BP703T · B.Pharm Semester VII · Theory Notes

AI in Clinical Applications

Unit IV — AI in Clinical Decision Support and Real-World Data (6 hours)

Unit IV (6 hours)

Pills & Python Notes

From-zero teaching standard: every technical term is defined before use.

Syllabus coverage: Unit IV (6 hours).

PILLS & PYTHON

Pharmacy • Data • AI • Healthcare

Instagram @pillsandpython · pillsandpython.com · YouTube @PillsAndPython

CHAPTER 9: Electronic Health Records, Clinical Decision Support Systems, and AI for Outcome Prediction and Risk Scoring

Learning Objectives

After studying this chapter, you will be able to:

1 Describe what an Electronic Health Record (EHR) is and list its main parts.

2 Explain what a Clinical Decision Support System (CDSS) is and how it helps doctors and pharmacists.

3 Distinguish between outcome prediction (regression) and risk scoring (classification).

4 Work through a small hand-calculated example of linear regression for outcome prediction.

5 Work through a small hand-calculated example of classification for risk scoring.

6 Explain what real-world data (RWD) and real-world data analytics mean in healthcare.

7 Describe the pharmacy relevance of EHRs and CDSS in Indian hospitals.

Definitions

- **Electronic Health Record (EHR):** A digital version of a patient's complete medical history, stored and updated on a computer system in a hospital or clinic.
- **Clinical Decision Support System (CDSS):** A computer program that combines patient information with medical knowledge to give doctors, pharmacists, and nurses helpful advice — such as alerts, reminders, and suggested diagnoses or doses.
- **Outcome prediction:** Using data to guess a future number, such as how many days a patient will stay in the hospital. This uses regression models.
- **Risk scoring:** Using data to guess whether something will happen or not (yes/no), such as whether a patient is at high risk of readmission. This uses classification models.
- **Real-world data (RWD):** Health data collected during everyday medical care (not from planned research experiments) — for example, hospital records, insurance claims, and pharmacy bills.
- **Real-world data analytics:** Studying real-world data to find useful patterns, such as which medicines work best in daily practice.

Terminology Table

Term	Simple Meaning	Example
EHR	Digital file of a patient's full medical history	The computer file showing a patient's diagnoses, medicines, and lab reports
CDSS	Computer helper that gives treatment advice	A pop-up that warns "this dose is too high for this patient's kidney function"

Term	Simple Meaning	Example
Outcome prediction	Guessing a future number from data	Predicting the number of days a patient will stay in the ICU
Risk scoring	Guessing yes/no risk from data	Predicting high vs low risk of hospital readmission
Real-world data (RWD)	Everyday care data, not research data	Prescriptions written in a hospital OPD last year
Predictor variable (feature/input)	The information we already know, used for guessing	Age, weight, blood pressure of a patient
Outcome variable (target/label)	The answer we want to guess	Whether the patient developed an ADR (yes/no)
Regression	A math method that predicts numbers	Predicting dissolution rate (a number, e.g., 78%)
Classification	A math method that predicts categories	Predicting ADR risk as high or low (a category)
Probability	A number from 0 to 1 (or 0% to 100%) showing how likely an event is	0.80 means an 80% chance
Training data	The old data used to teach the model	Past patient records used to build the risk model
Testing data	New data used to check whether the model works	Patient records the model has never seen before
Accuracy	Fraction of all guesses that were correct	85 correct guesses out of 100 patients
Confusion matrix	A 2×2 table counting correct and wrong yes/no guesses	TP, FP, TN, FN counts
Alert fatigue	When doctors start ignoring computer warnings because there are too many	A doctor clicks "OK" on every pop-up without reading it
Pharmacovigilance	The science of detecting and preventing side effects of medicines	Reporting ADRs to the national centre

9.1 Structure of Electronic Health Records (EHR)

Simple Explanation

An EHR is like a digital file cabinet for every patient. Instead of paper files scattered across departments, one computer record holds everything: who the patient is, what diseases they have, which medicines they take, what their lab tests show, and what happened at each visit.

Example

Ramesh, a 58-year-old patient with diabetes, visits a hospital. His EHR contains: name, age, and phone number (demographics); his diabetes and blood pressure diagnosis; his prescription of metformin 500 mg twice daily; his latest blood sugar report (fasting 142 mg/dL); his doctor's note from the last visit; and his allergy history (penicillin allergy).

Diagram: Main Parts of an EHR

```
+-------------------------------------------------------+
|              ELECTRONIC HEALTH RECORD (EHR)           |
+-------------------------------------------------------+
| 1. Demographics      : name, age, sex, address        |
| 2. Diagnosis list    : diabetes, hypertension         |
| 3. Medication list   : metformin 500 mg BD            |
| 4. Laboratory data   : blood sugar, kidney tests      |
| 5. Clinical notes    : doctor's visit notes           |
| 6. Allergy history   : penicillin allergy             |
| 7. Imaging           : X-ray, CT scan reports         |
| 8. Vital signs       : BP, pulse, temperature         |
+-------------------------------------------------------+
```

Key Takeaway

An EHR is a single digital record holding patient demographics, diagnoses, medicines, labs, notes, allergies, images, and vitals — the raw material on which AI models are built.

9.2 AI in Clinical Decision Support Systems (CDSS)

Simple Explanation

A CDSS is a smart assistant built on top of the EHR. It reads the patient's record, compares it with medical rules or an AI model, and then suggests or warns: "consider checking kidney function before giving this dose," or "this patient has a high risk of readmission."

Example

A doctor prescribes a high dose of an antibiotic to an elderly patient. The CDSS reads the patient's age and kidney test from the EHR, runs a rule, and shows an alert: "Reduce the dose — kidney function is low." The pharmacist sees the same alert and can double-check before dispensing.

