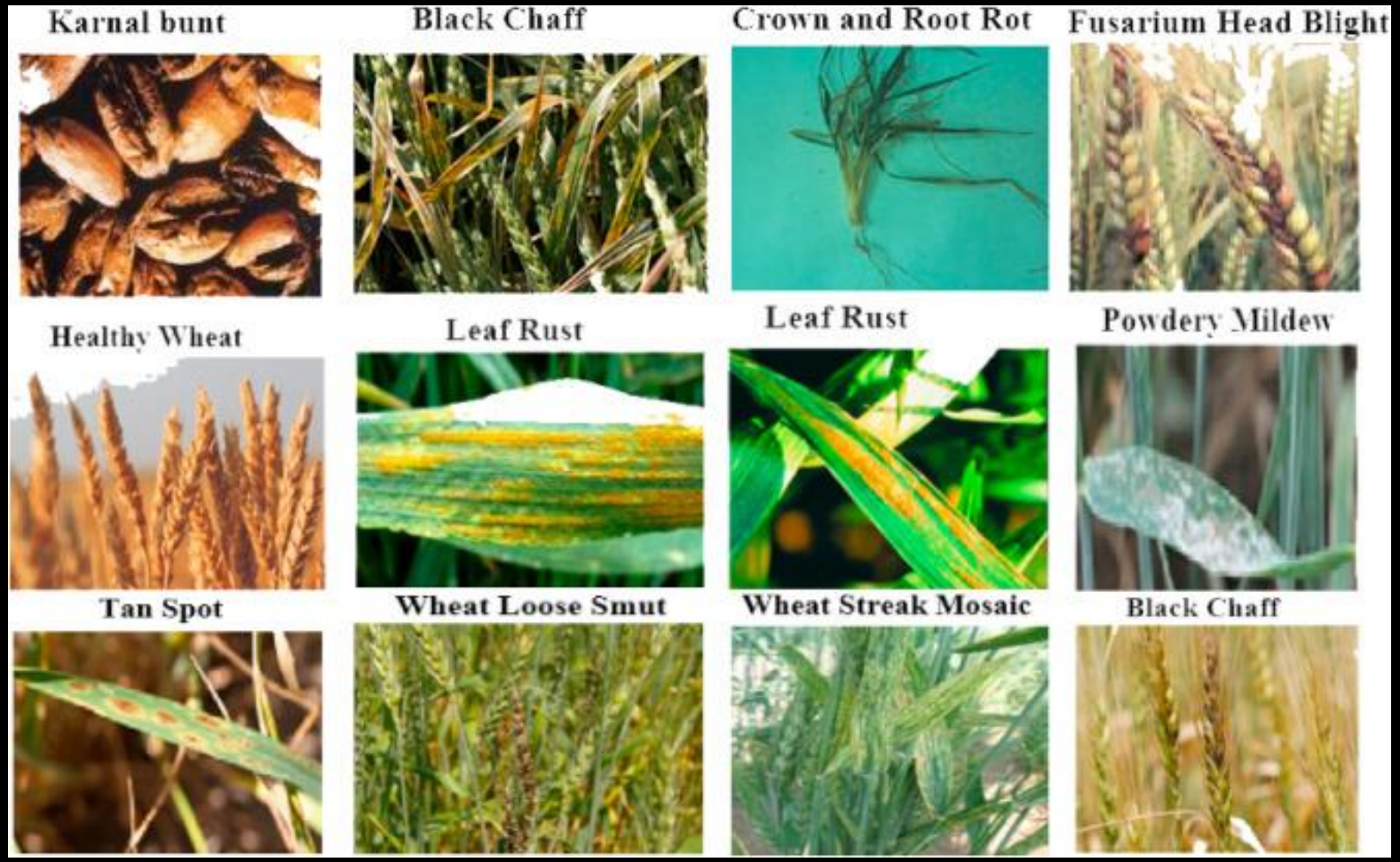
OUTPUT LAYER

ACTIVATED NEURONS

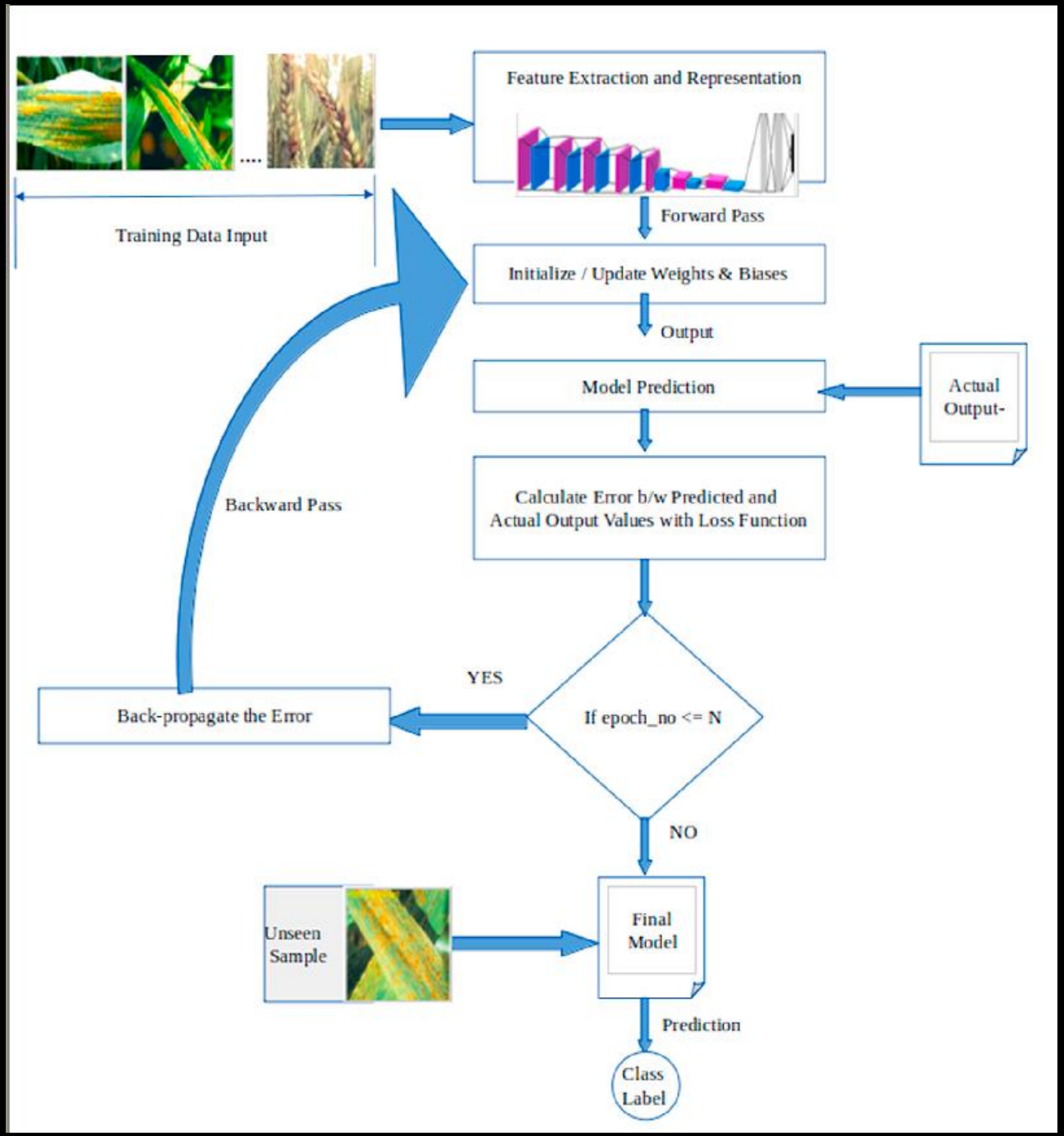
INPUT LAYER

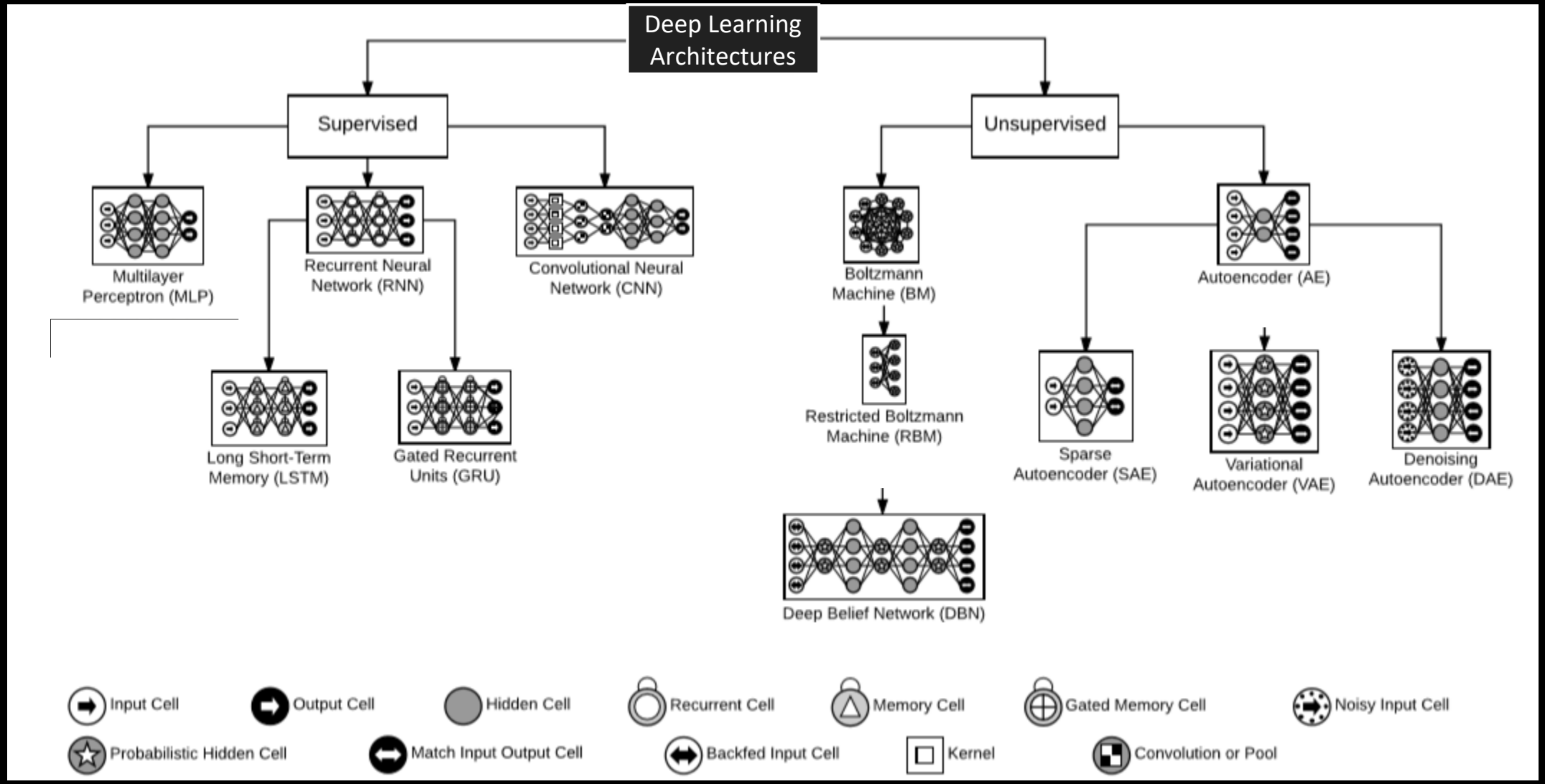
DEEP *neural* NETWORK

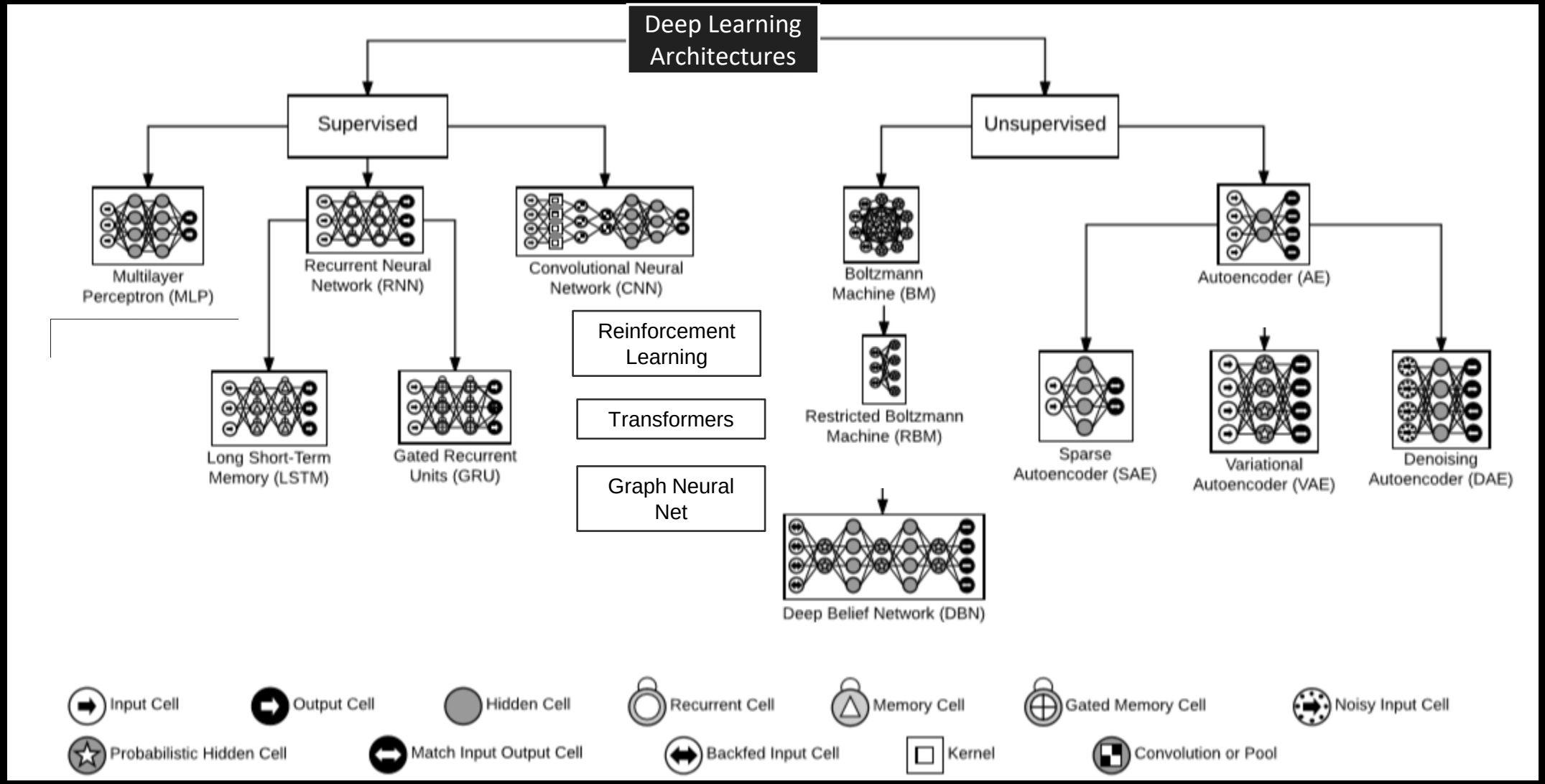
CNN based Wheat Disease Classification



Source: L. Goyal, CM Sharma et al, J. Informatics in Medicine Unlocked, Elsevier


Feature Extraction and Representation
Training Data Input
Forward Pass
Initialize / Update Weights & Biases
Output
Model Prediction
Actual Output-
Calculate Error b/w Predicted and Actual Output Values with Loss Function
Backward Pass
YES
If epoch_no <= N
Back-propagate the Error
NO
Unseen Sample
Final Model
Prediction
