Università “Magna Graecia” di Catanzaro

SYLLABUS DELL'INSEGNAMENTO

Data Mining e Bioimmagini

Metodi computazionali per l'analisi dei dati e delle immagini biomediche

---

1. Informazioni generali

Voce	Informazione
Insegnamento	“Sistemi di elaborazione delle Informazioni - C.I. Data Mining e Bioimmagini”
Anno accademico	2026/27
Corso di Studio	Corso di Laurea Magistrale in Medicina e Chirurgia TD
Dipartimento	Scienze della Salute
Docente	Domenico Costanzo
CFU	6
Ore di didattica	78
SSD	IINF-05/A
Tipologia dell'insegnamento	In presenza
Lingua di insegnamento	Italiano
Modalità di erogazione	Lezioni teoriche e attività pratiche di laboratorio

---

2. Presentazione dell'insegnamento

L'insegnamento di **Data Mining e Bioimmagini** introduce i fondamenti metodologici, statistici e computazionali necessari per l'analisi dei dati biomedici e delle immagini ottenute mediante tecniche di acquisizione diagnostica e sperimentale.

Il corso affronta il processo che conduce dall'osservazione di un fenomeno biologico o clinico alla sua rappresentazione computazionale, alla costruzione di modelli di apprendimento automatico e alla valutazione critica dei risultati.

Una prima parte è dedicata alla comprensione, organizzazione e preparazione dei dati: rappresentazione delle osservazioni, tipologie di variabili, analisi esplorativa, qualità dei dati, trattamento dei valori mancanti, identificazione delle osservazioni anomale, normalizzazione, standardizzazione e costruzione delle feature.

Successivamente vengono introdotti i principali paradigmi del Machine Learning, con particolare attenzione all'apprendimento supervisionato e non supervisionato, agli algoritmi di classificazione e clustering, alla riduzione della dimensionalità, alle reti neurali artificiali e all'apprendimento automatico delle rappresentazioni.

La seconda area tematica approfondisce le bioimmagini come particolare forma di dato biomedico. Vengono considerate le principali modalità di imaging, la struttura e l'interpretazione dei dati DICOM, le tecniche di preprocessing, filtraggio, enhancement, segmentazione ed estrazione di caratteristiche quantitative.

Il percorso introduce infine le reti neurali convoluzionali e le principali problematiche connesse all'applicazione dell'intelligenza artificiale alle immagini biomediche, con attenzione alla valutazione dei modelli, alla riproducibilità, all'interpretabilità e ai limiti dell'inferenza computazionale in ambito clinico.

Il corso adotta un'impostazione teorico-pratica: i concetti matematici e metodologici sono affiancati da esercitazioni realizzate in Python, attraverso notebook Jupyter e librerie per il calcolo scientifico, il Machine Learning e l'elaborazione delle immagini.

Idea guida dell'insegnamento

> **Comprendere il fenomeno biomedico, rappresentarlo attraverso dati strutturati, analizzarlo mediante metodi computazionali e interpretare criticamente i risultati ottenuti.**

---

3. Obiettivi formativi

L'obiettivo generale è fornire allo studente le conoscenze teoriche e le competenze operative necessarie per comprendere e utilizzare le principali metodologie di Data Mining e di analisi computazionale delle bioimmagini.

In particolare, l'insegnamento intende sviluppare la capacità di:

- comprendere il rapporto tra fenomeno biomedico, osservazione, dato e rappresentazione computazionale;
- organizzare e documentare dataset biomedici;

- utilizzare strumenti statistici e computazionali per l'analisi esplorativa;
- riconoscere e affrontare problemi di qualità dei dati;
- applicare tecniche di preprocessing e feature engineering;
- comprendere i principali algoritmi di apprendimento automatico;
- costruire e valutare modelli di classificazione e clustering;
- interpretare le metriche di valutazione e le strategie di validazione;
- comprendere il principio dell'apprendimento di rappresentazioni mediante reti neurali;
- leggere e interpretare dati di imaging in formato DICOM;
- applicare metodologie di base per il preprocessing e l'elaborazione quantitativa delle immagini;
- comprendere i principi delle reti convoluzionali e del Deep Learning applicato alle bioimmagini;
- riconoscere i limiti dei modelli computazionali e distinguere prestazione algoritmica, interpretabilità e validità clinica.