Classification: Types of CDSS

Type	What It Does	Example
Knowledge-based CDSS	Uses fixed medical rules written by experts	"If patient is allergic to penicillin, block the prescription"
Non-knowledge-based (AI) CDSS	Learns patterns from data	A model that learned which patients are likely to be readmitted
Drug-interaction checker	Checks medicine combinations	Warns when two medicines that should not be combined are prescribed together
Dose advisor	Suggests correct doses	Suggests a kidney-adjusted dose of a medicine
Diagnostic assistant	Suggests possible diagnoses	Lists likely causes of a set of symptoms

Flowchart: How a CDSS Works

```
Patient data (EHR)
       |
       v
+------------------+
|  CDSS engine     |<--- Medical knowledge / AI model
|  (rules or AI)   |
+------------------+
       |
       v
Advice shown to doctor/pharmacist:
 - Alert (warning)
 - Reminder (do the test)
 - Suggestion (recommended dose)
       |
       v
Clinician makes the final decision
```

Key Takeaway

A CDSS combines EHR data with rules or AI models to give alerts, reminders, and suggestions — but the clinician always makes the final decision.

9.3 Regression Models for Outcome Prediction

Simple Explanation

Regression answers "how much?" or "how many?" — a number question. A linear regression model draws the best straight line through old data points, and then uses that line to predict new numbers.

The line has a simple formula:

```
y = a + b × x
```

- **y** = the number we want to predict (outcome)
- **x** = the information we know (predictor)
- **b** = slope — how much y changes when x increases by 1
- **a** = intercept — the value of y when x is 0

Example

A hospital wants to predict how many days a patient will stay in hospital (y) based on the patient's age (x).

INFOGRAPHIC CONCEPT: "Drawing the best-fit line" — Imagine plotting patients' ages on the horizontal axis and hospital stay days on the vertical axis. Each patient is a dot. Linear regression draws one straight line that passes as close as possible to all the dots. To predict a new patient's stay, go up from their age to the line, then across to the number of days. The slope tells you how many extra days each extra year of age adds.

WORKED EXAMPLE 9.1 — Linear Regression for Hospital Stay Prediction (illustrative toy dataset)

Problem: Predict hospital stay (days) from patient age using 5 patients.

Illustrative toy dataset (invented for learning only; not real hospital data):

Patient	Age x (years)	Stay y (days)
A	30	4
B	40	5
C	50	7
D	60	8
E	70	10

Step 1 — Find the means.

- Mean of x = (30 + 40 + 50 + 60 + 70) / 5 = 250 / 5 = 50
- Mean of y = (4 + 5 + 7 + 8 + 10) / 5 = 34 / 5 = 6.8

Step 2 — Build the working table.

x	y	x − mean(x)	y − mean(y)	(x − mean)²	(x − mean)(y − mean)
30	4	−20	−2.8	400	56
40	5	−10	−1.8	100	18
50	7	0	0.2	0	0
60	8	10	1.2	100	12
70	10	20	3.2	400	64
Total				**1000**	**150**

Step 3 — Calculate slope b and intercept a.

- b = (total of last column) / (total of squared column) = 150 / 1000 = 0.15
- a = mean(y) − b × mean(x) = 6.8 − 0.15 × 50 = 6.8 − 7.5 = −0.7

Model: predicted stay = −0.7 + 0.15 × age

Step 4 — Make a prediction. For a 55-year-old patient: predicted stay = −0.7 + 0.15 × 55 = −0.7 + 8.25 = 7.55 ≈ 8 days.

Step 5 — Check the model fit (RMSE and R^2).

Predictions: for ages 30, 40, 50, 60, 70 → 3.8, 5.3, 6.8, 8.3, 9.8 days.

Actual y	Predicted	Error (actual − predicted)	Error²
4	3.8	0.2	0.04
5	5.3	−0.3	0.09
7	6.8	0.2	0.04
8	8.3	−0.3	0.09
10	9.8	0.2	0.04
		Total	**0.30**

- MSE = 0.30 / 5 = 0.06 → RMSE = √0.06 ≈ 0.24 days (average prediction error is only about a quarter of a day).
- $R^2 = (150)^2 / (1000 \times 22.8) = 22500 / 22800 \approx 0.99$. (Here 22.8 is the total of squared y-deviations.) R^2 = 0.99 means the model explains 99% of the variation in stay length — an excellent fit on this toy data.

Interpretation: Each extra year of age adds about 0.15 days to the predicted stay. The model fits the toy data very well ($R^2 \approx 0.99$, RMSE ≈ 0.24 days). On real data the fit would be much less perfect.

Key Takeaway

Linear regression finds the best-fit line y = a + b×x, makes number predictions, and is judged with RMSE (lower is better) and R^2 (closer to 1 is better).