Class Label


Deep Learning Architectures
Supervised
Unsupervised
Multilayer Perceptron (MLP)
Recurrent Neural Network (RNN)
Convolutional Neural Network (CNN)
Long Short-Term Memory (LSTM)
Gated Recurrent Units (GRU)
Boltzmann Machine (BM)
Restricted Boltzmann Machine (RBM)
Deep Belief Network (DBN)
Autoencoder (AE)
Sparse Autoencoder (SAE)
Variational Autoencoder (VAE)
Denoising Autoencoder (DAE)
Input Cell
Output Cell
Hidden Cell
Recurrent Cell
Memory Cell
Gated Memory Cell
Noisy Input Cell
Probabilistic Hidden Cell
Match Input Output Cell
Backfed Input Cell
Kernel
Convolution or Pool


Deep Learning Architectures
Supervised
Unsupervised
Multilayer Perceptron (MLP)
Recurrent Neural Network (RNN)
Convolutional Neural Network (CNN)
Boltzmann Machine (BM)
Autoencoder (AE)
Reinforcement Learning
Transformers
Graph Neural Net
Long Short-Term Memory (LSTM)
Gated Recurrent Units (GRU)
Restricted Boltzmann Machine (RBM)
Sparse Autoencoder (SAE)
Variational Autoencoder (VAE)
Denoising Autoencoder (DAE)
Deep Belief Network (DBN)
Input Cell
Output Cell
Hidden Cell
Recurrent Cell
Memory Cell
Gated Memory Cell
Noisy Input Cell
Probabilistic Hidden Cell
Match Input Output Cell
Backfed Input Cell
Kernel
Convolution or Pool

While it is a fact that great progress in research has been achieved in the development of powerful AI algorithms...

