(12th- 14th September 2023)

Workshop Overview
&
Introduction to Artificial Intelligence, Machine Learning & Neural Networks

Dr. Dinesh Gupta
Group Leader,
Translational Bioinformatics Group
ICGEB, New Delhi

Translational Bioinformatics Group

https://www.icgeb.org/dinesh-gupta.html

http://bioinfo.icgeb.res.in

Research interests:

- Artificial intelligence for development of classification models for complex biological data.
- Comparative Genomics of hosts and pathogens to study evolutionary basis of drug resistance, identify novel drug targets and vaccine candidates
- Molecular modeling and Simulations
- *in silico* screening for novel leads against drug targets in human pathogens
- Development of algorithms for Translational Bioinformatics
- Development of databases

International

Mauro Giacca, King's College, London, UK

Collaborators

Rita Tewari, University of Nottingham, UK
Arnab Pain, KAUST, Saudi Arabia
Pavel Karpov & Yaroslav Blume, Institute of Food Biotechnology and Genomics NAS of Ukraine, Ukraine

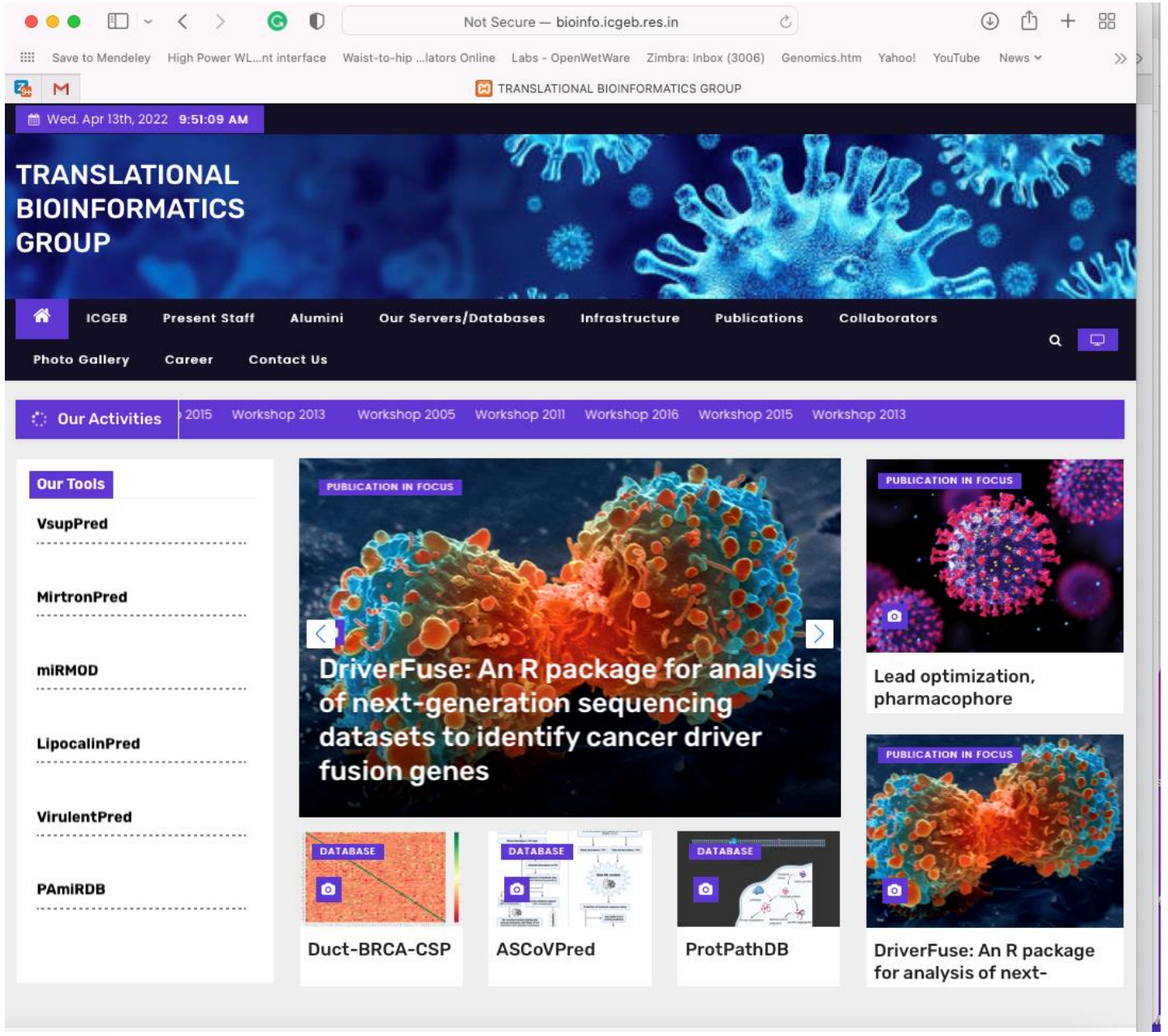
National

Arundhati Sharma, Shipra Agarwal, Radhika Tandon, Viney Gupta: AIIMS, New Delhi
Shashank Tripathi, IISc, Bangalore
Shakir Ali, Jamia Hamdard, New Delhi
Vijay Kumar and Puniti Mathur, Amity University, NOIDA

ICGEB

Asif Mohmmed, Parasite Cell Biology
Pawan Malhotra, Malaria Biology
Shams Yazdani,
Naseem Gaur
Neel S. Bhavesh, Transcriptional Regulation
Dhiraj Kumar, Cellular Immunology
Ranjan Nanda, Translational Health

Bioinformatics Tools (http://bioinfo.icgeb.res.in)



Workshop Overview

- **Cover state-of-art application of AI technologies in modern biology**
- **With large components of practical after-theory explanation**
- **Exploring AI with Google Colab and Advanced Libraries like Keras, PyTorch, etc**
- **Expert Lecture series:** Sharing the research work that will help the participants to understand the AI applications in different fields.
- **25 participants with a diverse backgrounds from all over India**
- **Preference M.Sc. Students > Ph.D.(initial years)>Young faculty**

Learning objectives

- Basics of AI/ML/DL techniques and their application in biology.
- Lectures from experts in the field.
- Hands on session using advanced libraries Pytorch, Keras etc.
- Hands on session using google colab.
- Experience to work with Kaggle datasets.

Final List of Participants (out of 900 applicants!)