---

4. Risultati di apprendimento attesi

Al termine dell'insegnamento lo studente dovrà aver acquisito conoscenze e capacità coerenti con i seguenti risultati di apprendimento.

4.1 Conoscenza e capacità di comprensione

Lo studente sarà in grado di:

- descrivere le principali fasi del processo di Data Mining;
- distinguere dati numerici, categorici, segnali e immagini;
- comprendere i fondamenti matematici e statistici dei principali algoritmi trattati;
- descrivere il funzionamento dei principali modelli di Machine Learning;
- comprendere i principi delle reti neurali artificiali;
- riconoscere le caratteristiche delle principali modalità di bioimaging;
- comprendere la struttura informativa di un oggetto DICOM;
- descrivere i principi fondamentali dell'elaborazione e dell'analisi quantitativa delle immagini.

4.2 Capacità di applicare conoscenza e comprensione

Lo studente sarà in grado di:

- utilizzare Python e le principali librerie scientifiche;
- costruire, ispezionare e documentare dataset;
- effettuare analisi esplorative e statistiche;
- applicare tecniche di preprocessing;
- costruire matrici di feature e variabili target;

- implementare algoritmi di classificazione e clustering;
- applicare tecniche di riduzione della dimensionalità;
- costruire e addestrare semplici reti neurali;
- leggere immagini DICOM e accedere ai relativi metadata e pixel;
- eseguire operazioni di base di preprocessing e visualizzazione delle bioimmagini;
- produrre elaborati computazionali riproducibili.

4.3 Autonomia di giudizio

Lo studente dovrà sviluppare la capacità di:

- valutare la qualità e l'adeguatezza dei dati rispetto al problema scientifico;
- riconoscere possibili errori e distorsioni nei processi di acquisizione e preparazione;
- scegliere metodologie coerenti con la natura dei dati;
- interpretare criticamente le metriche di valutazione;
- riconoscere overfitting, data leakage e limiti di generalizzazione;
- distinguere risultati statistici, risultati computazionali e conclusioni cliniche;
- riconoscere le problematiche relative alla riservatezza e alla gestione dei dati sanitari;
- evitare interpretazioni cliniche non supportate dalle evidenze disponibili.

4.4 Abilità comunicative

Lo studente dovrà essere in grado di:

- descrivere un problema biomedico in termini computazionali;
- documentare le caratteristiche di un dataset;
- spiegare le scelte metodologiche effettuate;
- presentare risultati mediante tabelle e visualizzazioni;
- illustrare il funzionamento generale di un modello;
- discutere i limiti di un'analisi e le condizioni di validità dei risultati.

4.5 Capacità di apprendimento

Lo studente dovrà acquisire la capacità di:

- consultare documentazione tecnica e scientifica;
- utilizzare autonomamente librerie e strumenti computazionali;
- approfondire algoritmi e metodologie non trattati in modo esaustivo;
- trasferire le competenze acquisite a nuovi problemi biomedici;
- aggiornare le proprie conoscenze nell'ambito del Data Mining, del Machine Learning e del Deep Learning.

---

5. Prerequisiti

Sono consigliate conoscenze di base di:

- matematica generale;
- algebra e funzioni;
- elementi di algebra lineare;
- statistica descrittiva;
- concetti introduttivi di probabilità;
- utilizzo del computer e gestione di file e directory.

È inoltre utile possedere conoscenze introduttive di programmazione.

Le competenze operative di Python necessarie per il corso saranno consolidate attraverso le guide introduttive e le prime esercitazioni.

Non è richiesta una precedente esperienza approfondita di Machine Learning o di elaborazione delle immagini.

---

6. Programma dell'insegnamento

Il programma è organizzato in aree tematiche progressive. La successione degli argomenti segue il processo che conduce dalla rappresentazione dei dati alla costruzione, valutazione e interpretazione dei modelli, fino all'analisi delle immagini biomediche.