9.4 Classification Models for Risk Scoring

Simple Explanation

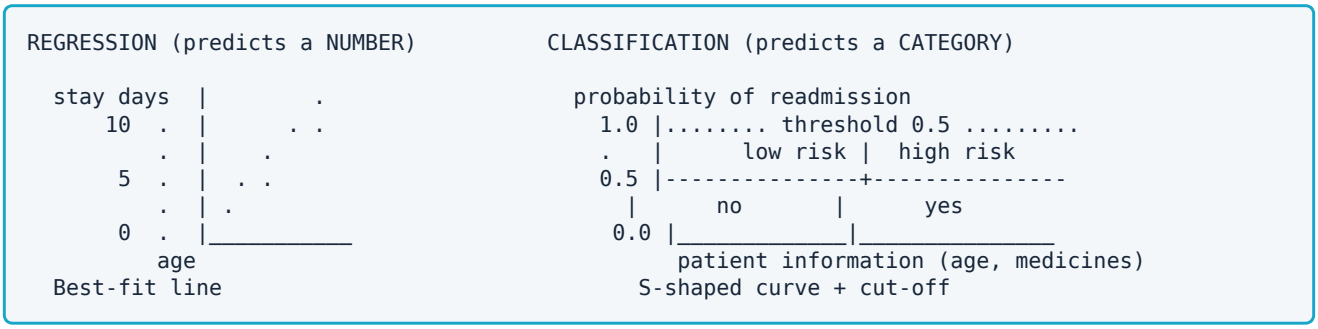
Classification answers "which group?" — a category question. In risk scoring, the model looks at a patient's information and places them into a risk group, for example "high risk of readmission" or "low risk of readmission."

Logistic regression is the most common classification model in medicine. Instead of drawing a straight line for numbers, it converts the information into a **probability** (0 to 1) of belonging to the "yes" group, and then applies a cut-off (usually 0.5): probability ≥ 0.5 → "yes/high risk"; probability < 0.5 → "no/low risk."

Example

A pharmacist wants a model that predicts whether a discharged patient will be readmitted within 30 days (yes/no) using age and number of medicines. The model outputs a probability like 0.72 (72% chance) → classified as "high risk," so the pharmacist plans a follow-up call.

Diagram: Regression vs Classification



INFOGRAPHIC CONCEPT: "From risk factors to a risk score" — Picture a funnel: at the top, patient facts pour in (age, number of medicines, previous admissions, lab values). In the middle, the model weighs each fact — some push risk up, some push it down. At the bottom, a single number comes out: the risk score, e.g., "72% chance of readmission = HIGH RISK." Below the funnel, a small confusion matrix (2×2 table) shows how the model did: correct yes-guesses (TP), correct no-guesses (TN), and the two kinds of mistakes (FP, FN).

WORKED EXAMPLE 9.2 — Classification for Readmission Risk (illustrative toy dataset)

Problem: Classify 8 discharged patients as high risk (will be readmitted) or low risk using one predictor: number of medicines they take.

Illustrative toy dataset (invented for learning only; not real patient data). Suppose a trained logistic model gives these readmission probabilities:

Patient	Medicines	True outcome	Model probability	Predicted (≥0.5 = high)
1	2	No	0.20	Low
2	3	No	0.35	Low
3	5	Yes	0.60	High
4	6	Yes	0.75	High
5	2	No	0.30	Low
6	7	Yes	0.85	High
7	4	No	0.55	High
8	5	Yes	0.45	Low

Step 1 — Build the confusion matrix (compare predicted vs true).

	True: Yes	True: No
Predicted: High	TP = 3 (patients 3, 4, 6)	FP = 1 (patient 7)
Predicted: Low	FN = 1 (patient 8)	TN = 3 (patients 1, 2, 5)

Step 2 — Calculate metrics.

- Accuracy = (TP + TN) / total = (3 + 3) / 8 = 6/8 = 0.75 (75%)
- Sensitivity (recall) = TP / (TP + FN) = 3 / 4 = 0.75 (75%) — of the patients truly readmitted, how many did we catch?
- Specificity = TN / (TN + FP) = 3 / 4 = 0.75 (75%) — of the patients truly not readmitted, how many did we correctly call low risk?
- Precision = TP / (TP + FP) = 3 / 4 = 0.75 (75%) — of the patients we flagged high risk, how many were truly readmitted?

Interpretation: The toy model correctly classifies 75% of patients. Patient 8 is a false negative (missed high-risk patient — the dangerous error), and patient 7 is a false positive (unnecessary follow-up — the costly but safer error). In real clinical use, missing a high-risk patient is usually worse than an extra follow-up call.

Key Takeaway

Classification puts patients into risk groups (high/low); the confusion matrix counts the two correct outcomes and the two errors, and sensitivity, specificity, precision, and accuracy summarize performance.

9.5 Real-World Data Analytics

Simple Explanation

Clinical trials are like examinations held in a quiet hall with strict rules. Real-world data is like observing students in their daily classroom — messier, but it shows what really happens. Real-world data analytics

studies this everyday data to answer questions trials cannot: "Does this medicine work as well in elderly Indian patients as it did in the trial?"

Sources of Real-World Data

Source	What It Contains	Pharmacy Example
EHRs	Diagnoses, prescriptions, labs	Finding which antibiotic is most prescribed for UTI
Insurance claims	Bills for treatments and medicines	Comparing costs of two diabetes regimens
Pharmacy dispensing records	What was actually dispensed	Measuring medication adherence (refill gaps)
Disease registries	Long-term records of one disease	Tracking TB treatment outcomes
Pharmacovigilance databases	ADR reports	Detecting a new side-effect signal

Flowchart: Real-World Data Analytics Pipeline

```
Collect RWD (EHR, claims, pharmacy records)
       |
       v
Clean the data (remove duplicates, fix errors)
       |
       v
Analyze (regression, classification, trends)
       |
       v
Interpret in clinical context
       |
       v
Decision: change a prescription policy? issue a safety alert?
```

Key Takeaway

Real-world data analytics turns messy everyday healthcare records into evidence about how medicines actually perform outside clinical trials.