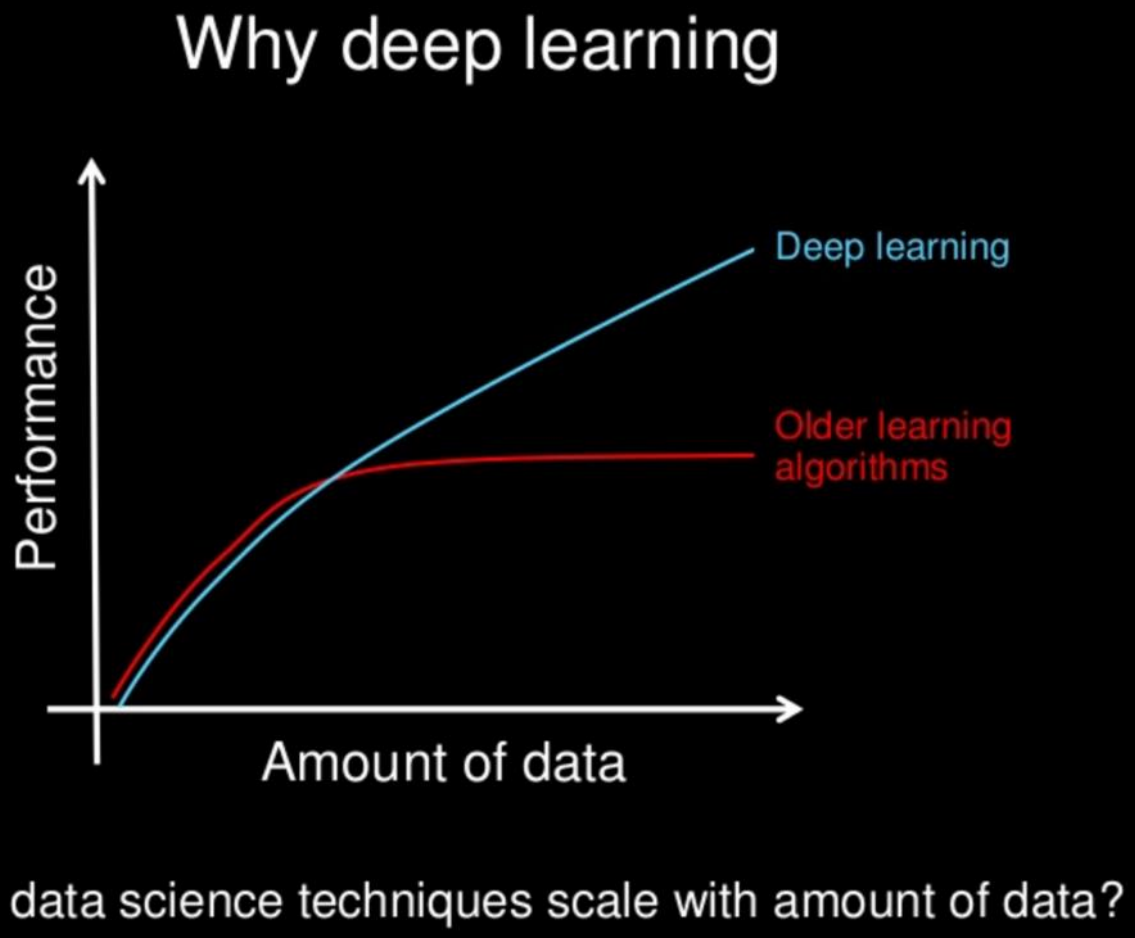
...the need of the hour is ...

Shifting the focus of AI practitioners from model / algorithms to the

quality of data they use to train the models

More Data more performance

Next Natural Resource ?

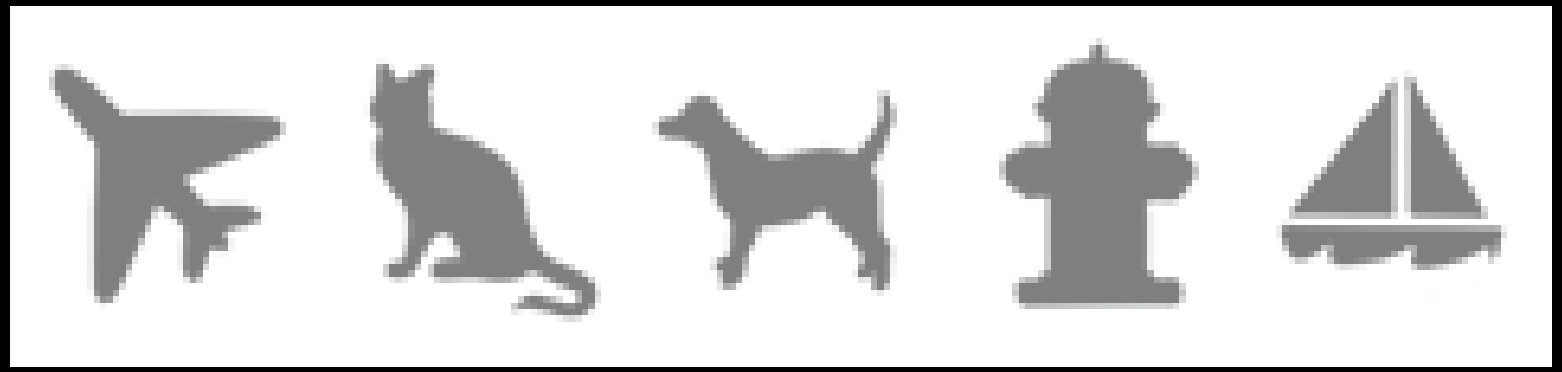


Credit: Prof. Andrew Ng

Data available as Digital Public Goods

- IoT sensors collect raw data
- Open Source Data Repositories
- ImageNet
- CIFAR
- MNIST
- COCO dataset

https://data.gov.in/



COCO (Common Objects in Context) is a large-scale object detection dataset.