Sr No	Participant Name	Institute
1	Devlina Sarkar	University of Calcutta
2	Pooja Chaudhary	Amity University, Mumbai
3	MANISHA RAJPUT	Dayalbagh Educational Institute
4	Ankita kumari	Ranchi university department of botany, Ranchi
5	Sreerupa Mitra	Bioinformatics Center, Savitribai Phule Pune University
6	Kunal Rai	IIT Hyderabad
7	Devvrat	School of Life Science, Khandari, Agra, Uttar Pradesh.
8	RITVIK GUPTA	Manipal Institute of Technology
9	Payal Gupta	Amity University
10	Arsh Roy	Delhi Technological University
11	HARSHITA TIWARI	Banaras Hindu University
12	Biswarup Mahato	Indraprastha Institute of Information Technology,Delhi
13	Dhirendra Singh Yadav	Central Forensic Science Laboratory (CFSL), Directorate of Forensic Science Services, Ministry of Home Affairs, Govt. of India, Posted at CFSL Pune
14	CHAINEE DAS	TEZPUR UNIVERSITY
15	Anshul verma	Central University of South Bihar
16	SHIVANI SHARMA	MMH College, CCS University
17	Ankita Murukesan	Pondicherry University
18	Shaban Ahmad	Jamia Millia Islamia
19	RITIKA	PONDICHERRY UNIVERSITY
20	Shilpa Sharma	Bennett University
21	Pramod Kumar	Army Hospital Research and Referral, New Delhi
22	Vikas Shukla	Indian Council of Medical Research, HQ, New Delhi
23	Govinda Rao Dabburu	South Campus, University of Delhi
24	Aakriti Jain	University of Delhi
25	Renu Jakhar	Department of Biotechnology Indira Gandhi University Meerpur Rewari Haryana

Workshop Organizing Committee and Volunteers

Organizer:

Dr. Dinesh Gupta

Organizing Committee:

Dr. Ankit Singhal

Dr. Shweta Birla Dhakonia

Ms. Minakshi Sharma

Ms.Chhaya Gajra

Ms. Sushmita Sharma

Volunteers:

Dr. Ashish Sharma

Ms. Tamseel Fatma

Mr. Neeraj Chaturvedi

Mr. Hemant Kumar Joon

Ms. Neetu Tyagi

Ms. Lalita Dagar

Dr. Deeksha Pandey

Ms. Nimisha Tiwari

Speakers

Dinesh Gupta

Group Leader,
Translational Bioinformatics Group,
ICGEB, New Delhi

Shweta Birla Dhakonia

Senior Project Scientist,
Translational Bioinformatics Group,
ICGEB, New Delhi

Ankit Singhal

Project Scientist -I,
Translational Bioinformatics Group,
ICGEB, New Delhi

Nimisha Tiwari

Project Associate,
Translational Bioinformatics Group,
ICGEB, New Delhi

D. Sundar

Institute Chair Professor,
IIT, New Delhi

Abhishek Sengupta

Assistant Professor,
Amity University, Noida

Priyanka Narad

Assistant Professor,
Amity University, Noida

Chetan Arora

Professor,
IIT, New Delhi

Ayon Roy

Executive Data Scientist,
NielsenIQ

Neeraj Chaturvedi

Senior Research Fellow,
Translational Bioinformatics Group,
ICGEB, New Delhi

Neetu Tyagi

Pre-Doctoral Fellow,
Translational Bioinformatics Group,
ICGEB, New Delhi

Hemant Kumar Joon

Pre-Doctoral Fellow
Translational Bioinformatics Group,
ICGEB, New Delhi

Ashish Sharma

Principal Project Scientist,
Translational Bioinformatics Group,
ICGEB, New Delhi

Vijay Kumar

PHD Student,
IIT, New Delhi

Manish Kumar

Associate Professor,
Delhi University

Deeksha Pandey

Project Assistant,
Translational Bioinformatics Group,
ICGEB, New Delhi

GPS Raghava

Professor (HAG),
IIIT, Delhi

Day 1 (12.09.2023)		
Timing	**Title**	**Speaker**
09:00 - 09:30 am	**Registration**	
09:30 - 10:00 am	Workshop Inauguration by Director ICGEB	Dr. Ramesh V. Sonti, ICGEB
	Introductory Talk by the Organizer	**Dr. Dinesh Gupta, ICGEB**
10:00 - 10:30 am	Introduction to Linear Classifiers and Key Deep Learning Concepts	**Dr. Shweta Birla Dhakonia, ICGEB**
10:30 - 11:00 am	Hands-on Application of Linear classifier of Perceptron Vs linear SVM for patient treatment classification using Electronic Health Records.	**Dr. Ankit Singhal,ICGEB** *Assisted by Dr. Shweta & Ms. Nimisha*
11:00 - 11:30 am	**Tea Break**	
11:30- 12:00 pm	Basics of Image (computer vision), Convolution Neural Network & Tansfer Learning Technique for Image Classification	**Ms. Nimisha Tiwari, ICGEB**
12:00 - 01:00 pm	Hands-on Image Classification using CNN & Transfer Learning Models	**Ms. Nimisha Tiwari, ICGEB** *Assisted by Dr. Shweta, Ms. Tamseel & Dr. Ankit*
01:00 - 02:00 pm	**Lunch**	
02:00 - 03:00 pm	Leveraging representation learning for drug discovery	**Dr. D. Sundar, IITD**
03:00 - 04:00pm	Revolutionizing Reproductive Health: The Role of AI in Optimizing Fertility and Maternal Care	**Dr Abhishek Sengupta/Dr. Priyanka** Narad, Amity University
04:00 - 04:30 pm	**Tea Break**	
04:30 - 05:30 pm	Hands-on AI application using Exploratory Data Analysis (EDA), data visualization, and ML algorithms for prediction modeling.	**Dr Abhishek Sengupta/Dr. Priyanka** Team

Day 2 (13.09.2023)		
Timing	**Title**	**Speaker**
09:00-10:00 am	Learnable Query Initialization for Surgical Instrument Instance Segmentation	**Dr. Chetan Arora, IITD**
10:00-11:00 am	Exploring Parkinson's Disease Progression Prediction Dataset on Kaggle	**Mr. Ayon Roy, NielsenIQ**
11:00 - 11:30 am	**Tea break**	
11:30 - 12:00	Deep Learning models for identifying angle dysgenesis in-vivo using glaucoma ASOCT images	Dr. Shweta Birla Dhakonia, ICGEB
12.00-1.00 pm	Self-Practice/Problem-Solving/Informal session	All of us
01:00 - 02:00 pm	**Lunch**	
02:00 - 03:00 pm	Talk & Demo-on Machine Learning Assisted Drug Discovery for SARS-CoV-2	**Mr. Neeraj Chaturvedi, ICGEB** *Assisted by Ms. Neetu, Ms. Chhaya, Ms. Nimisha & Dr. Ashish*
03:00- 04:00 pm	Hands-on Vision Transformers for Brain Tumor Classification in MRI Images.	**Mr. Rohan & Ms. Ankita, IITD**
04:00 - 04:30 pm	**Tea break**	
04:30 - 05:30 pm	Session continued	**Mr. Rohan & Ms. Ankita, IITD**