9.6 Pharmacy Relevance

- **Dispensing safety:** A CDSS checks every prescription against the EHR — drug interactions, allergies, duplicate therapy, and dose limits — before the medicine reaches the patient.
- **Dose individualization:** Regression models can suggest doses adjusted for age, weight, and kidney function, helping the pharmacist counsel correctly.
- **Risk-based follow-up:** Classification models flag high-risk patients (e.g., likely readmission or ADR) so the pharmacist can prioritize counselling and follow-up calls.
- **Medication adherence:** Real-world dispensing data shows which patients skip refills; analytics helps target adherence programs.
- **Inventory and quality:** Outcome prediction models can forecast medicine demand, reducing stock-outs of essential drugs.

9.7 India Context

- India's Ayushman Bharat Digital Mission (ABDM) is building a national digital health ecosystem where health records can be linked across hospitals, creating the foundation for EHR-based analytics.
- Large Indian hospitals already run CDSS-style drug-interaction and allergy alerts in their pharmacy software; the national pharmacovigilance programme collects ADR reports that feed real-world safety analytics.
- Challenges in India include mixed paper-digital records in smaller hospitals, data in multiple languages, and uneven internet connectivity — all of which make real-world data messier and preprocessing more important.

Key Points

1 An EHR stores patient demographics, diagnoses, medicines, labs, notes, allergies, images, and vitals in one digital record.

2 A CDSS reads EHR data and gives alerts, reminders, and suggestions; the clinician makes the final decision.

3 CDSS types: knowledge-based (fixed rules), AI-based (learned patterns), drug-interaction checkers, dose advisors, diagnostic assistants.

4 Regression predicts numbers ($y = a + b \times x$); judged by RMSE (lower is better) and R^2 (closer to 1 is better).

5 Classification predicts categories (high/low risk); judged with a confusion matrix, accuracy, sensitivity, specificity, and precision.

6 False negatives (missing a truly at-risk patient) are usually the most dangerous error in clinical risk scoring.

7 Real-world data comes from everyday care (EHRs, claims, pharmacy records, registries, ADR databases) and shows how medicines perform outside trials.

8 Pharmacists use CDSS for dispensing safety and use prediction models to prioritize counselling and follow-up.

Common Confusions

Confusion	Clarification
"EHR and CDSS are the same thing"	No. The EHR is the data store; the CDSS is the advisory program that reads the EHR and gives advice.
"Regression is for yes/no questions"	No. Regression predicts numbers. Classification predicts categories like yes/no.
"A probability of 0.8 means the event will definitely happen"	No. It means an 80% chance — 2 out of 10 such patients would still not have the event.
"Real-world data is better than trial data"	Not better — different. Trials are clean but narrow; RWD is broad but messy. Both are needed.
"The CDSS decides the treatment"	No. The CDSS advises; the doctor or pharmacist decides and is responsible.

Common Mistakes

1 Mixing up regression and classification — remember: regression = "how much?", classification = "which group?".

2 Judging a classification model by accuracy alone when the classes are unbalanced (e.g., 95% "no ADR") — sensitivity matters more.

3 Thinking a high R^2 on training data guarantees good real-world performance (it may be overfitting).

4 Forgetting that a CDSS alert must be read — alert fatigue causes real errors.

5 Using a risk score without checking whether the model was trained on patients like yours.

Mnemonics

- **EHR parts — "My Daily Medicines, Labs, Notes, Allergies, Images, Vitals"** → Demographics, Diagnosis, Medicines, Labs, Notes, Allergies, Images, Vitals.
- **CDSS advice — "ARS"**: Alerts, Reminders, Suggestions.
- **Regression vs Classification — "Numbers Need Regression, Groups Get Classification"** (NNR-GGC).
- **Confusion matrix metrics — "All Sick People Should Take medicine"**: Accuracy (all), Sensitivity (sick caught), Precision (people flagged correctly), Specificity (healthy correctly cleared).

Exam Summary

Chapter 9 covers the data foundation of clinical AI. EHRs store the complete digital patient record (demographics, diagnoses, medicines, labs, notes, allergies, images, vitals). CDSS tools read EHR data and provide alerts, reminders, and suggestions using fixed rules or AI models, while clinicians retain final authority. Regression models ($y = a + b \times x$) predict numerical outcomes such as length of stay and are evaluated with RMSE and R^2; classification models (e.g., logistic regression with a 0.5 cut-off) assign patients to risk groups and are evaluated with the confusion matrix, accuracy, sensitivity, specificity, and precision. Real-world data from everyday care (EHRs, claims, pharmacy records, registries, ADR databases) is analyzed to learn how medicines perform outside trials. For pharmacists, these tools support dispensing safety, dose individualization, risk-based follow-up, and adherence monitoring; in India, ABDM is building the digital backbone for such analytics.

One-Minute Revision

EHR = digital patient file; CDSS = computer advisor (alerts, reminders, suggestions); regression predicts numbers (RMSE, R^2); classification predicts groups (confusion matrix: sensitivity, specificity, precision, accuracy); missing a high-risk patient (false negative) is the worst error; real-world data = everyday care data used to study true medicine performance; pharmacist uses CDSS for safe dispensing and risk models for follow-up.

Short-Answer Questions

1 List any four components of an Electronic Health Record.

2 What is the difference between a knowledge-based and an AI-based CDSS?

3 Define outcome prediction and give one pharmacy example.

4 Define risk scoring and give one clinical example.

5 Write the formula of a simple linear regression line and name each term.

6 What do RMSE and R^2 tell us about a regression model?

7 Draw a 2×2 confusion matrix and label TP, FP, TN, FN.

8 Define sensitivity and specificity in your own words.

9 Name three sources of real-world data.

10 Why is a false negative more dangerous than a false positive in readmission risk scoring?

Long-Answer Questions

1 Describe the structure of an Electronic Health Record with a labelled diagram. Explain how each component supports AI-based clinical decision support.

2 Explain Clinical Decision Support Systems: their types, working with a flowchart, and their role in safe dispensing by pharmacists.