- Object segmentation
- Recognition in context
- 330K images (>200K labeled)
- 1.5 million object instances
- 80 object categories
- 5 captions per image

“A data-centric approach, allows people in manufacturing, hospitals, farms, to customize the data, making it more feasible for someone without technical training in AI to feed it into an open-source model”

- Prof. Andrew Ng

Design Thinking elements

in

Data Driven Innovation

Like Design Thinking,

Data Driven Innovation

follows the process of
problem discovery, research, ideation, prototyping and usability testing

with strong emphasis on

Creating Human Centred Solutions to Real World Problems

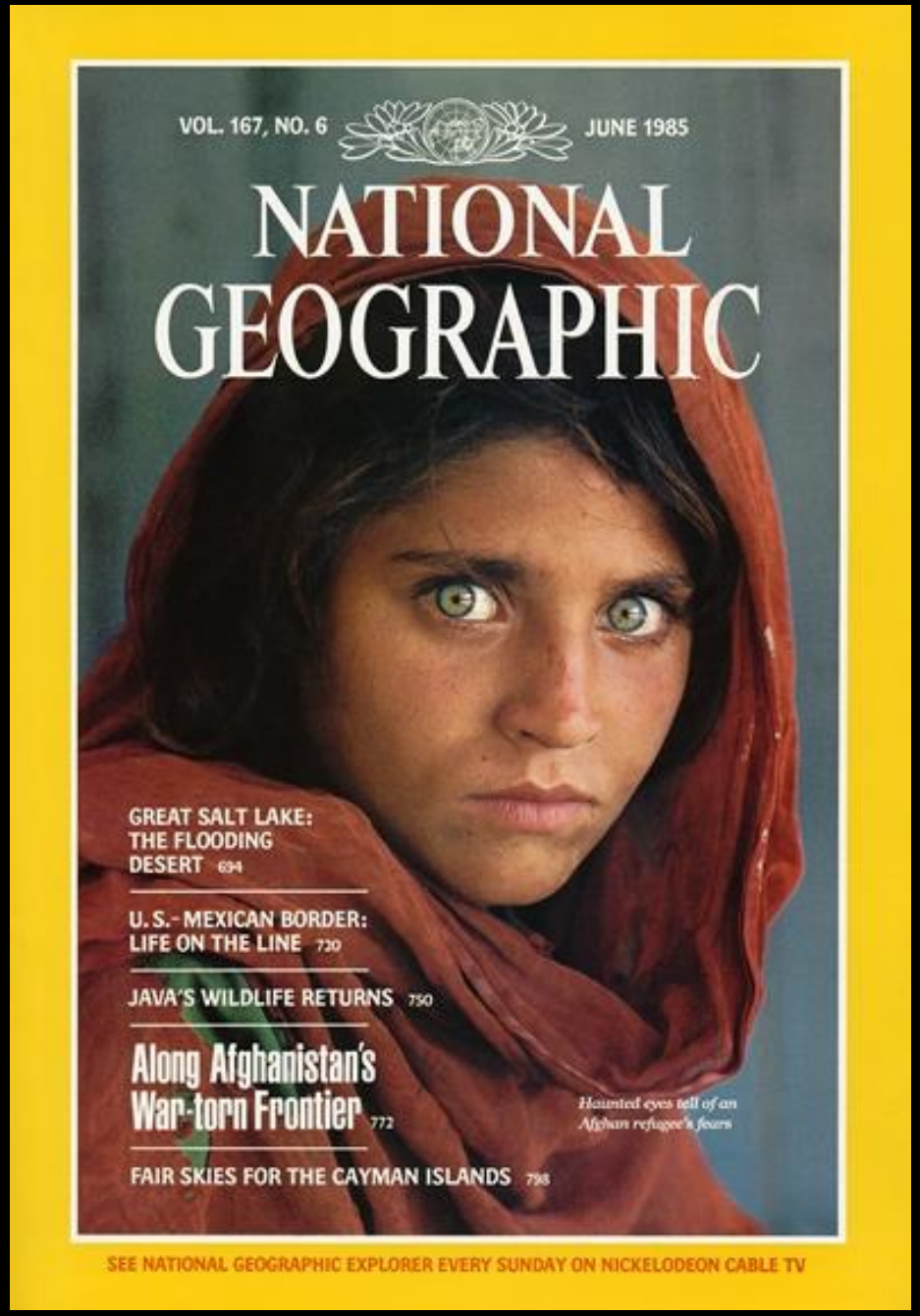
The case in point is …

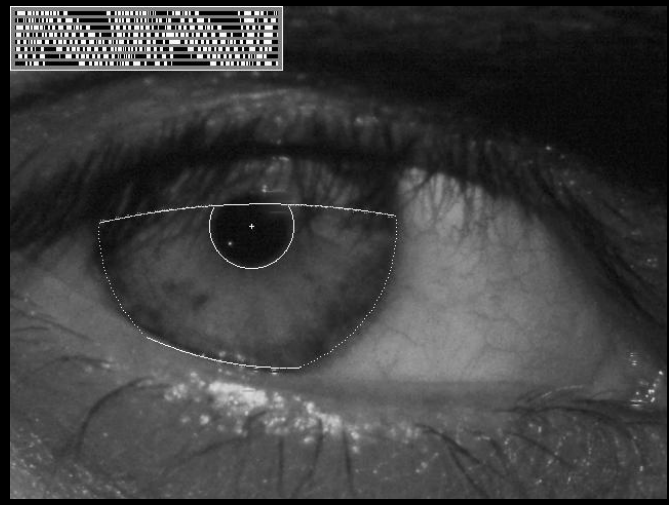
Seeing is Believing

“bionic eyes”