Day 3 (14.09.2023)		
Timing	**Title**	**Speaker**
09:00 - 09:45 am	Lecture on Role of AI in Unlocking Gene Expression Patterns	**Ms. Neetu Tyagi, ICGEB**
09:45 - 10:45 am	Hands-on Session for Handling Gene Expression Data and analysing it via AI	**Mr. Hemant, Ms. Neetu & Ms. Lalita** ICGEB
10:45 - 11:15 am	Type your text **Tea break**	
11:15 - 12:00 pm	Hands-on Implication of AI driven classification models on metabolomics profiling dataset: Breast cancer case study	**Dr. Ashish Sharma, ICGEB** *Assisted by Ms. Tamseel, Ms. Chhaya & Ms. Deeksha*
12:00 - 01:00 pm	Deep Learning Assisted Retinopathy of Prematurity Screening	**Mr. Vijay Kumar,IITD**
01:00 - 02:00 pm	**Lunch**	
02:00 - 03:00 pm	Using Machine-Learning to Discern the Antimicrobial Resistance Profile of Microbes	**Dr. Manish Kumar, DU**
03:00 - 04:00 pm	Hands-on Approach to Modelling and Understanding Hidden Markov Models	**Dr. Deeksha Pandey, ICGEB** *Assisted by Dr. Sangeeta, Dr. Sonu & Ms. Tamseel*
04:00 - 05:00 pm	Expert Lecture Series	**Dr. GPS Raghava, IIITD**
05:00 - 05:30 pm	**Clossing Session with High Tea**	

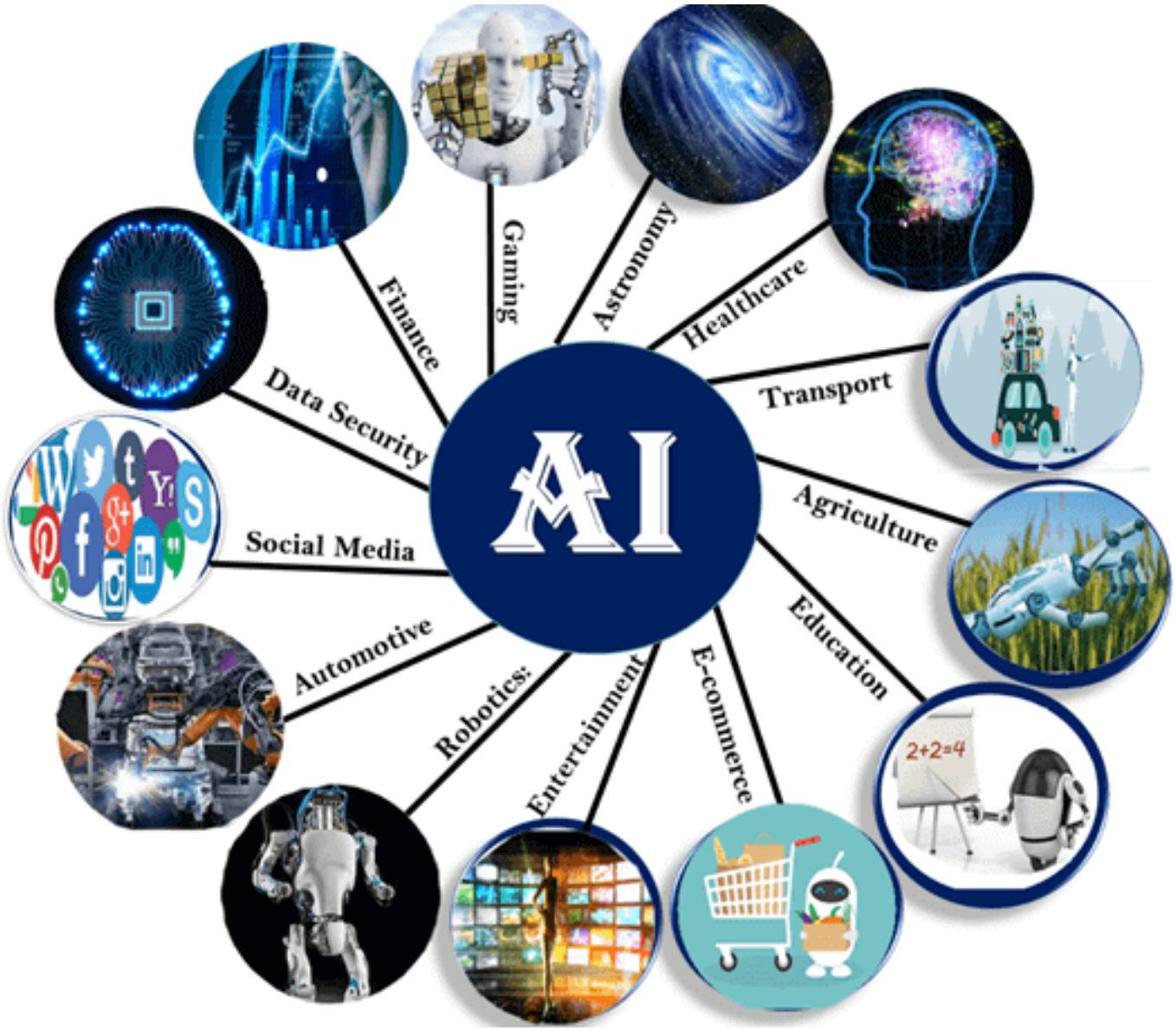
Workshop information

- Website URL: https://apexbtic.icgeb.res.in/aiworkshop2023/
- YouTube Streaming: https://www.youtube.com/channel/UCA-JOGIXz2krAOWl_A1BfdQ