Unità 1 — Ambiente computazionale e rappresentazione dei dati biomedici

- Introduzione al Data Mining e alle Bioimmagini.
- Ambiente computazionale Python.
- Interprete, pacchetti, ambienti virtuali e kernel.
- JupyterLab e notebook.
- Librerie scientifiche Python.
- Dal fenomeno biomedico all'osservazione.
- Unità osservazionale e variabili.
- Identificatori, feature e target.
- Dataset tabellari e data dictionary.
- Formati di memorizzazione e organizzazione dei dati.

Attività pratica: preparazione dell'ambiente Python, costruzione di un dataset biomedico sintetico e documentazione delle variabili.

Unità 2 — Data Understanding e qualità dei dati

- Ispezione e comprensione dei dataset.
- Analisi esplorativa dei dati (EDA).
- Statistiche descrittive.
- Distribuzioni e visualizzazioni.
- Variabili numeriche e categoriche.
- Dati mancanti.

- Completezza, validità e coerenza.
- Outlier statistici e interpretazione biomedica.
- Rumore e variabilità biologica.
- Errori di acquisizione e di registrazione.
- Normalizzazione e standardizzazione.
- Preprocessing e feature engineering.

Attività pratica: analisi esplorativa, identificazione dei problemi di qualità e realizzazione di una prima pipeline di preprocessing.

Unità 3 — Fondamenti statistici e rappresentazione matematica dei dati

- Statistica descrittiva e distribuzioni.
- Media, mediana, varianza e deviazione standard.
- Percentili e quartili.
- Correlazione e dipendenza.
- Concetto di probabilità e variabili casuali.
- Campionamento e variabilità.
- Matrici, vettori e spazi delle feature.
- Distanze e misure di similarità.
- Basi matematiche per gli algoritmi di apprendimento.

Attività pratica: calcolo di statistiche descrittive e analisi delle relazioni tra variabili mediante NumPy, Pandas e Matplotlib.

Unità 4 — Introduzione al Machine Learning

- Definizione e obiettivi del Machine Learning.
- Apprendimento supervisionato e non supervisionato.
- Training set, validation set e test set.
- Feature matrix e target.
- Modelli, parametri e iperparametri.
- Classificazione e regressione.
- Pipeline computazionale.
- Generalizzazione.
- Overfitting e underfitting.
- Data leakage.

Attività pratica: preparazione del dataset per un primo problema di classificazione.

Unità 5 — Apprendimento supervisionato: algoritmi di classificazione

- k-Nearest Neighbors (k-NN).
- Logistic Regression.
- Decision Tree.
- Random Forest.

- Support Vector Machine (SVM).
- Pattern Recognition.
- Funzioni decisionali e confini di classificazione.
- Scelta degli iperparametri.
- Interpretazione del comportamento dei classificatori.

Attività pratica: implementazione e confronto di differenti classificatori su un dataset biomedico.

Unità 6 — Valutazione e validazione dei modelli

- Confusion matrix.
- True Positive, True Negative, False Positive e False Negative.
- Accuracy.
- Precision.
- Recall e sensibilità.
- Specificità.
- F1-score.
- Soglia decisionale.
- Curve ROC e Area Under the Curve (AUC).
- Cross-validation e Stratified K-Fold.
- Selezione e confronto dei modelli.
- Generalizzazione e validazione esterna.
- Limiti delle metriche nel dominio biomedico.

Attività pratica: confronto delle prestazioni dei classificatori, analisi degli errori e valutazione mediante cross-validation e ROC/AUC.

Unità 7 — Apprendimento non supervisionato e clustering

- Apprendimento non supervisionato.
- Concetto di clustering.
- Similarità e distanze.
- K-Means.
- Centroidi e variabilità intra-cluster.
- Scelta del numero di cluster.
- Metodo del gomito.
- Silhouette score.
- Clustering gerarchico.
- Dendrogrammi e strategie di linkage.
- DBSCAN.
- Rumore e osservazioni non assegnate.
- Interpretazione dei gruppi nei dati biomedici.
- Limiti del clustering nella definizione di sottogruppi clinici.

- Apprendimento supervisionato e non supervisionato delle rappresentazioni.