3 Using the worked toy dataset in this chapter, explain step by step how a linear regression model is built and evaluated (slope, intercept, RMSE, R^2), and interpret each result.

4 Explain classification for risk scoring with a confusion matrix. Calculate and interpret accuracy, sensitivity, specificity, and precision for a given toy dataset.

5 Discuss real-world data analytics: its sources, the analysis pipeline, and how it differs from clinical trial data. Give pharmacy examples.

MCQs

1 Which of the following is NOT a component of an EHR?

a) Allergy history b) Laboratory reports c) Medicine advertising brochure d) Clinical notes

1 A CDSS alert warning about a drug interaction is an example of:

a) A knowledge-based CDSS function b) A laboratory test c) Real-world data collection d) A regression model

1 Predicting the number of ICU stay days from patient age is a:

a) Classification problem b) Regression problem c) Clustering problem d) Data cleaning problem

1 In a confusion matrix, a false negative means:

a) A healthy patient correctly called low risk b) An at-risk patient wrongly called low risk c) A healthy patient wrongly called high risk d) An at-risk patient correctly called high risk

1 Sensitivity of a risk model is calculated as:

a) TP / (TP + FP) b) TN / (TN + FP) c) TP / (TP + FN) d) (TP + TN) / total

Answer Key: 1-c, 2-a, 3-b, 4-b, 5-c

Viva Questions

1 What does EHR stand for, and what is stored in it?

2 Who makes the final treatment decision — the CDSS or the clinician?

3 What is the difference between regression and classification?

4 What does an R^2 of 0.99 tell you?

5 Which is worse in risk scoring — a false positive or a false negative? Why?

6 Name two sources of real-world data.

7 What is alert fatigue?

Application-Based Questions

1 A hospital pharmacist sees a CDSS alert: "Patient has penicillin allergy — prescribed amoxicillin." What should the pharmacist do, step by step?

2 You are given a toy dataset of 6 patients with their weights and drug clearance values. Outline how you would build a regression model to predict clearance for a new patient.

3 A readmission-risk model flags 20 patients as high risk, but only 12 are truly readmitted. Calculate precision and explain what it means for the follow-up team.

References

- Aggarwal & Reddy, Healthcare Data Analytics
- Bonate, Pharmacokinetic–Pharmacodynamic Modeling and Simulation
- Campbell, Machin & Walters, Medical Statistics: A Textbook for the Health Sciences
- Schmidt & Derendorf, Applied Pharmacometrics
- Steyerberg, Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating
- Strom, Kimmel & Hennessy, Pharmacoepidemiology

CHAPTER 10: Limitations of AI in Clinical Environments, Accountability, and Interpretability

Learning Objectives

After studying this chapter, you will be able to:

1 List and explain the main limitations of AI models in clinical practice.

2 Define bias in AI and give healthcare examples with a small toy illustration.

3 Define overfitting and underfitting in simple words.

4 Explain what interpretability means and why doctors need it.

5 Distinguish between interpretable models and black-box models.

6 Explain accountability in AI-assisted healthcare — who is responsible when something goes wrong.

7 Describe practical safeguards pharmacists can apply before trusting an AI recommendation.

Definitions

- **Limitation:** A weakness or boundary of a method — something the method cannot do reliably.
- **Bias (in AI):** A systematic error in a model that makes it unfair or wrong for some groups of patients — for example, performing well for one age group but poorly for another.
- **Overfitting:** When a model memorizes the training data (including its noise) instead of learning the real pattern, so it fails on new patients.
- **Underfitting:** When a model is too simple to capture the real pattern, so it performs poorly even on the training data.
- **Generalization:** The ability of a model to work well on new, unseen patients — the true test of a clinical model.
- **Interpretability:** How easily a human can understand *why* a model made a particular prediction.
- **Black-box model:** A model whose internal reasoning is too complex for a human to follow (e.g., a large neural network).
- **Accountability:** Clear responsibility for decisions — knowing who answers for the outcome when AI assists a clinical decision.
- **Confounding:** When a hidden factor distorts the apparent relationship between predictor and outcome.
- **Data drift:** When the real-world data slowly changes over time so the model becomes outdated.

Terminology Table

Term	Simple Meaning	Example
Limitation	What the model cannot do well	A model trained on adults may fail for children
Bias	Systematic unfair error	A risk model underestimates risk in women because the training data had few women
Overfitting	Memorizing instead of learning	A model that "knows" training patients by heart but fails on new ones

Term	Simple Meaning	Example
Underfitting	Too simple to learn the pattern	Using only age to predict a complex disease
Generalization	Working on new patients	A model tested successfully in a different hospital
Interpretability	Understanding the "why" behind a prediction	"The model flagged high risk because of age 75 + 8 medicines"
Black box	Cannot see the reasoning inside	A deep neural network with millions of parameters
Accountability	Who is answerable for the decision	The prescribing doctor remains responsible, not the software
Confounding	A hidden factor confusing the result	Patients on drug X seem sicker because drug X is given to sicker patients
Alert fatigue	Ignoring warnings due to too many of them	Clicking past a dose alert without reading it
Validation	Testing the model on independent data	Checking the model in a second hospital before using it
Human-in-the-loop	A human checks every AI recommendation	A pharmacist reviews each AI-suggested dose

10.1 Why AI Models Fail in Clinical Environments — The Limitations

Simple Explanation

AI models are students who only know what they were taught from old books. If the books were incomplete, old, or one-sided, the student gives wrong answers on the exam of real life. Clinical environments are messy: missing records, different patient populations, changing diseases, and tired staff. Each of these limits what AI can safely do.