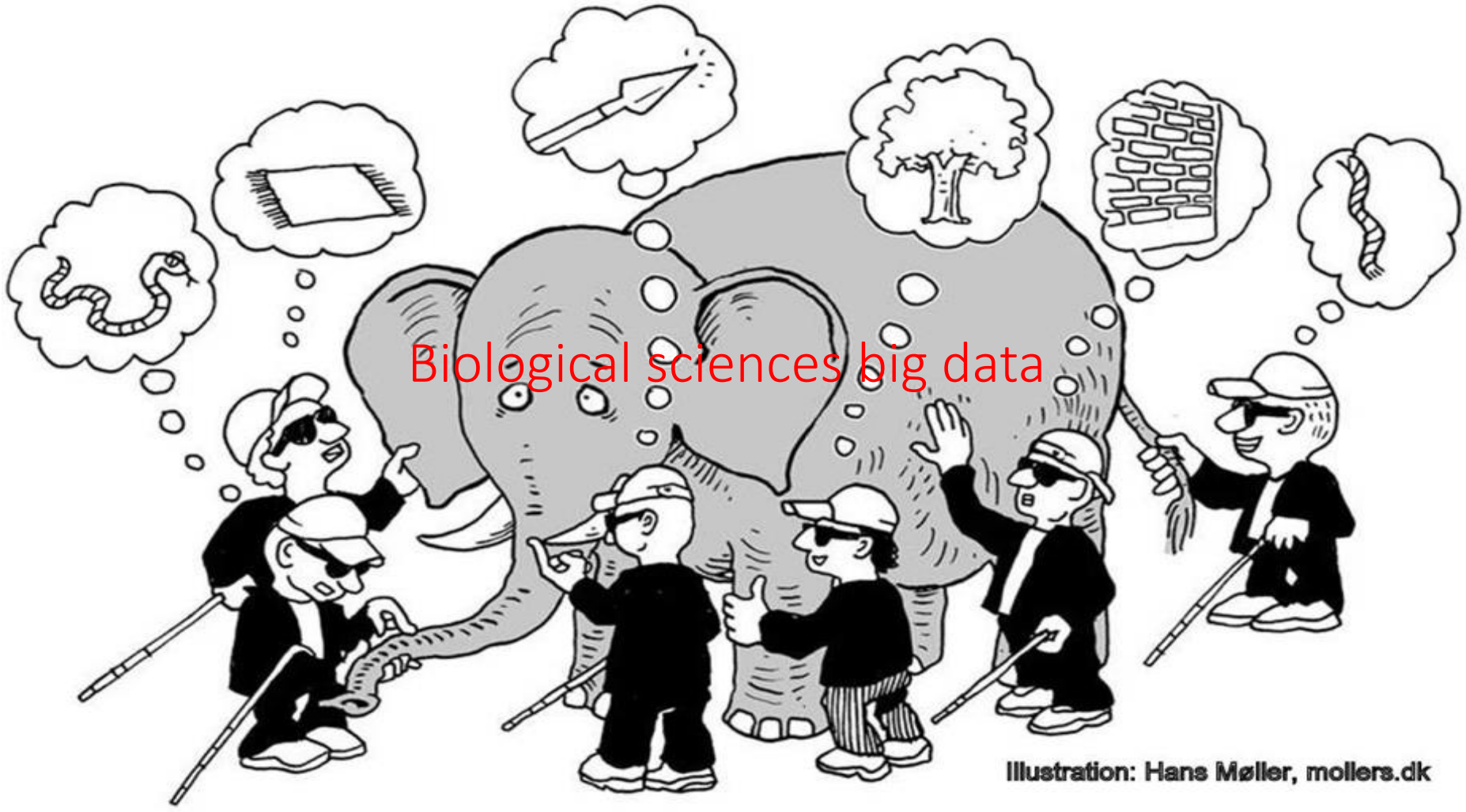
Introduction

Artificial Intelligence, Machine Learning & Neural Networks

Applications of AI

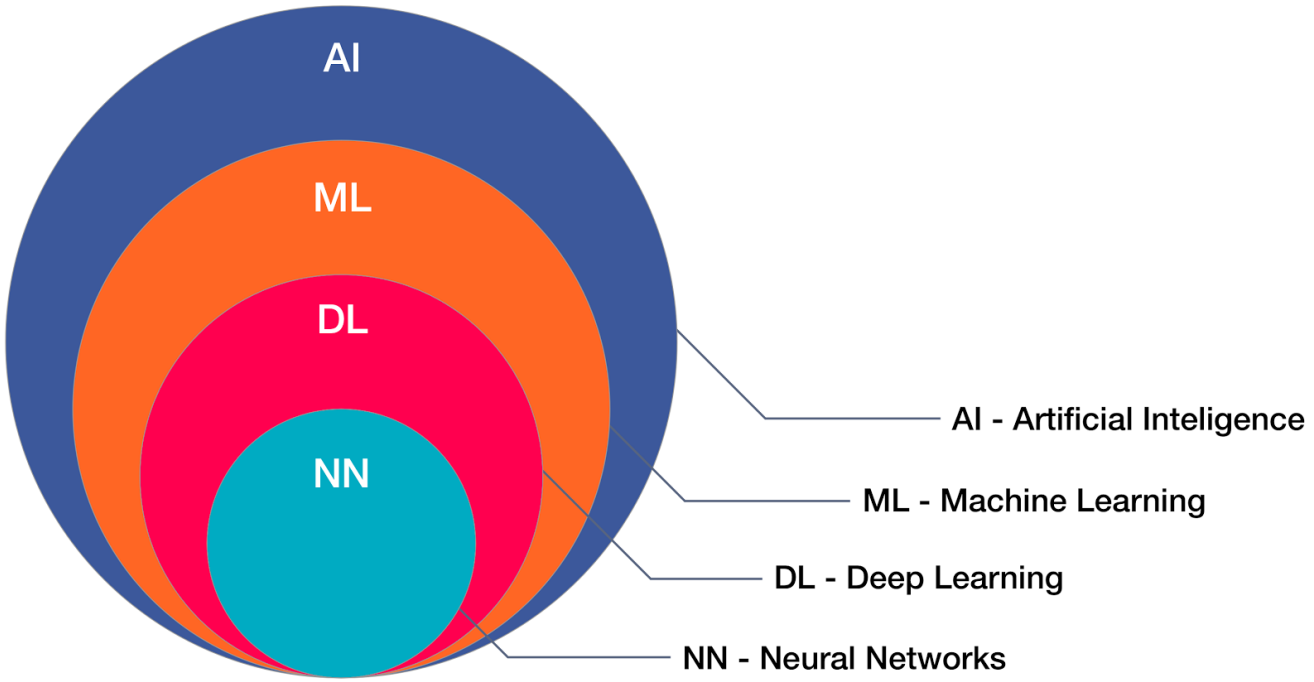


(javapoint.com)