Attività pratica: individuazione e interpretazione di gruppi in un dataset biomedico senza utilizzare il target durante l'apprendimento.

Unità 8 — Reti neurali artificiali e Deep Learning

- Neurone artificiale.
- Pesi e bias.
- Funzioni di attivazione.
- Perceptron e Multi-Layer Perceptron (MLP).
- Strati di input, hidden e output.
- Forward propagation.
- Funzioni di loss.
- Gradient descent.
- Backpropagation.
- Autodifferenziazione.
- Ottimizzatori.
- Epoche, batch e iterazioni.
- Early stopping e regolarizzazione.
- Rappresentazioni latenti.
- Representation learning.
- Limiti di interpretabilità delle reti neurali.

Attività pratica: costruzione e addestramento di una rete MLP mediante PyTorch, con estrazione e visualizzazione della rappresentazione latente.

Unità 9 — Fondamenti di bioimaging

- Il ruolo delle immagini nei processi biomedici.
- Immagini digitali e matrici di pixel.
- Immagini bidimensionali e volumi tridimensionali.
- Risoluzione spaziale e risoluzione in intensità.
- Pixel e voxel.
- Modalità di imaging:
    - radiografia;
    - tomografia computerizzata (CT);
    - risonanza magnetica (MRI);
    - ecografia;
    - PET e SPECT;
    - OCT.
- Principi generali di acquisizione.
- Informazione anatomica e funzionale.
- Differenze tra modalità e significato dei valori di intensità.

Attività pratica: visualizzazione e confronto di immagini e rappresentazioni sintetiche di differenti modalità di imaging.

Unità 10 — DICOM e rappresentazione strutturata dell'immagine clinica

- Introduzione allo standard DICOM.
- Oggetto DICOM e file DICOM.
- Gerarchia Patient, Study, Series e Instance.
- Metadata e Pixel Data.
- Data Elements, Tag e Value Representation.
- Unique Identifiers (UID).
- SOP Class e SOP Instance.
- Dimensioni dell'immagine.
- Pixel Spacing e Slice Thickness.
- Image Position e Image Orientation.
- Spazio dei pixel e spazio fisico.
- Rescale Slope e Rescale Intercept.
- Organizzazione delle serie e ricostruzione volumetrica.
- Privacy, anonimizzazione e pseudonimizzazione.
- Lettura dei dati DICOM mediante pydicom.

Attività pratica: caricamento di oggetti DICOM, ispezione dei metadata, estrazione dei pixel e costruzione di un report tecnico.

Unità 11 — Image Processing e preprocessing delle bioimmagini, Segmentazione e analisi quantitativa delle immagini

- Rappresentazione numerica delle immagini.
- Istogrammi e distribuzioni delle intensità.
- Windowing e contrast enhancement.
- Normalizzazione delle intensità.
- Filtraggio spaziale.
- Kernel e convoluzione.
- Filtro media e filtro mediano.
- Filtro Gaussiano.
- Riduzione del rumore.
- Riconoscimento e gestione degli artefatti.
- Concetto di segmentazione.
- Segmentazione manuale, semiautomatica e automatica.
- Thresholding.
- Segmentazione basata su regioni.
- Operazioni morfologiche.
- Contorni e regioni di interesse.
- Maschere binarie.

- Misure di area, perimetro e forma.
- Segmentazione di strutture anatomiche e patologiche.
- Valutazione della segmentazione.
- Introduzione alle metriche di sovrapposizione, tra cui Dice coefficient e Intersection over Union.

Attività pratica: visualizzazione, filtraggio, confronto tra immagini originali e trasformate e analisi dell'effetto del preprocessing. Segmentazione introduttiva di regioni di interesse e calcolo di semplici caratteristiche geometriche.

7. Metodologie didattiche

L'insegnamento integra lezioni teoriche, dimostrazioni computazionali ed esercitazioni pratiche.

Lezioni frontali

Sono dedicate alla presentazione dei concetti, dei modelli matematici e dei principi metodologici.

Particolare attenzione viene riservata alla comprensione delle ipotesi alla base degli algoritmi e delle condizioni necessarie per interpretarne correttamente i risultati.

