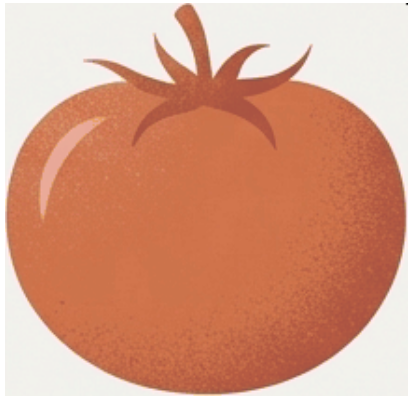


Machine Learning applicato alla metabolomica LC-MS: nuove strategie per autenticità e tracciabilità degli alimenti

Fabiola Eugelio, PhD

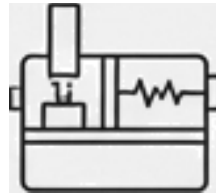
Metabolomica: svelare l'impronta digitale molecolare dei sistemi biologici

La metabolomica è una scienza "omica" che si concentra sullo studio di piccole molecole (metaboliti) in un campione biologico. Cattura gli effetti a valle di fattori genetici e ambientali sulla composizione del cibo, delle piante e degli organismi animali fornendo una visione olistica della loro chimica.



Alimento

Analisi strumentale



Impronta Digitale Molecolare

È come scoprire "l'impronta digitale" unica che definisce la qualità, la sicurezza e il valore nutrizionale di un alimento.

La Foodomica integra scienze "omiche" per migliorare la salute e la fiducia dei consumatori.

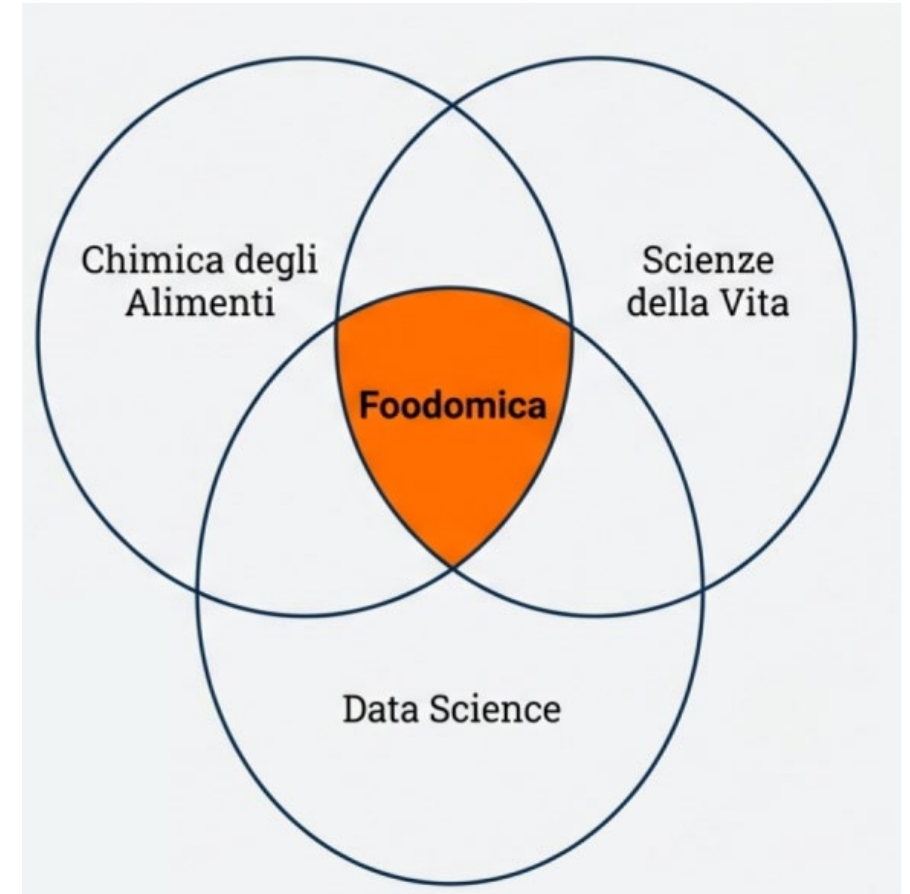
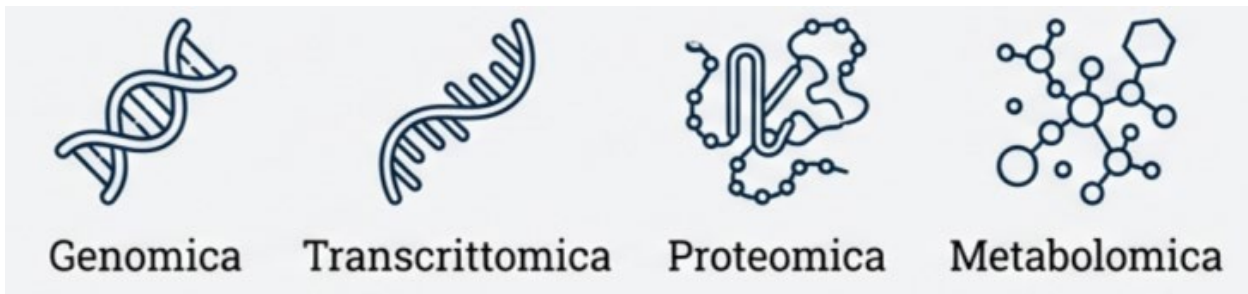
La Foodomica è una disciplina che studia i sistemi alimentari e nutrizionali attraverso l'applicazione di tecnologie omiche avanzate (Genomica, Transcrittomica, Proteomica, Metabolomica).

