AI and ML in Healthcare and Lifesciences: The Jargons (Part 2)

To the curious friends of science,

This blog is the continuation of "**The Jargons**", used in Artificial Intelligence (AI) and Machine Learning (ML). Each term is explained in two ways — its actual meaning and a simpler or science-based analogy that connects with our world of experiments, data, and discovery.

From a fellow biotechy to new biotechies! Let's read!

21. Deep Learning

- A subset of ML using multi-layered neural networks to learn complex patterns in large data.
- Stacking building blocks — each layer adds more detail until you see the full picture.

22. Neural Networks

- A system of interconnected nodes (neurons) designed to simulate the way the human brain processes information.
- A relay race — each runner (neuron) passes the baton (information) to the next until the race (decision) is complete.

23. Epochs

- The number of complete passes a model makes through the entire training dataset.
- Revising the same textbook multiple times during examinations — each round helps the model remember better.

24. Loss Function

- A mathematical function that measures how much the model's predictions ***deviate*** from the actual values.
- A feedback report showing how many diagnostic errors were made — the model uses this to improve next time.

25. Evaluation Metrics

- Quantitative measures used to assess a model's performance.
- Just like how teachers use marks to see how well students learned, ML models use metrics to see how well they perform.

26. Accuracy

- The ratio of correct predictions made by the model to the total predictions.
- Similar to checking how many answers you got right in a test.

27. Precision

- Of all the positive results predicted, how many are actually correct.
- A doctor who says someone is sick only when they're really sure — and is usually right.

28. Sensitivity (Recall)

- Of all actual positive cases, how many the model correctly identified.
- A teacher who marks almost every wrong answer but sometimes marks a few right ones wrong too.

29. Specificity

- Of all the actual negatives, how many were correctly identified as negative.
- A teacher checking exams — specificity is how well they recognize students who didn't cheat (the true negatives) and don't wrongly accuse them.

30. F1 Score

- A balance between precision and recall; used when both false positives and false negatives matter.
- A cricket player (all-rounder) who balances batting and bowling — the F1 score values both precision (batting) and recall (bowling) equally.

31. Confusion Matrix

- A table that shows correct and incorrect predictions — with counts of true positives, false positives, true negatives, false negative.
- A scoreboard showing **hits, misses, false alarms,** and **correct rejections** in a game of identifying objects

32. ROC Curve (Receiver Operating Characteristic)

- A graph showing how well a model distinguishes between classes (like diseased vs healthy).
- Checking which student can best tell apart right and wrong answers — the sharper their judgment, the better their score on the curve.

33. AUC (Area Under Curve)

- A single number summarizing how good the ROC curve is; higher means better discrimination ability.
- Grading a student's performance — 1.0 means perfect answers every time, 0.5 means just guessing.

34. Cross-validation

- Splitting data into parts to train and test the model multiple times for more reliable performance.
- Testing a new drug in different patient groups to ensure it works consistently.

35. Feature Engineering

- The process of creating or selecting the right input variables to improve model performance.
- Choosing the right biomarkers that actually tell you something useful about disease progression.

36. Dimensionality Reduction

- Reducing the number of input features while keeping the most important information.
- Focusing on key lab parameters (WBC, RBC, Platelets) instead of analyzing every molecule in blood.

37. Principal Component Analysis (PCA)

- A statistical method to reduce data complexity by identifying main patterns (principal components).
- Summarizing a long research paper into a short abstract, keeping only the key points

38. Hyperparameter Tuning

- Optimizing the settings that control how a model learns (not learned from data itself).
- Adjusting PCR conditions — temperature, cycles to get the best amplification results.

39. Gradient Descent

- An optimization algorithm that adjusts model parameters to minimize the error (loss).
- Walking downhill step by step to reach the lowest point — the point of minimum error.

40. Batch Size

- Number of training samples the model processes before updating its parameters.
- Testing samples in small batches instead of the entire batch at once to improve precision and efficiency.