Esercitazioni guidate

Le esercitazioni consentono allo studente di applicare progressivamente le conoscenze acquisite attraverso programmi Python e notebook Jupyter.

Le attività comprendono:

- costruzione e analisi di dataset;
- elaborazione statistica;
- implementazione di algoritmi;
- confronto tra modelli;
- analisi di immagini;
- lettura di file DICOM;
- visualizzazione e interpretazione dei risultati.

Laboratorio computazionale

Il laboratorio utilizza un ambiente Python organizzato e verificabile.

Lo studente impara a documentare:

- ambiente software;
- librerie e versioni;
- dati di ingresso;

- parametri degli algoritmi;
- procedure di preprocessing;
- risultati prodotti.

La riproducibilità viene considerata una componente trasversale dell'intero insegnamento.

Discussione dei risultati

Le esercitazioni non si concludono con la sola esecuzione del codice.

Lo studente deve essere in grado di interpretare i risultati, riconoscere eventuali anomalie, formulare osservazioni e discutere i limiti delle procedure utilizzate.

---

8. Strumenti software e tecnologici

Le attività pratiche saranno svolte principalmente mediante Python e librerie open source.

Strumento	**Utilizzo didattico**
Python	Linguaggio di programmazione
JupyterLab	Ambiente interattivo per notebook
NumPy	Calcolo numerico e matriciale
Pandas	Gestione dei dati tabellari
Matplotlib	Visualizzazione
SciPy	Calcolo scientifico
Scikit-learn	Machine Learning e preprocessing
PyTorch	Reti neurali e Deep Learning
pydicom	Lettura e gestione di dati DICOM
scikit-image	Elaborazione delle immagini
SimpleITK	Elaborazione di immagini e volumi medicali, ove previsto

Le librerie e le versioni effettivamente utilizzate saranno specificate nelle istruzioni operative delle esercitazioni.

L'uso di acceleratori hardware, GPU o dispositivi specializzati non costituisce un prerequisito per le attività introduttive.

---

9. Materiale didattico

Il materiale didattico comprende:

- syllabus dell'insegnamento;
- guide introduttive all'ambiente Python;
- dispense e presentazioni delle lezioni;
- notebook delle esercitazioni;
- dataset didattici;
- documentazione tecnica delle librerie;
- riferimenti scientifici e normativi.

Le dispense e i notebook predisposti per il corso costituiscono il riferimento operativo principale per lo svolgimento delle esercitazioni.

Testi e riferimenti bibliografici consigliati

Data Mining, statistica e Machine Learning

1. Bishop, C. M., *Pattern Recognition and Machine Learning*, Springer, 2006.
2. Hastie, T., Tibshirani, R., Friedman, J., *The Elements of Statistical Learning*, Springer, 2009.
3. Géron, A., *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, O'Reilly, 3rd edition, 2022.

Deep Learning

1. Goodfellow, I., Bengio, Y., Courville, A., *Deep Learning*, MIT Press, 2016.
2. Documentazione ufficiale PyTorch, con particolare riferimento a `torch.nn`, Autograd e agli algoritmi di ottimizzazione.

Bioimmagini e DICOM

1. DICOM Standards Committee, *Digital Imaging and Communications in Medicine (DICOM) Standard*.
2. Documentazione ufficiale pydicom.
3. Documentazione ufficiale scikit-image.
4. Documentazione ufficiale SimpleITK.

Documentazione software

1. Documentazione ufficiale NumPy.
2. Documentazione ufficiale Pandas.
3. Documentazione ufficiale SciPy.
4. Documentazione ufficiale Scikit-learn.
5. Documentazione ufficiale Matplotlib.

---

10. Modalità di verifica dell'apprendimento

La verifica deve accertare sia la comprensione teorica sia la capacità di applicare correttamente gli strumenti computazionali.

Si propone una valutazione articolata in due componenti.

10.1 Prova pratica

La prova pratica consiste nello sviluppo o nella discussione di un'attività computazionale coerente con gli argomenti del corso.