Obiettivo principale:

Comprenderne qualità, autenticità e proprietà biologiche

Studiare la relazione tra alimentazione, salute e malattie.

Pilastri fondamentali:



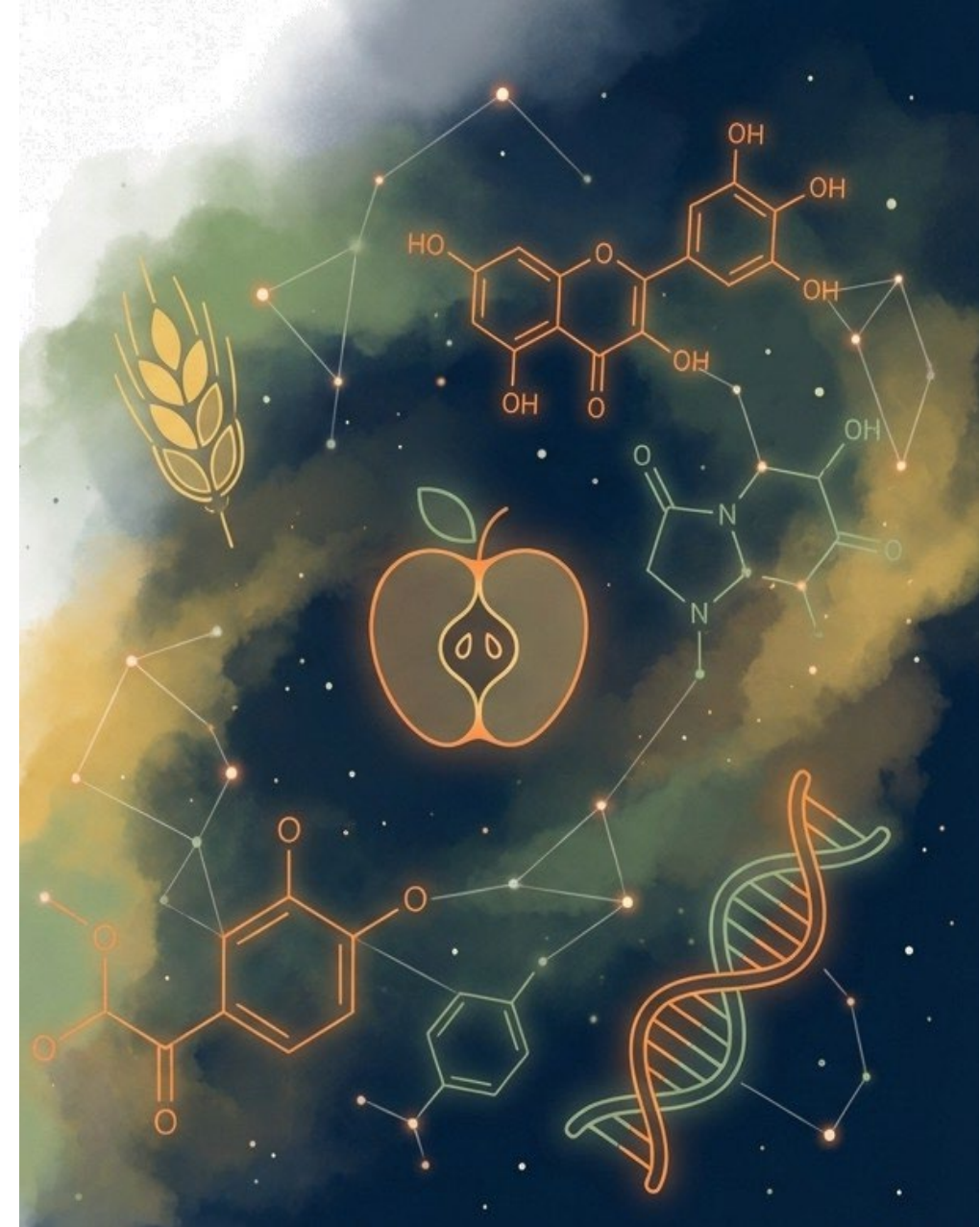
Il cibo è un vasto e complesso universo di dati