Biological sciences big data
Illustration: Hans Møller, mollers.dk

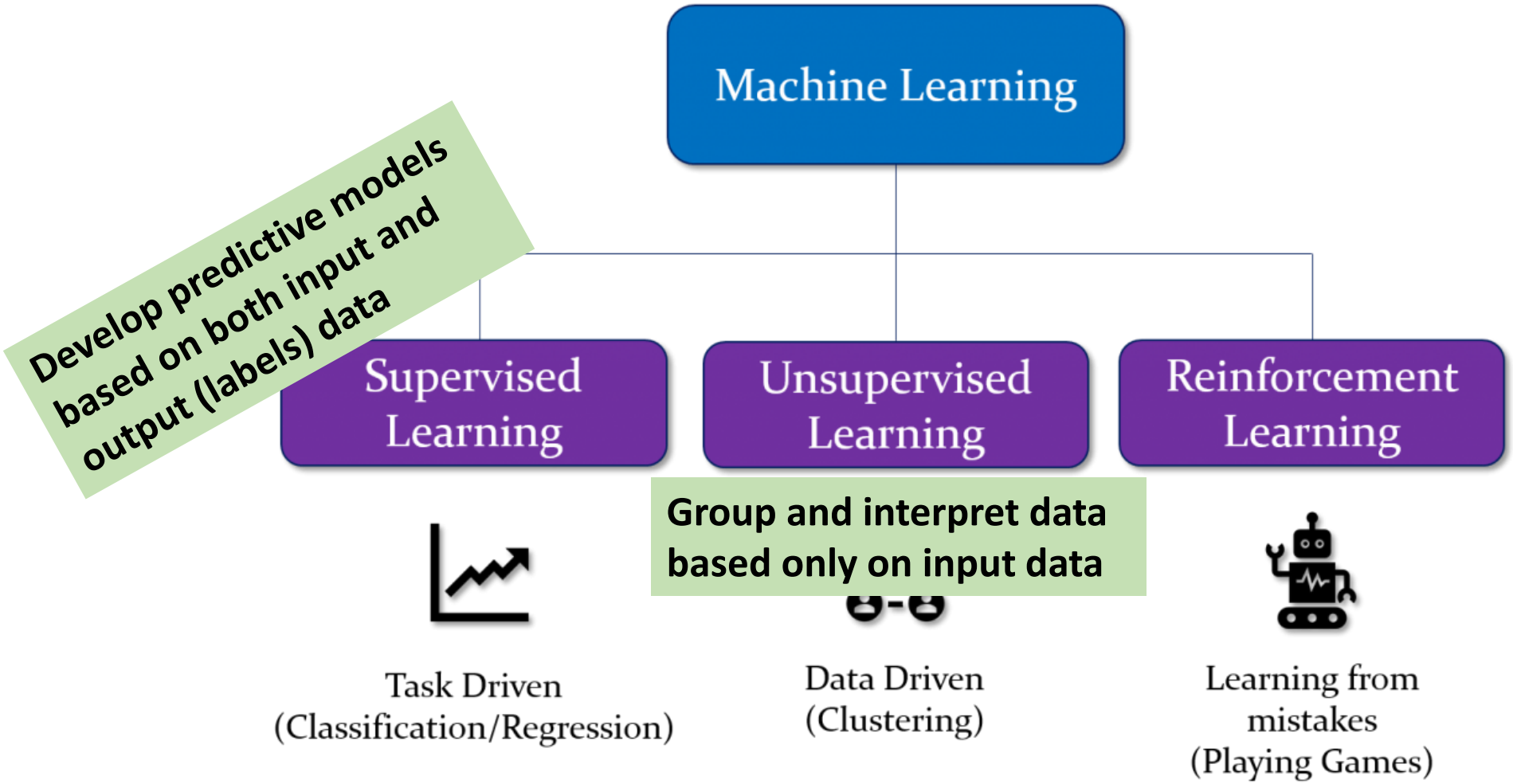
Introduction: Artificial Intelligence and its Subsets

- **Artificial intelligence (AI)** "*simulation of human intelligence processes by machines*" - development of algorithms and models that enable computers to perform tasks that typically require human intelligence e.g. problem-solving, decision-making, understanding, and learning from experience.
- **Machine learning (ML)**is a subset of AI, which gives computers the ability to learn without being explicitly programmed. It learns patterns and relationships from data and uses this knowledge to make predictions or decisions.
- **Deep Learning (DL)** is a branch of Machine learning which makes use of layers of artificial neural networks, inspired by human brains.
- **Neural Networks (NN)** are a subset of Deep Learning, which is a network designed to work like human brains on large data.



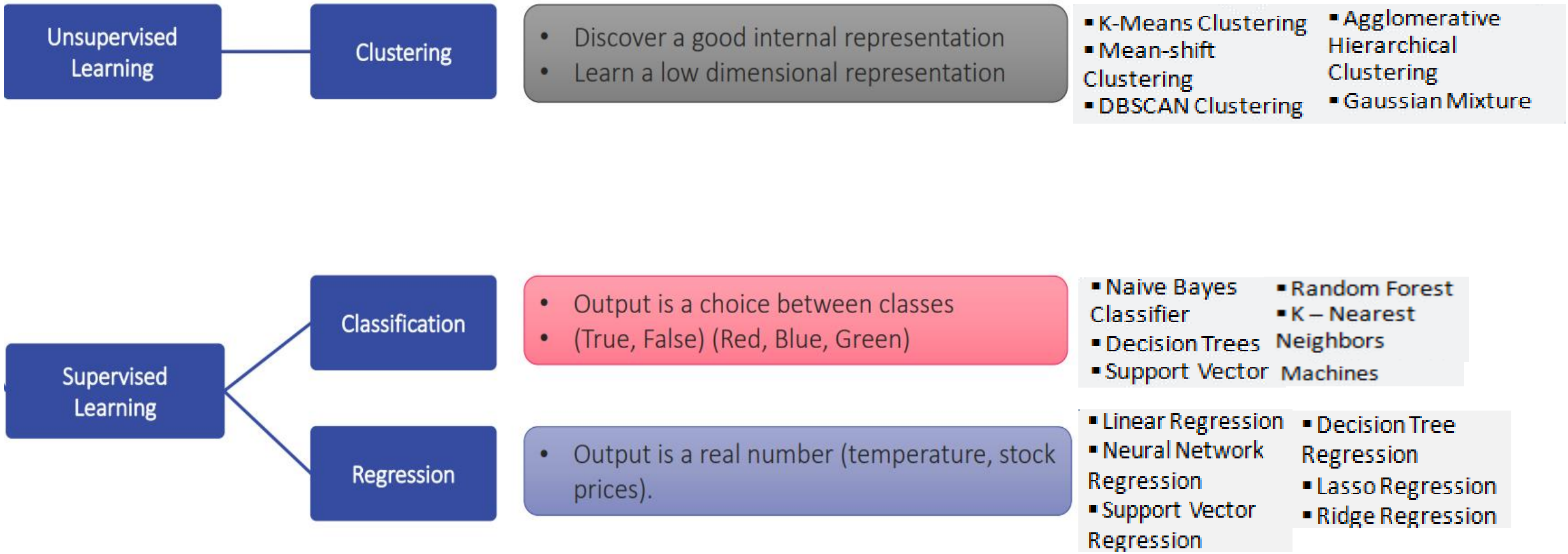
Types of Machine Learning

"Machine learning is essentially a statistical model"