Potrà comprendere:

- caricamento e organizzazione di un dataset;
- analisi esplorativa;
- preprocessing;
- costruzione di un modello;
- valutazione delle prestazioni;
- elaborazione di un'immagine;
- lettura di metadata DICOM;
- interpretazione dei risultati.

Lo studente dovrà dimostrare di saper utilizzare correttamente gli strumenti e di comprendere le scelte metodologiche effettuate.

10.2 Colloquio teorico

Il colloquio accerta:

- conoscenza dei concetti fondamentali;
- comprensione dei modelli matematici;
- capacità di descrivere il funzionamento degli algoritmi;
- capacità di interpretare le metriche;
- comprensione delle problematiche di qualità dei dati;
- consapevolezza dei limiti dei metodi applicati;
- capacità di collegare le metodologie ai problemi biomedici.

10.3 Criteri generali di valutazione

La valutazione terrà conto di:

1. correttezza dei concetti teorici;
2. comprensione dei fondamenti matematici e statistici;
3. correttezza dell'implementazione;
4. appropriatezza delle scelte metodologiche;
5. capacità di valutare e interpretare i risultati;
6. autonomia nell'affrontare problemi computazionali;

7. capacità di documentare e riprodurre l'esperimento;
8. consapevolezza dei limiti e delle implicazioni biomediche.

La modalità definitiva di esame, la ripartizione del punteggio e l'eventuale obbligatorietà delle esercitazioni saranno definite in conformità al regolamento del Corso di Studio.

---

11. Attività pratiche e prodotti attesi

Durante il percorso lo studente sarà guidato nella realizzazione di una serie progressiva di elaborati computazionali.

Tra i prodotti didattici previsti:

- notebook di verifica dell'ambiente Python;
- dataset biomedico strutturato;
- data dictionary;
- report di analisi esplorativa;
- report di qualità dei dati;
- pipeline di preprocessing;
- modelli di classificazione;
- report di valutazione e confronto dei modelli;
- risultati di clustering e riduzione della dimensionalità;
- modello MLP e rappresentazione latente;
- notebook di analisi di immagini;
- report di lettura e interpretazione DICOM;
- risultati di preprocessing e segmentazione;
- estrazione di caratteristiche quantitative;
- esperimento introduttivo con reti convoluzionali;
- report di interpretabilità e discussione critica.

Gli elaborati saranno organizzati in un progetto Python strutturato, nel quale dati, notebook, codice, figure e report saranno mantenuti separati.

---

12. Approccio metodologico e principi trasversali

L'insegnamento adotta alcuni principi che devono accompagnare tutte le attività teoriche e pratiche.

12.1 Comprendere prima di modellare

Prima di applicare un algoritmo è necessario comprendere il fenomeno, l'unità osservazionale e il significato delle variabili.

12.2 La qualità del dato condiziona il risultato

La correttezza di un modello dipende anche dalla qualità, dalla provenienza e dalla rappresentatività dei dati utilizzati.

12.3 La valutazione è parte integrante dell'apprendimento

La prestazione sul training set non costituisce una dimostrazione di generalizzazione.

La valutazione deve utilizzare procedure appropriate e mantenere separati i dati impiegati per addestramento, selezione e test.

12.4 Una rappresentazione quantitativa non coincide con il fenomeno biologico

Le feature, le componenti principali, le coordinate latenti e le predizioni sono rappresentazioni computazionali.

La loro interpretazione richiede conoscenza del dominio e verifica scientifica.

12.5 Correttezza computazionale e validità clinica sono distinte

Un codice può essere eseguito correttamente e produrre metriche elevate senza che il risultato possieda automaticamente validità clinica.

12.6 Riproducibilità

Ogni esperimento dovrebbe essere documentato attraverso:

- dati utilizzati;
- ambiente software;
- versioni delle librerie;
- procedure applicate;
- parametri;
- risultati;
- limiti dell'analisi.

12.7 Protezione dei dati e responsabilità

I dati biomedici e le immagini cliniche possono contenere informazioni personali e sensibili.

Il loro utilizzo deve rispettare le regole di riservatezza, sicurezza e gestione dei dati applicabili al contesto didattico e scientifico.