Ogni alimento, dal più semplice al più complesso, contiene un'immensa quantità di informazioni chimiche.

La "**Foodomica**" è la disciplina che utilizza tecnologie avanzate per decodificare questo universo, trasformando il cibo da semplice nutrimento a una fonte di dati preziosi per il benessere e la salute.

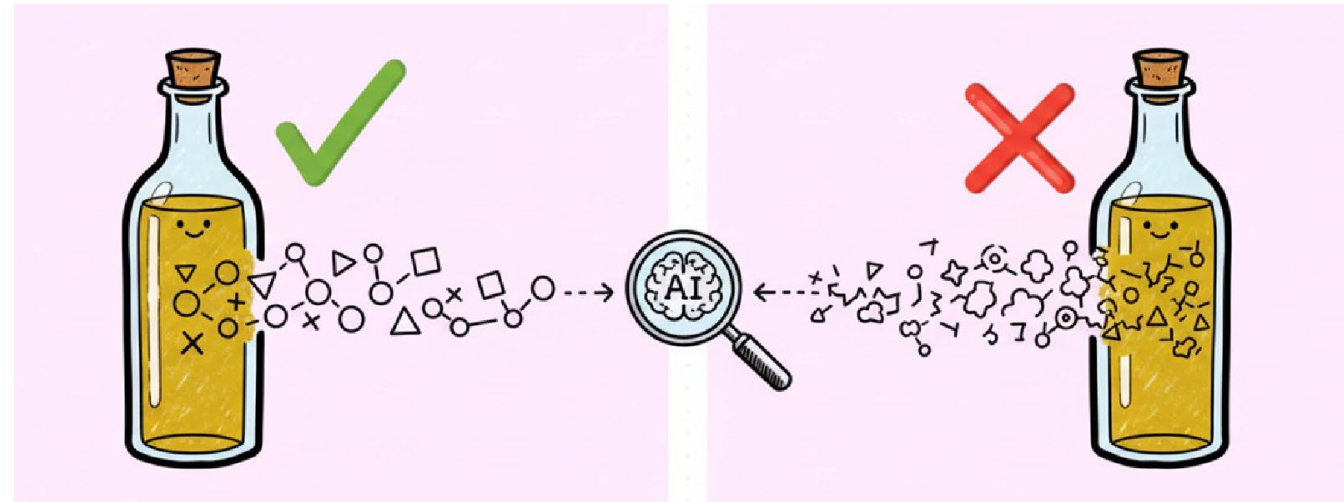
"Foodome":

È un insieme di tutti i composti presenti in un campione alimentare e nel sistema biologico che interagisce con esso.



Una Trasformazione Fondamentale nelle Scienze degli Alimenti

L'unione tra scienze omiche e intelligenza artificiale sta inaugurando un nuovo approccio, aprendo possibilità senza precedenti per tracciabilità, controllo qualità e comprensione della salute. In questo scenario, il cibo non è più solo nutrimento, ma diventa informazione biologica.