(adapted from: towardsdatascience.com)

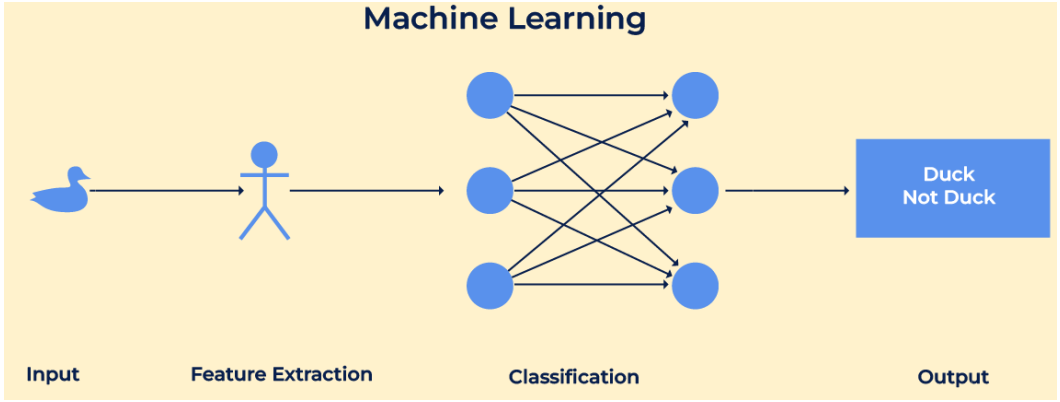
Subtypes of Machine Learning



Basic ML Steps

- **Dataset preparation**: Cleaning the dataset, imputing missing data, removing outliers
- **Preparation of the training and testing dataset**: The training dataset should be representative of the significant features of the data
- **Dimensionality reduction**: identifying the critical variables, removing the unnecessary variables, combining multiple dependent variables
- **Identifying the appropriate mathematical model**: depends on the size and nature of the data
- **Training, testing and validation**: Training the model using the training set to maximize the predictive capability while avoiding overfitting, performance metrics (accuracy, precision, recall, F1-score).
- **Hyperparametric optimization** to simplify the model and increase its interpretability

Machine Learning vs Deep Learning



•DL outperforms ML on large datasets.

•Demands more time and hardware resources.

•DL eliminates the need for complex feature extraction.

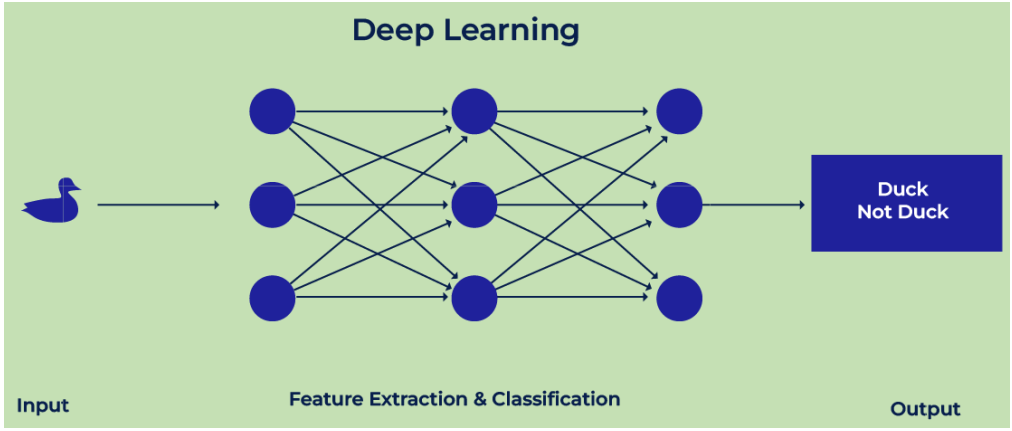
•Performs end-to-end learning directly from data (images, text, signals).

•Often viewed as a "blackbox" with limited interpretability.

•ML is superior to DL on small datasets, faster, and cheaper hardware requirements.

•Requires intricate feature engineering, incurring time and expertise costs.

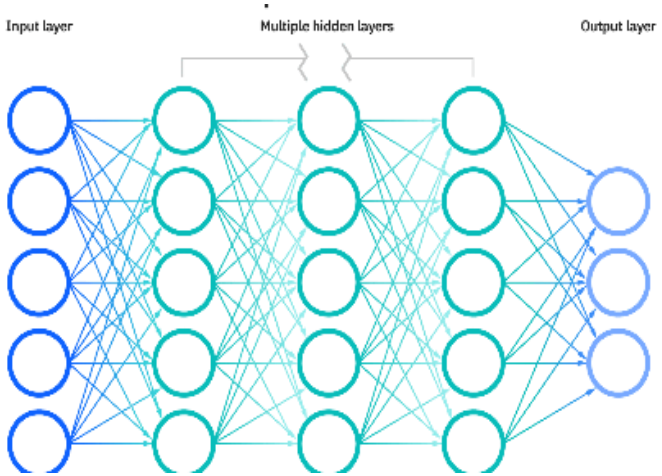
•Generally easier to interpret than DL models, providing insights into decision-making like "*How* and *why* the ML algorithm arrived at an outcome"



Deep Learning & its Types

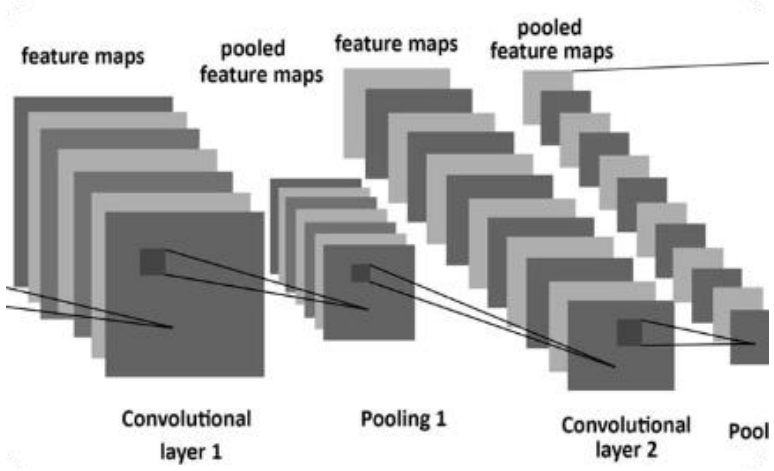
- ***Deep learning*** is a type of supervised machine learning that focuses on training neural networks to learn and make decisions from data.