---

13. Collegamento tra Data Mining e Bioimmagini

Il corso è concepito come un percorso integrato.

Il Data Mining fornisce gli strumenti per organizzare, rappresentare, analizzare e modellare i dati.

Il Bioimaging introduce una particolare classe di dati, caratterizzata da struttura spaziale, modalità di acquisizione, geometria e metadata.

Le due aree si incontrano nel processo di estrazione di informazione quantitativa dalle immagini.

Il percorso complessivo può essere rappresentato così:

```
FENOMENO BIOMEDICO
        ↓
OSSERVAZIONE
        ↓
ACQUISIZIONE DEL DATO
        ↓
RAPPRESENTAZIONE
        ↓
DATA UNDERSTANDING
        ↓
DATA QUALITY
        ↓
PREPROCESSING
        ↓
FEATURE ENGINEERING
        ↓
MACHINE LEARNING
        ↓
DEEP LEARNING
        ↓
BIOIMAGING E DICOM
        ↓
IMAGE PROCESSING
        ↓
SEGMENTAZIONE
        ↓
FEATURE EXTRACTION
        ↓
VALIDAZIONE
        ↓
INTERPRETAZIONE BIOMEDICA
```

Questo schema evidenzia la continuità tra la rappresentazione del fenomeno, la sua elaborazione quantitativa e la costruzione dei modelli.

---

14. Articolazione temporale

La distribuzione degli argomenti e delle esercitazioni sarà definita sulla base dei CFU e delle ore di didattica assegnate.

A titolo organizzativo, il corso può essere articolato in quattro macrofasi:

Fase	Contenuto	Risultato atteso
I — Fondamenti		Comprensione e preparazione dei dataset

Fase	Contenuto	Risultato atteso
	Python, rappresentazione dei dati, statistica, Data Understanding e qualità	
II — Data Mining	Machine Learning supervisionato e non supervisionato, valutazione, riduzione dimensionale	Costruzione e analisi di modelli
III — Deep Learning	Reti neurali, MLP e apprendimento delle rappresentazioni	Comprensione dei modelli basati su reti neurali
IV — Bioimmagini	Modalità di imaging, DICOM, Image Processing, segmentazione, feature extraction e CNN	Analisi computazionale delle immagini biomediche

La progressione è modulare: ogni fase utilizza le competenze acquisite nella precedente.

---

15. Destinatari e prospettive formative

L'insegnamento è rivolto a studenti interessati all'applicazione dei metodi computazionali allo studio dei fenomeni biologici, fisiologici e clinici.

Le competenze acquisite potranno essere utilizzate come base per approfondimenti in:

- analisi dei dati sanitari;
- ricerca biomedica;
- bioinformatica;
- Machine Learning applicato alla medicina;
- elaborazione di segnali fisiologici;
- analisi di immagini radiologiche;
- radiomica;
- Deep Learning per bioimmagini;
- sistemi di supporto alla ricerca e alla decisione clinica.

Il corso non ha l'obiettivo di formare sistemi diagnostici autonomi né di sostituire l'interpretazione del professionista sanitario. Mira a fornire una preparazione metodologica e computazionale che consenta di comprendere, progettare e valutare strumenti di analisi dei dati biomedici.

---

16. Profilo delle competenze in uscita

Al termine del percorso, lo studente avrà acquisito una prima capacità di affrontare un problema biomedico seguendo una sequenza metodologica completa:

1. descrivere il fenomeno da studiare;
2. individuare le osservazioni e le variabili rilevanti;
3. costruire o acquisire un dataset;
4. verificarne qualità e coerenza;
5. effettuare l'analisi esplorativa;
6. preparare i dati per l'elaborazione;
7. scegliere un approccio di apprendimento;
8. addestrare un modello;
9. valutare le prestazioni;
10. interpretare i risultati;
11. documentare l'esperimento;
12. riconoscere i limiti delle conclusioni ottenute.