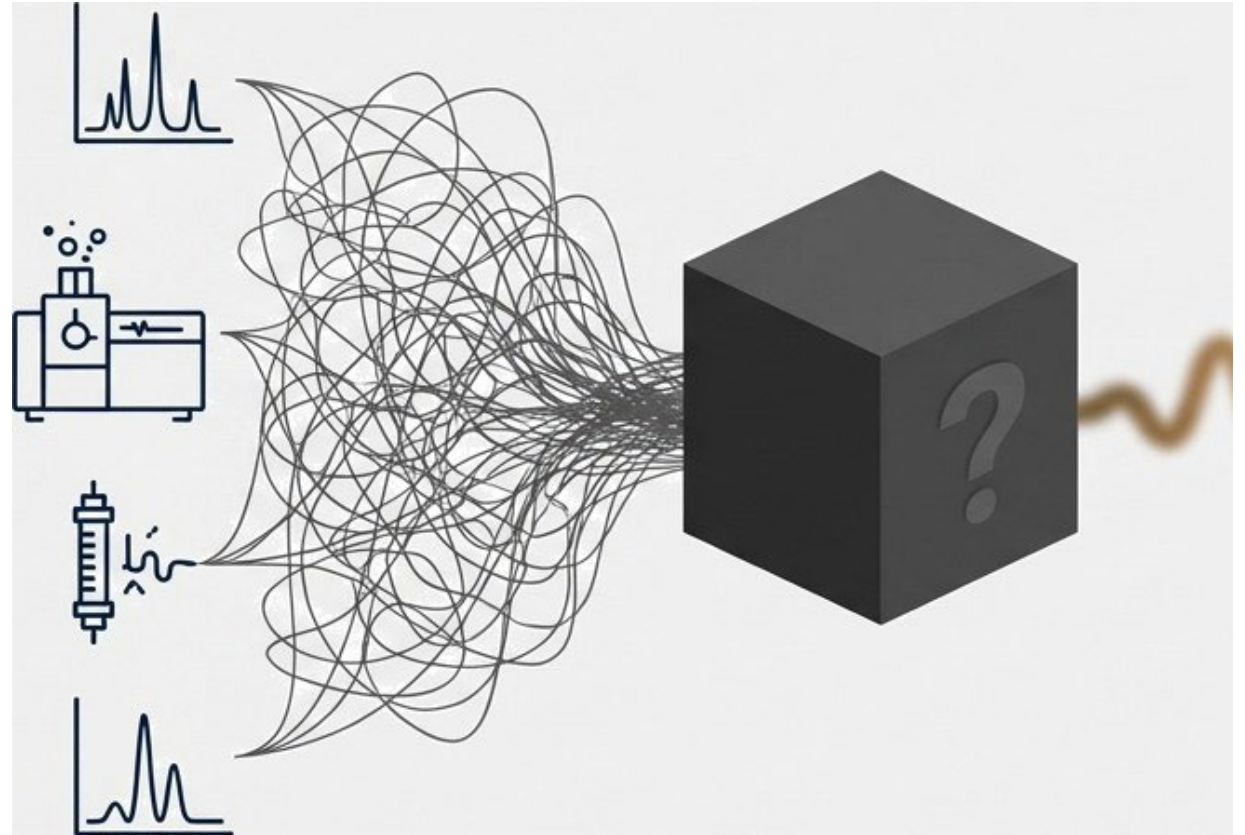


Il ruolo dell'analisi dei dati negli studi di Foodomica

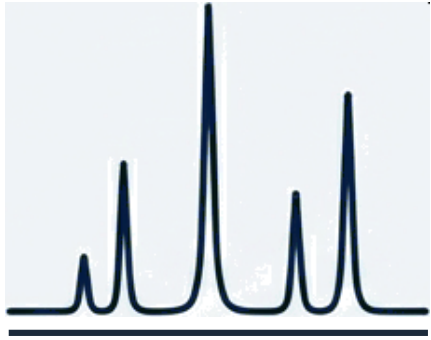
La Sfida del "**Big Data**" Analitico: Le piattaforme moderne generano dati multivariati e megavariati. Un singolo campione può produrre migliaia di segnali strumentali, rendendo complessa l'interpretazione manuale.