Example

A readmission-risk model trained in a big city hospital is installed in a rural clinic. It performs badly because rural patients differ in age mix, diseases, and record quality. The model learned the city's pattern, not a universal truth.

Classification: Seven Key Limitations

#	Limitation	Simple Meaning	Example
1	Poor data quality	Missing, wrong, or incomplete records	Half the allergy fields are blank, so the model misses allergies
2	Biased training data	Training data does not represent all patients	Few elderly patients in training → poor predictions for the elderly
3	Overfitting	Memorized training noise	99% accuracy on old data, 60% on new patients
4	Data drift	The world changes after training	A model trained before a new treatment guideline becomes outdated

#	Limitation	Simple Meaning	Example
5	Lack of context	The model misses what a doctor sees	The model cannot see that the patient looks very unwell today
6	Alert fatigue	Too many warnings → ignored warnings	Staff click past a critical interaction alert
7	False confidence	Users trust the model too much	A pharmacist skips double-checking because "the computer said so"

Flowchart: From Data Problem to Clinical Harm

```
Biased or poor-quality training data
       |
       v
Model learns the wrong pattern (bias / overfitting)
       |
       v
Wrong prediction for a real patient
       |
       v
Clinician over-trusts or ignores the alert
       |
       v
Wrong dose / missed interaction / missed high-risk patient
       |
       v
Patient harm  -->  Who is accountable? (see 10.4)
```

Key Takeaway

AI limitations start with data (quality, bias, drift), continue through modeling (overfitting, no context), and end with human factors (alert fatigue, over-trust).

10.2 Bias and Confounding — Worked Toy Illustration

Simple Explanation

Bias means the model systematically favours or harms one group. Confounding means a hidden factor creates a fake relationship. Both come from the data, not from the math.

Example

A hospital builds an ADR-prediction model. The training records come mostly from one ward where young patients are common. The model works well for young patients but misses ADRs in elderly patients — this is bias. Separately, the data shows that patients taking vitamin supplements have fewer ADRs — but the hidden truth is that healthier patients take vitamins; the vitamin is not protective — this is confounding.

INFOGRAPHIC CONCEPT: "Where bias hides" — Picture three funnels feeding one model: Funnel 1 (who is in the data) shows 90 young patients and 10 elderly patients — sampling bias. Funnel 2 (how the data was recorded) shows missing allergy fields in handwritten notes — measurement bias. Funnel 3 (what changed since) shows last year's prescribing pattern vs this year's — drift. The model at the bottom inherits all three funnels' problems, and its single output number cannot warn you about them.

WORKED EXAMPLE 10.1 — Spotting Bias in a Toy Dataset (illustrative toy dataset)

Problem: Check whether a toy ADR-risk model performs equally for younger and older patients.

Illustrative toy dataset (invented for learning only; not real patient data). The model predicts "ADR likely" or "ADR unlikely" for 20 patients:

Younger patients (10):

Patient	True ADR	Predicted
Y1–Y7	Yes (7 patients)	ADR likely (all 7 correct)
Y8–Y10	No (3 patients)	ADR unlikely (all 3 correct)

Older patients (10):

Patient	True ADR	Predicted
O1–O5	Yes (5 patients)	ADR likely (3 correct, 2 missed)
O6–O10	No (5 patients)	ADR unlikely (2 correct, 3 wrongly flagged)

Step 1 — Accuracy for younger patients: (7 + 3) / 10 = 1.00 (100%).

Step 2 — Accuracy for older patients: (3 + 2) / 10 = 0.50 (50%).

Step 3 — Sensitivity for older patients: 3 / 5 = 0.60 (60%) — the model misses 40% of true ADRs in the elderly.

Interpretation: The model looks excellent overall (15/20 = 75% accuracy) but hides a serious bias: it fails half the time for older patients, the group most vulnerable to ADRs. Overall accuracy masked the problem; checking performance group-by-group revealed it.

Key Takeaway

Always check model performance separately for important patient groups (age, sex) — a good overall number can hide dangerous bias against one group.

10.3 Overfitting and Underfitting

Simple Explanation

Imagine teaching a student with 5 practice questions. **Overfitting** is the student memorizing the exact answers without understanding — they score 100% on practice but fail the real exam. **Underfitting** is the student barely studying — they fail both practice and the real exam. A good model is like a student who understands the concept: good on practice and good on the exam (new data).

Diagram

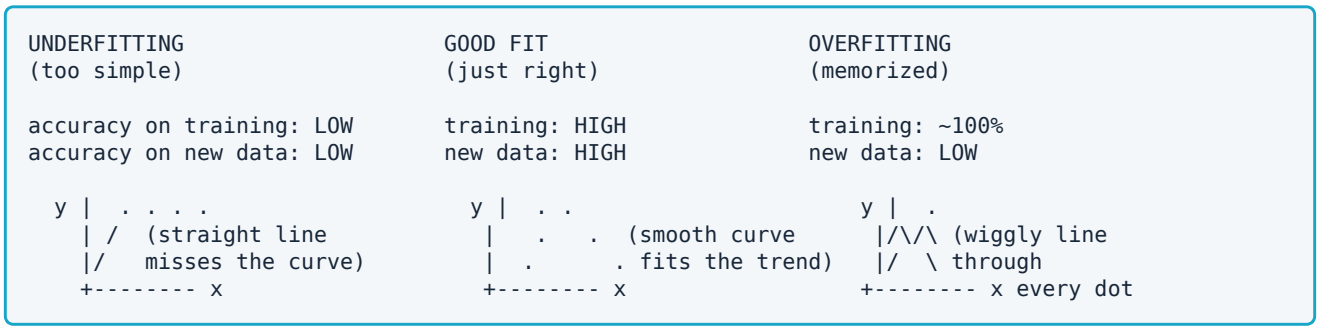


Table: Detecting the Three Situations

Situation	Training performance	New-data performance	Remedy (concept)
Underfitting	Poor	Poor	Use a richer model / add relevant predictors
Good fit	Good	Good	Keep and monitor
Overfitting	Excellent	Poor	Simplify the model; test on separate data

Key Takeaway

Overfitting = memorizes training data, fails on new patients. Underfitting = too simple, fails everywhere. Only separate testing on new data reveals the difference.