Per i dati di imaging, lo studente dovrà inoltre comprendere il passaggio:

```
IMMAGINE CLINICA
      ↓
DICOM E METADATA
      ↓
PIXEL E GEOMETRIA
      ↓
PREPROCESSING
      ↓
SEGMENTAZIONE
      ↓
FEATURE EXTRACTION
      ↓
RAPPRESENTAZIONE QUANTITATIVA
      ↓
ANALISI E MODELLAZIONE
```

---

17. Conclusione del Syllabus

Data Mining e Bioimmagini propone un percorso formativo che unisce statistica, programmazione, Machine Learning, Deep Learning ed elaborazione delle immagini biomediche.

Il filo conduttore dell'insegnamento è la trasformazione dell'informazione biomedica in rappresentazioni quantitative che possano essere analizzate mediante metodi computazionali, mantenendo sempre la consapevolezza del significato e dei limiti dei risultati.

L'obiettivo non è soltanto insegnare a utilizzare algoritmi e librerie software, ma sviluppare una capacità di ragionamento metodologico: comprendere il problema, scegliere gli

strumenti, verificare i risultati e interpretare correttamente ciò che i modelli sono effettivamente in grado di mostrare.

Principio guida

> **Prima di applicare un algoritmo dobbiamo comprendere il dato; prima di interpretare il risultato dobbiamo comprendere il modello; prima di attribuire un significato biomedico dobbiamo disporre di evidenze adeguate.**

Questo principio costituisce il fondamento metodologico dell'intero insegnamento.

---

18. Modalità di verifica dell'apprendimento e criteri di Valutazione

Organizzazione della didattica

Ore

Totali	Didattica frontale	Pratica (laboratorio, campo, esercitazione, altro)	Studio individuale
248	*39*	*39*	*170*

CFU/ETCS

6

Modalità di verifica dell'apprendimento e criteri di Valutazione

Modalità di verifica dell’apprendimento	Due prove in itinere per accertare la comprensione e l'applicazione del programma. Una prova orale, a discrezione del docente, per valutare la capacità di esporre in modo chiaro e approfondito i concetti appresi.
Criteri di valutazione	Tutte le prove sono orientate a verificare:

- Conoscenza e capacità di comprensione dei contenuti teorici.
- Applicazione delle conoscenze ai problemi pratici.
- Autonomia di giudizio: capacità di sviluppare soluzioni autonome e analisi di contesto.
- Abilità comunicative: chiarezza nell'esposizione e nella spiegazione.
- Capacità di apprendimento: attitudine ad apprendere e approfondire i concetti.

Criteri di misurazione dell'apprendimento e di attribuzione del voto finale

Criteri di Valutazione

Prove scritte e pratiche: Superate con un punteggio minimo di 18/30.

Gli studenti che ottengono un punteggio tra 15/30 e 17/30 sono ammessi alla prova orale con riserva.

Valutazione finale: La media dei punteggi ottenuti nelle prove scritte, pratiche e orali, espressa in trentesimi, determinerà il voto finale.

Lode: Per ottenere la lode, lo studente deve aver raggiunto una media superiore a 27/30 nelle prove scritte e pratiche e dimostrare una conoscenza completa durante le prove orali

Tabella Riassuntiva.

	Conoscenza e comprensione argomento	**Capacità di analisi e sintesi**	**Utilizzo di referenze**
Non idoneo	Importanti carenze. Significative in accuratezze	Irrilevanti. Frequenti generalizzazioni. Incapacità o scarsa capacità di sintesi	Inappropriato
Idoneo	Conoscenza routinaria	E’ in grado di analisi e sintesi corrette.	Utilizza le referenze standard

		Argomenta in modo logico e coerente	
18-20	A livello soglia. Imperfezioni evidenti	Capacità appena sufficienti	Appena appropriato
21-23	Conoscenza routinaria	E’ in grado di analisi e sintesi corrette. Argomenta in modo logico e coerente	Utilizza le referenze standard
24-26	Conoscenza buona	Ha capacità di a. e s. buone gli argomenti sono espressi coerentemente	Utilizza le referenze standard
27-29	Conoscenza più che buona	Ha notevoli capacità di a. e s.	Ha approfondito gli argomenti
30-30L	Conoscenza ottima	Ha notevoli capacità di a. e s.	Importanti approfondimenti