Il Rischio del «**Black Box**». Molti ricercatori si affidano a software predefiniti senza una profonda conoscenza metodi di data mining, trattando l'analisi come una "scatola nera". Questo può portare a intuizioni biologiche errate.

Servono competenze e strumenti avanzati di pre-processamento e di analisi multivariata per interpretare criticamente i risultati, per estrarre informazioni significative e non evidenti, nascoste in questi vasti dataset.

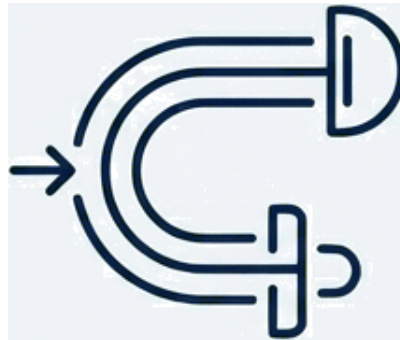


Piattaforme analitiche avanzate generano dataset ampi e complessi



Spettroscopia di Risonanza Magnetica Nucleare (NMR)

Fornisce informazioni complete sulla composizione chimica. È una tecnica non distruttiva e altamente riproducibile.



Spettrometria di Massa (MS)

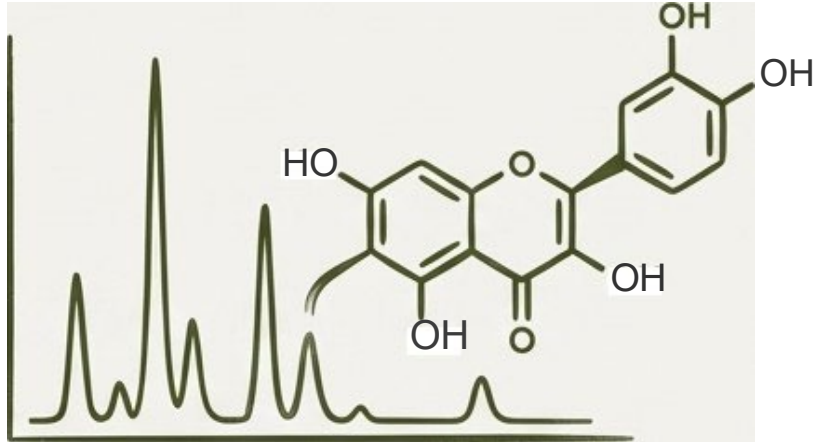
Accoppiata con tecniche cromatografiche, offre altissima sensibilità e specificità per l'identificazione e la quantificazione di migliaia di metaboliti.



Cromatografia (GC & LC)

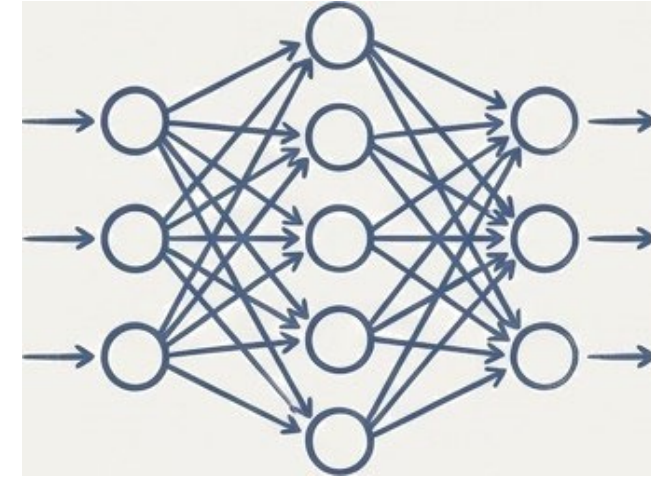
Separa i componenti di miscele complesse. Le tecniche come GC ed LC aumentano esponenzialmente la capacità di risoluzione, indispensabile per analizzare il Foodome.

Scienze omiche – Machine Learning



Analisi -omica

La capacità di catturare l'intera impronta digitale chimica di un alimento. Invece di cercare composti specifici, l'analisi -omica (in particolare la metabolomica) mappa centinaia o migliaia di piccole molecole simultaneamente, rivelando un profilo completo e unico.