Artificial/Deep Neural Networks



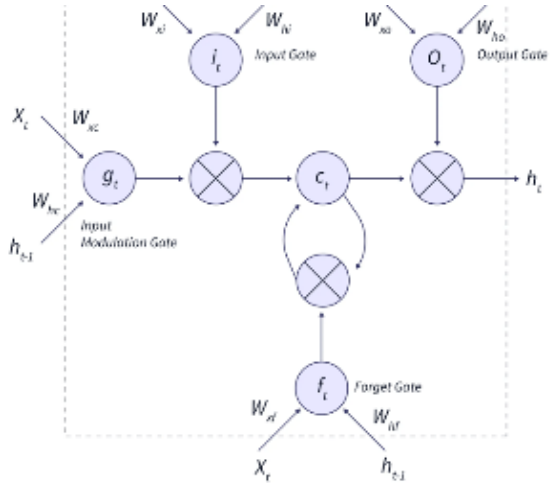
ANN: NN with many layers, capable of learning complex representations of data.

Convolutional Neural Networks



CNNs are a type of deep neural network used for image and video processing and are trained to extract visual features from input.

Recurrent Neural Networks

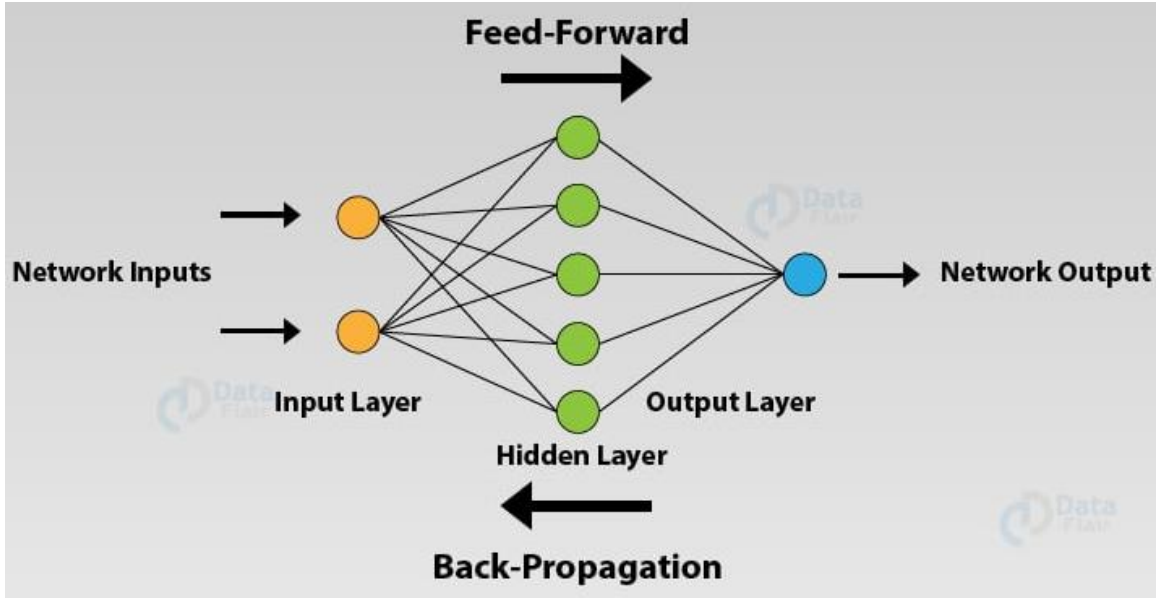


RNNs are used for sequential data analysis, such as natural language processing or time series prediction.

Artificial neural networks (ANN)

- Deep learning is usually implemented using an artificial ***neural network***.
- Artificial neural networks (ANN), comprise an input layer, multiple hidden layers, and an output layer. Each node in one layer is intricately connected to every node in the subsequent layer.
- The term “deep” refers to the ***number of layers*** in the network—the more layers, the deeper the network.

Artificial neural networks (ANN)



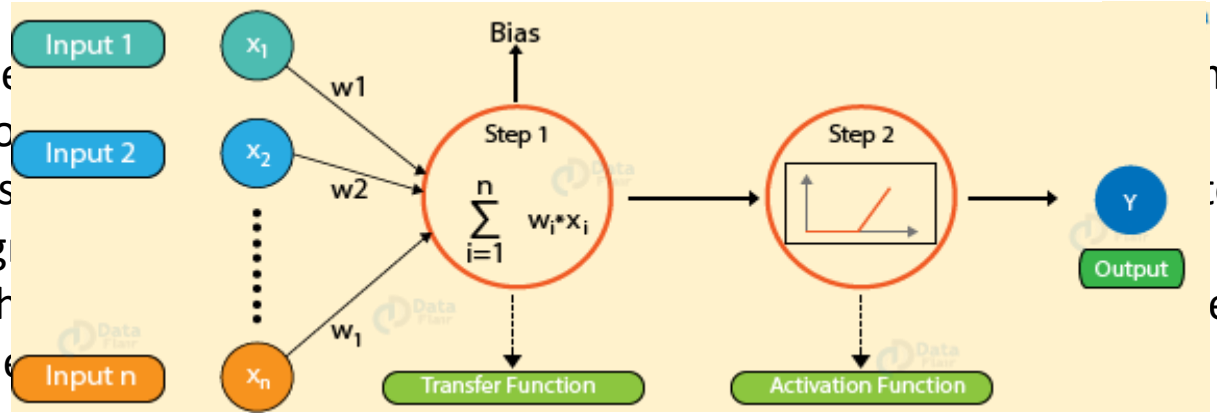
(data-flair.training)

ANN: Fundamental Processes

Two fundamental processes in training artificial neural networks.