10.4 Interpretability — Understanding the "Why"

Simple Explanation

Interpretability means the model can explain itself in human words. A doctor will not act on "the computer says high risk" unless they also hear "because the patient is 78, takes 9 medicines, and was admitted twice this year." Interpretability builds the trust needed for clinical action.

Example

Two models predict readmission risk:

- **Model A (interpretable — logistic regression):** "Risk = 72% because age contributes +15%, number of medicines +20%, previous admissions +25%, baseline 12%." The doctor understands and acts.
- **Model B (black box — deep neural network):** "Risk = 72%." The doctor cannot verify the reasoning and hesitates.

Classification: Interpretable vs Black-Box Models

Feature	Interpretable Models	Black-Box Models
Examples	Linear regression, logistic regression, small decision trees	Deep neural networks, large ensembles
Can it explain a single prediction?	Yes — shows each factor's contribution	Usually no — too complex
Accuracy on complex data	Moderate	Often higher
Trust in clinic	Easier to earn	Harder to earn

Feature	Interpretable Models	Black-Box Models
When preferred	Dosing, risk scoring, regulated decisions	Image analysis where explanation is less critical

Flowchart: Interpretability in a Clinical Decision

```
Model predicts "HIGH RISK (78%)"
       |
       v
Can the model explain WHY?
      / \
    Yes   No (black box)
     |       |
     v       v
Doctor checks     Doctor must decide whether
the reasons and    to trust an unexplained number
agrees/disagrees   --> higher risk of error or rejection
     |
     v
Informed clinical decision
```

Key Takeaway

Interpretable models show their reasons (each factor's contribution); black-box models give answers without reasons. In patient care, an explainable reason is often more valuable than a slightly more accurate mystery.

10.5 Accountability — Who Is Responsible?

Simple Explanation

Accountability answers: "If the AI-assisted decision harms a patient, who answers for it?" The clear rule in healthcare is: **the licensed clinician who acts on the advice remains responsible.** The software is a tool, like a calculator — if you use a calculator wrongly, the mistake is yours.

Example

A CDSS suggests a dose, the pharmacist dispenses it without checking, and the patient is harmed. Accountability rests with the pharmacist and the prescribing doctor, not the software vendor — because a trained professional is expected to verify.

Table: Accountability of Each Party

Party	Responsibility
Doctor / pharmacist (user)	Verify every AI recommendation against clinical judgment; final decision-maker
Hospital / institution	Choose validated models, train staff, monitor performance, keep human-in-the-loop
Model developer / vendor	Provide honest documentation of limitations, testing, and intended use
Regulator / ethics committee	Set rules for safe deployment and audit of clinical AI

Practical Safeguards for the Pharmacist (Checklist)

1. **Verify, don't trust blindly** — recheck dose, interactions, and allergies yourself.
2. **Know the model's limits** — ask: which patients was it tested on? What does it not cover?
3. **Keep human-in-the-loop** — never let software auto-dispense or auto-prescribe without review.
4. **Watch for alert fatigue** — read every alert; report useless alerts so they can be fixed.
5. **Document your decision** — record why you accepted or overrode the AI advice.
6. **Report failures** — a wrong AI suggestion is a safety incident; report it like any error.

Key Takeaway

The clinician is accountable for AI-assisted decisions. Safety comes from verification, knowing the model's limits, human-in-the-loop review, and documenting decisions.

10.6 Pharmacy Relevance

- **Dispensing:** A pharmacist who understands overfitting will not trust a vendor's "99% accurate" claim without asking how it was tested.
- **Counselling:** Interpretability lets the pharmacist explain to the patient *why* they are flagged high-risk, in plain language.
- **Pharmacovigilance:** Recognizing bias helps the pharmacist question ADR models that underperform in elderly or female patients.
- **Hospital practice:** The accountability rule protects patients only if pharmacists keep human-in-the-loop and document overrides.

10.7 India Context

- Many Indian hospitals use software with built-in interaction checkers; understanding their limitations helps pharmacists avoid both blind trust and alert fatigue.
- Data bias is a real concern: models trained on Western populations may not generalize to Indian patients with different genetics, diets, and disease patterns — local validation matters.
- India's pharmacovigilance programme and hospital ethics committees provide the natural home for reporting AI-related safety concerns and auditing model use.

Key Points

1. AI limitations: poor data quality, biased training data, overfitting, data drift, missing clinical context, alert fatigue, and false confidence.
2. Bias = systematic unfair error for a patient group; confounding = a hidden factor creating a fake relationship.
3. A good overall accuracy can hide poor performance in one group — always check group-wise performance.