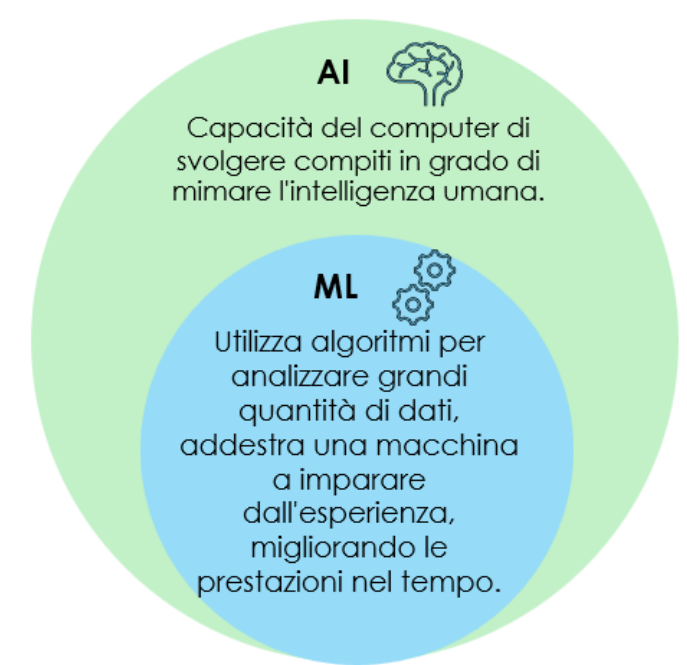
Machine Learning (ML)

L'intelligenza artificiale può analizzare queste complesse impronte digitali. Gli algoritmi di ML sono addestrati per identificare pattern, correlazioni e relazioni predittive nei vasti dataset -omici, trasformando dati grezzi in informazioni.

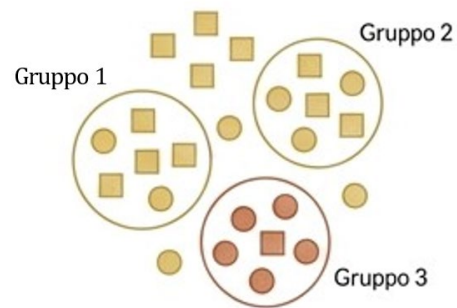


Machine Learning

I dati della metabolomica sono vasti e complessi. Il Machine Learning (ML), una branca dell'intelligenza Artificiale, fornisce algoritmi in grado di scoprire pattern, correlazioni e relazioni predittive in questi dataset su larga scala, estraendo informazioni significative che sarebbero impossibili da identificare per un essere umano.

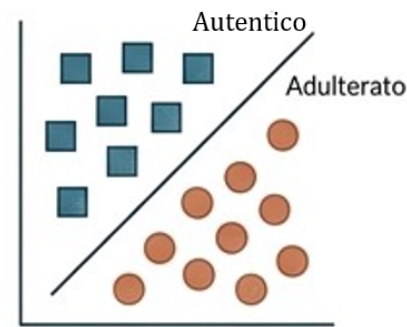


Apprendimento Non Supervisionato



Trova pattern e strutture nascoste in dati non etichettati (es. raggruppare campioni di frutta in base al profilo di "sapore").

Apprendimento Supervisionato



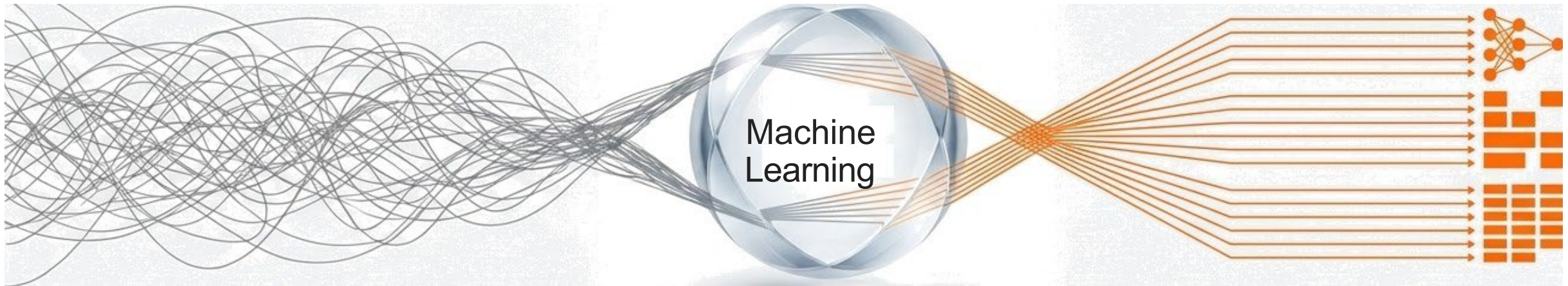
Apprende da dati etichettati per fare previsioni (es. classificare un olio come 'autentico' o 'adulterato').