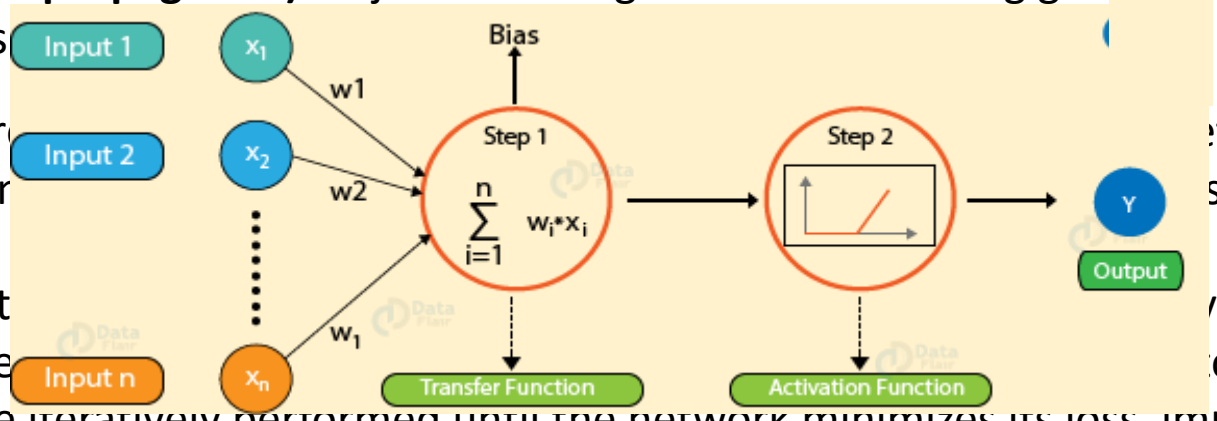
Forward Propagation: proce…nake predictions.

- *Input Layer*: Initial data o…
- *Hidden Layers*: Data pass…e weighted sums, apply activation functions (e.g., ReLU, sig…
- *Output Layer*: Data reach…ed on learned patterns and weights. Output from the…



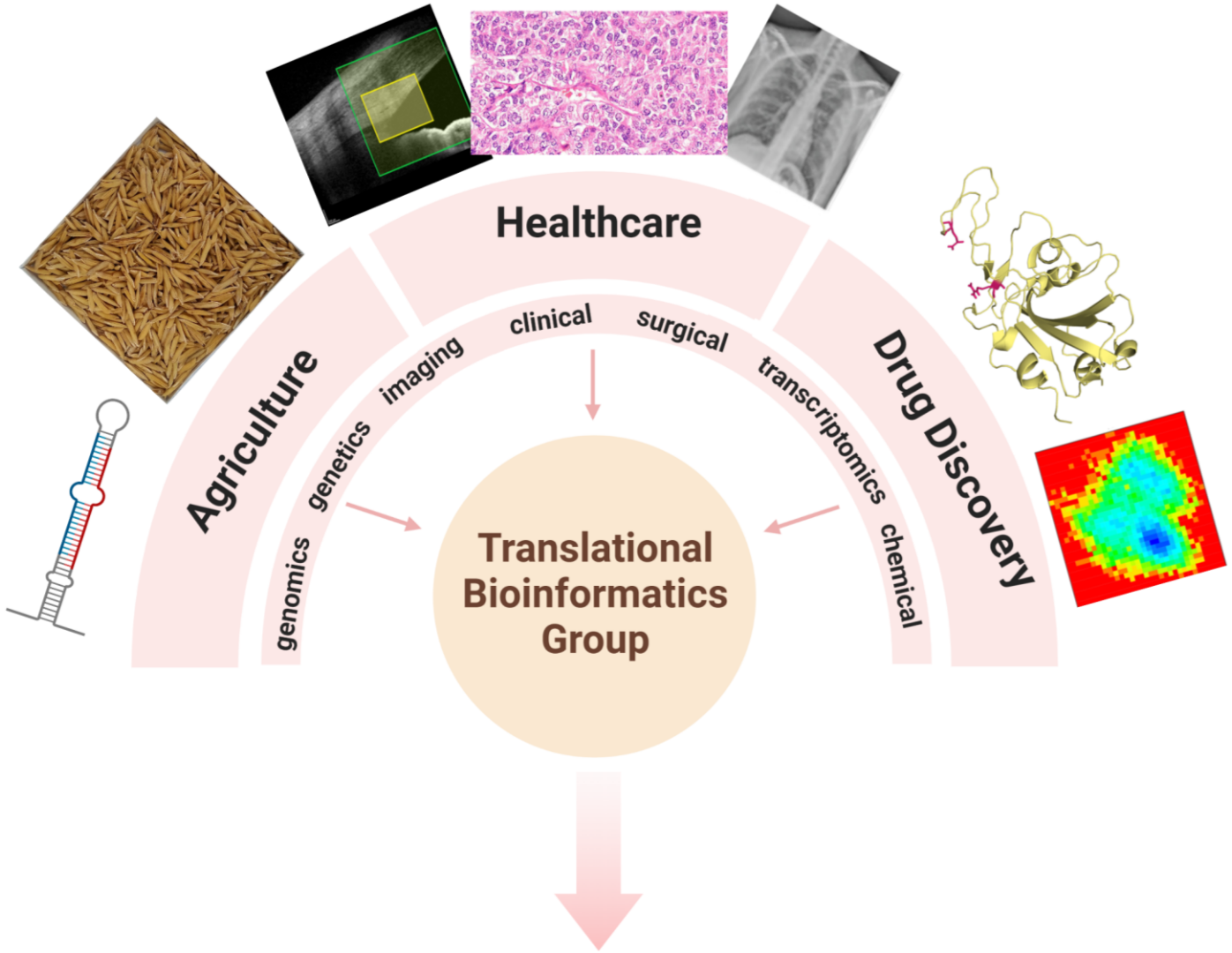
Backward Propagation (Backpropagation): adjusts NN weights and biases using gradients to minimize prediction errors. It's how the network learns…

- *Loss Calculation*: Measur…ts using a loss function.
- *Gradient Calculation*: Cor…s (weights and biases), indicating how to adjust them.
- *Backward Pass*: Propagat…vers, using the chain rule.
- *Weight Update*: Optimize…o reduce the loss.
- *Iteration*: These steps are iteratively performed until the network minimizes its loss, improving its performance.



(data-flair.training)

Use of NN in TBG lab



https://bioinfo.icgeb.res.in

Upcoming lectures

- Linear Classifiers
- Details of activation functions
- Common terms used in Deep Learning: epoch, learning rate, gradient descent, etc
- CNN
- Image classification

Enjoy Learning!