Data Privacy and Human Rights

- Respect for Privacy and protection of Personal Information
- Human Rights and Data Privacy
- Transparent and open discussion on Data privacy & Human Rights


VOL. 167, NO. 6
JUNE 1985
NATIONAL GEOGRAPHIC
GREAT SALT LAKE: THE FLOODING DESERT 694
U.S.-MEXICAN BORDER: LIFE ON THE LINE 720
JAVA'S WILDLIFE RETURNS 750
Along Afghanistan's War-torn Frontier 772
Haunted eyes tell of an Afghan refugee's fears
FAIR SKIES FOR THE CAYMAN ISLANDS 798
SEE NATIONAL GEOGRAPHIC EXPLORER EVERY SUNDAY ON NICKELODEON CABLE TV



Data Driven Smart Solutions

AI based

Pest Management System for Cotton Farming

Based on : “Pest management in Cotton Farms…”, Aman Dalmia, et. al, Applied Data Science Track KDD 20, USA

Good yield of certain crops depend on identification and manual counting of pests and controlling them.

These projects are challenging when it comes to identifying pests accurately and offering guidance to the farmers

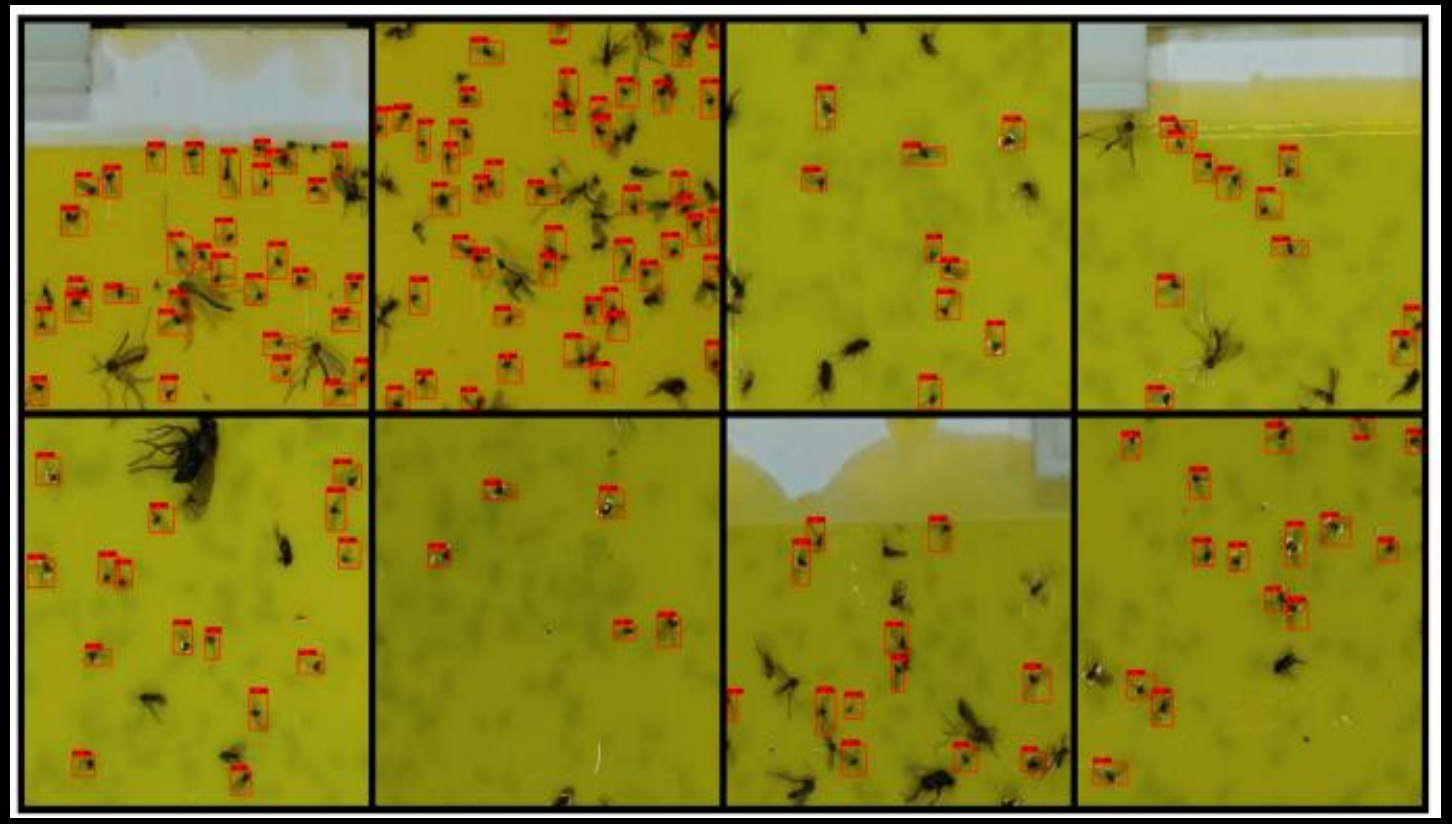
- By constantly monitoring for pests, it may be possible to detect an infestation before it occurs.

- Early detection of pests using traps can also lessen damage to plants.