4 Overfitting memorizes training data and fails on new patients; underfitting is too simple and fails everywhere.

5 Interpretability = understanding *why* a model predicted something; essential for clinical trust.

6 Interpretable models (regression, small trees) explain themselves; black-box models (deep networks) usually cannot.

7 Accountability rule: the clinician who acts on AI advice remains responsible.

8 Pharmacist safeguards: verify, know limits, keep human-in-the-loop, fight alert fatigue, document, report failures.

Common Confusions

Confusion	Clarification
"A 99% accurate model is safe to use"	Only if tested on new, representative data — 99% on training data may be overfitting.
"The software company is responsible for a wrong dose"	No — the prescribing clinician and dispensing pharmacist are accountable.
"Interpretable means less accurate, so it is worse"	Not necessarily worse — in the clinic, a trustworthy explanation often beats a mysterious extra 2% accuracy.
"Bias means the model is prejudiced on purpose"	No — bias usually comes silently from unrepresentative data, not from intent.
"Once deployed, the model stays good forever"	No — data drift means models need regular re-checking.

Common Mistakes

1 Trusting overall accuracy without checking performance in vulnerable groups.

2 Accepting a CDSS recommendation without independent verification.

3 Ignoring alerts because "most alerts are useless" — one may be critical.

4 Using a model trained on one population for a very different population without re-testing.

5 Forgetting to document why an AI suggestion was accepted or overridden.

Mnemonics

- **Seven limitations — "Poor Boys Often Drink Cold Apple Falooda"**: Poor data, Bias, Overfitting, Drift, (missing) Context, Alert fatigue, False confidence.
- **Accountability — "CASH-V"**: Clinician decides, hospital Administers safely, vendor Shares limits, Human-in-the-loop, Verify and document.
- **Over vs under — "Memorizer fails new, Lazy fails all"**: the overfitter memorizes (fails new data), the underfitter is lazy (fails everything).

Exam Summary

Chapter 10 covers the limits of clinical AI. Models fail because of poor or biased data, overfitting, data drift, missing bedside context, alert fatigue, and clinician over-trust. Bias makes a model systematically worse for one patient group, and a good overall accuracy can hide it — performance must be checked group by group. Overfitting means memorizing training data (excellent on training, poor on new patients); underfitting means the model is too simple (poor everywhere). Interpretability — knowing *why* a model predicted something — is essential for clinical trust; regression and small trees are interpretable, deep networks are black boxes. Accountability rests with the clinician who acts on the advice; hospitals must validate models and keep humans in the loop. Pharmacists protect patients by verifying every AI suggestion, knowing model limits, reading alerts, documenting decisions, and reporting failures.

One-Minute Revision

Limits: bad data, bias, overfitting, drift, no context, alert fatigue, over-trust; bias hides behind good overall accuracy — check groups; overfit = memorizes, underfit = too simple; interpretability = knowing why (regression yes, deep nets no); clinician is accountable; pharmacist safeguards: verify, know limits, human-in-the-loop, read alerts, document, report.

Short-Answer Questions

1 List any four limitations of AI in clinical environments.

2 Define bias in AI with one healthcare example.

3 What is the difference between overfitting and underfitting?

4 Why can a high overall accuracy hide bias? Explain with the toy example from this chapter.

5 Define interpretability and give one example each of an interpretable and a black-box model.

6 Who is accountable when an AI-assisted decision harms a patient? Why?

7 What is alert fatigue, and why is it dangerous?

8 What is data drift?

9 What is confounding? Give one pharmacy example.

10 List three practical safeguards a pharmacist should follow when using a CDSS.

Long-Answer Questions

1 Describe the seven key limitations of AI in clinical environments with examples, and explain how each can lead to patient harm.

2 Using the worked toy example, explain how bias hides behind overall accuracy. How should performance be checked to detect such bias?

3 Explain overfitting and underfitting with diagrams. How is each detected, and what are the conceptual remedies?

4 Discuss interpretability in clinical AI: what it means, why clinicians need it, and the trade-off between interpretable and black-box models.

5 Discuss accountability in AI-assisted healthcare: the responsibilities of the clinician, hospital, developer, and regulator, plus the pharmacist's safety checklist.

MCQs

1 A model with 99% accuracy on training data but 60% on new patients is showing:

a) Underfitting b) Overfitting c) Good generalization d) Data drift

1 A risk model performs well for men but poorly for women because the training data had few women. This is:

a) Overfitting b) Bias c) Alert fatigue d) Interpretability

1 Which model is most interpretable?

a) Deep neural network b) Logistic regression c) Large ensemble d) None of these

1 When a CDSS-suggested dose harms a patient, who is primarily accountable?

a) The software vendor b) The clinician who accepted and acted on the suggestion c) The computer hardware d) Nobody

1 Alert fatigue leads to:

a) Better reading of every alert b) Ignoring alerts, including critical ones c) Faster computers d) Higher model accuracy

Answer Key: 1-b, 2-b, 3-b, 4-b, 5-b

Viva Questions

1 Name three limitations of AI in clinical use.

2 What is the difference between bias and confounding?

3 How do you detect overfitting?

4 Why does a doctor need to know *why* a model predicted high risk?

5 Who is accountable for an AI-assisted prescribing error?

6 What should you do when a CDSS alert looks wrong?

7 Give one example of data drift.

Application-Based Questions

1 A vendor claims their ADR model is "98% accurate" based on training data from one hospital. As the hospital pharmacist, list five questions you would ask before allowing its use.

2 Your hospital's CDSS generates 200 alerts a day and doctors now ignore most. Propose three practical steps to reduce alert fatigue.

3 A readmission model flags elderly patients as high risk far less often than they are actually readmitted. Identify the problem (bias type) and suggest two corrective actions.

References

- Aggarwal & Reddy, Healthcare Data Analytics
- Bonate, Pharmacokinetic–Pharmacodynamic Modeling and Simulation
- Campbell, Machin & Walters, Medical Statistics: A Textbook for the Health Sciences
- Schmidt & Derendorf, Applied Pharmacometrics
- Steyerberg, Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating
- Strom, Kimmel & Hennessy, Pharmacoepidemiology