- A lure based on pheromone, attracts the male moths of the pink bollworm.

- When set up in cotton fields, prone for bollworm infestation, **the trap** competes with the female bollworm moths for the male's attention, disrupting mating and curbing population growth of the pest.

- The male moth lured by the pheromone gets trapped.



Can AI assist farmers in accurately counting pink bollworm moths ?

Cotton Farmer takes photo of Trap containing moths with a phone camera

Photo containing trapped moths

Bounding Box put by AI Object Detection SSD model based on CNN algorithm

- AI Algorithm – CNN
- Data – custom annotated
- Object Detection Model – SSD

- The object detection model puts bounding box and also counts the pink bollworm moths

Using this pest-count and with the help of entomologists, the system guides the farmer on volume, type and timing of pesticide. This advisory is provided in local language

Spray only when both insects and plant-damage are at the threshold levels.

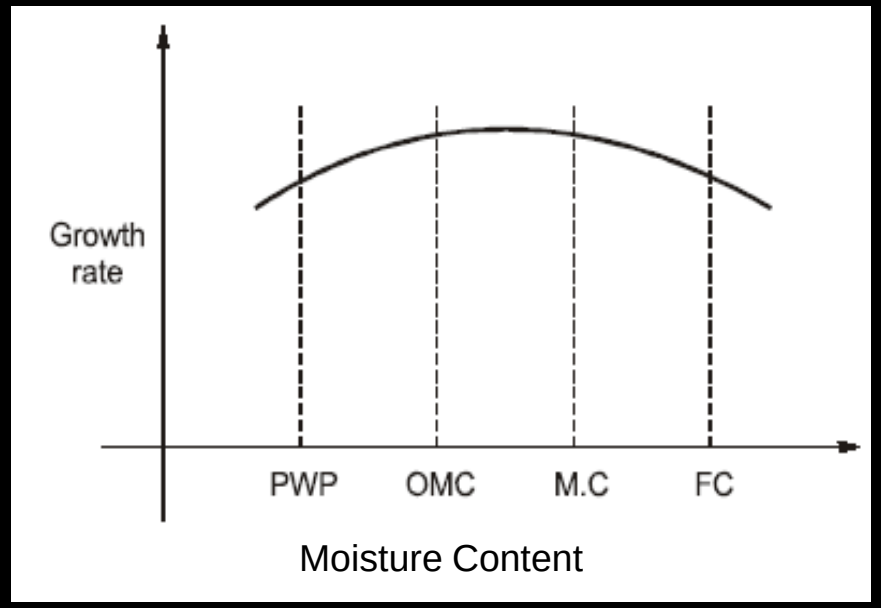
This solution can be used to provide lakhs of farmers with timely, localised advice, reducing crop loss and over-use of pesticides by improving the timing of usage.

During 2020 cultivation , the solution was deployed in four districts across 3 of cotton-producing states- Gujarat, Maharashtra and Telangana.

The team reached out to nearly 15,000 farmers who saw a benefit through increase in profit as well as a reduction in pesticide cost compared to 2019.

Data Driven Precison Irrigation

Replace guess / intuition work with Data-Driven innovation.



PWP - water content below which plant can't extract water (Permanent Wilting Point)

OMC - Optimum Moisture / Water content easily extracted by plant

M.C - Moisture content

FC - Max water that can be held by a field

A sensor-based approach is the most ideal way to measure soil moisture in real-time

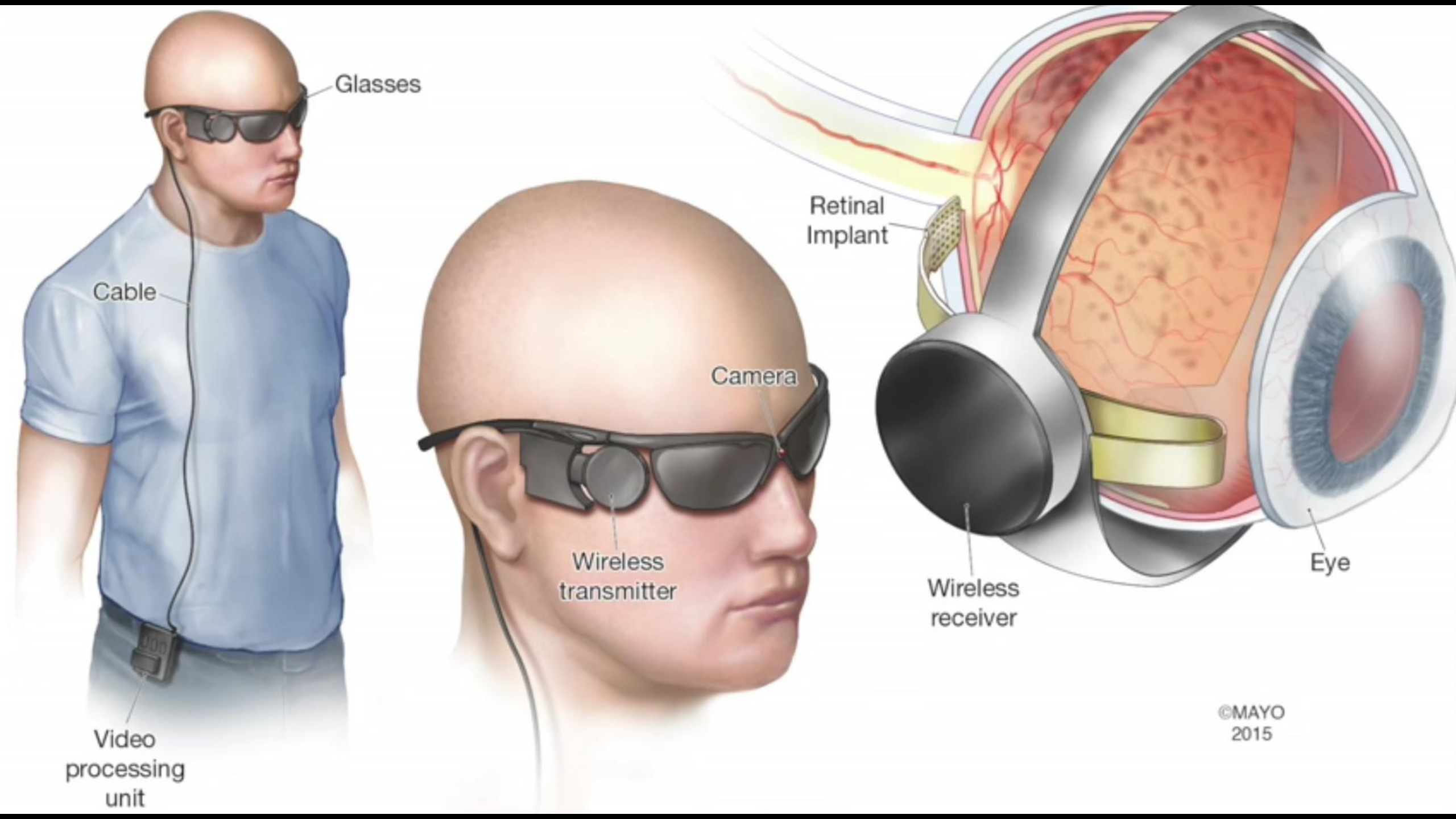
Irrometer Watermark Sensor

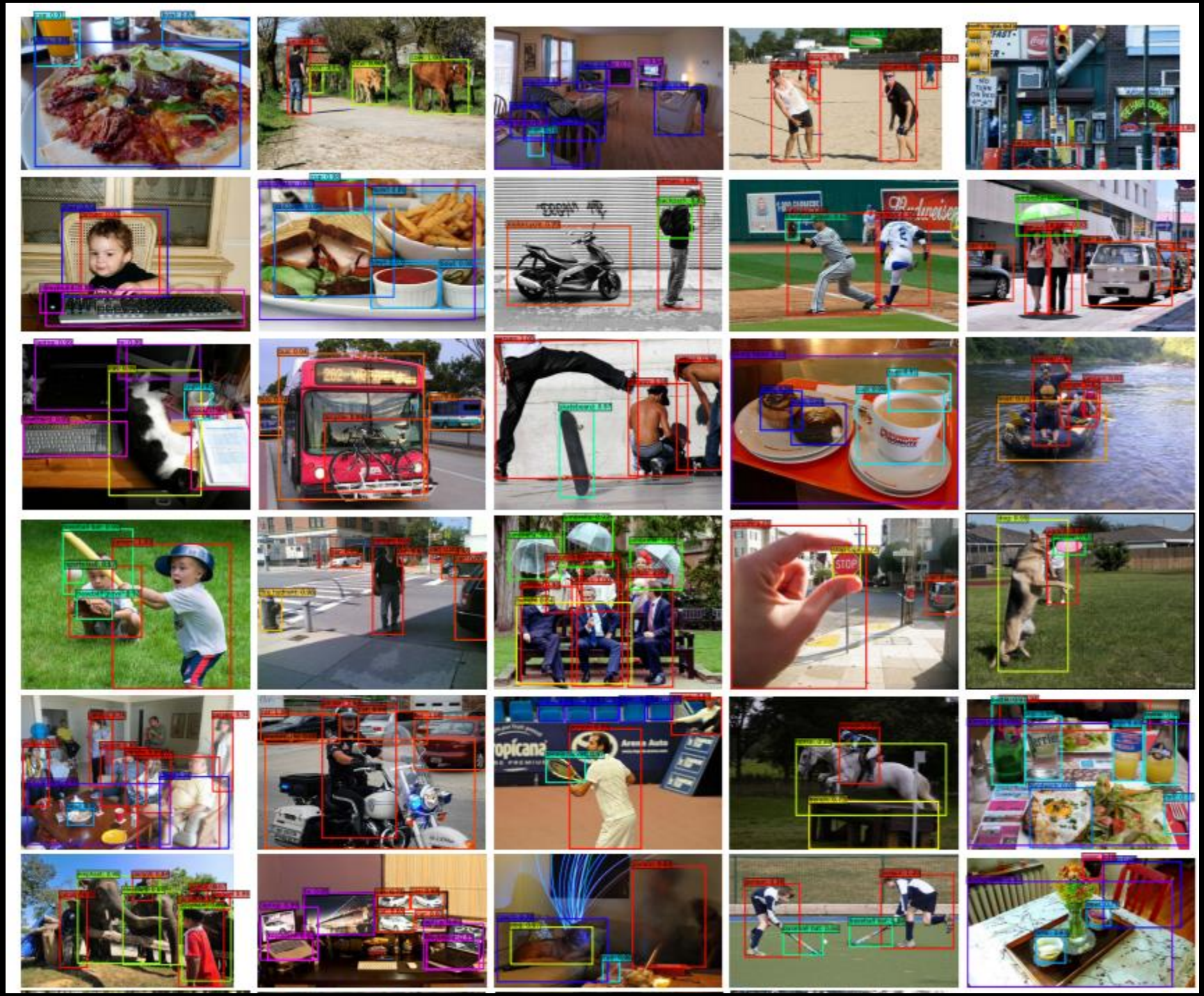
Collect Data from Irrometer, analyze it and alert farmers about watering schedule through a Mobile App

Source/credit: irrometer.com

Value to farmers

Optimal irrigation leads to better quality and a better quantity of crop yield.


Glasses
Cable
Video processing unit
Camera
Wireless transmitter
Retinal Implant
Wireless receiver
Eye
©MAYO 2015


STOP