Apprendimento Profondo (Deep Learning)



Utilizza reti neurali per analizzare dataset molto grandi e complessi (es. riconoscere l'aroma di un alimento da un'immagine spettrale).

Il Machine Learning agisce come una lente intelligente per decifrare la complessità dei dati.

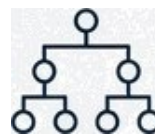


Il Machine Learning fornisce metodi computazionali per trattare vasti dati in modo intelligente.

Approcci Principali:

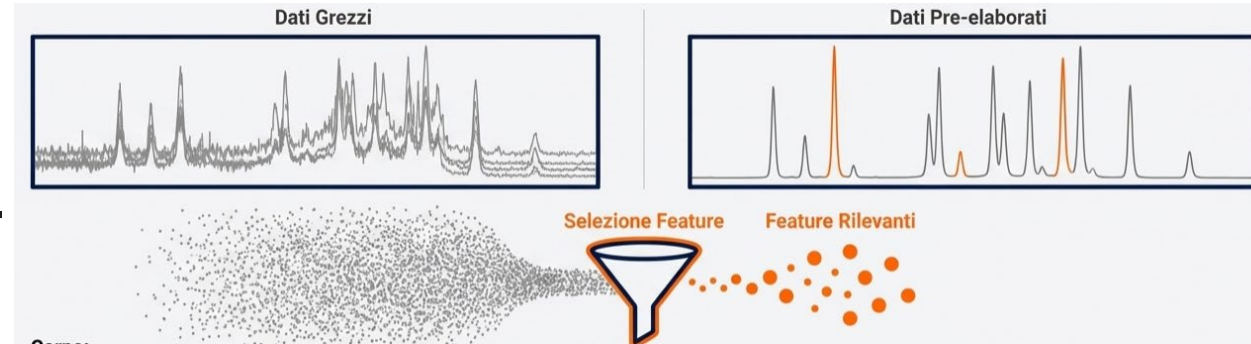
Non Supervisionato (Esplorazione): Esplora la presenza di raggruppamenti naturali tra i campioni senza informazioni pre-esistenti." esempio: Principal Component Analysis (PCA), Hierarchical Clustering Analysis (HCA)

Supervisionato (Classificazione e Predizione): Costruisce un modello conoscendo a priori una caratteristica dei campioni, per poi classificare nuovi campioni sconosciuti. "Esempi: Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN).



Costruzione di un modello di Machine Learning

Obiettivo: ottenere modelli robusti e affidabili.



1. Raccolta e Pre-elaborazione Dati



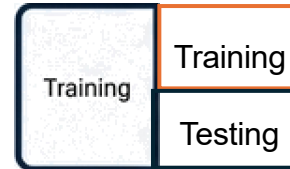
Pulizia del rumore, correzione della linea di base, normalizzazione per migliorare la qualità dei dati grezzi.

2. Estrazione/Selezione delle Caratteristiche (Features)



Riduzione della dimensionalità o selezione delle variabili più importanti per evitare l'overfitting e migliorare l'accuratezza.

3. Divisione del Dataset



Suddivisione dei dati in set di addestramento e di test per valutare la capacità di generalizzazione del modello.

4. Addestramento del Modello e Ottimizzazione



Selezione dell'algoritmo, addestramento sul training set e ottimizzazione degli iperparametri tramite cross-validazione

5. Valutazione del Modello



Misurazione delle performance sul test set utilizzando metriche come accuratezza, precisione, e la matrice di confusione.

Le applicazioni del Machine Learning alla Foodomica



Previsione del Sapore

Prevedere gli attributi sensoriali (dolcezza, aroma) dai dati metabolomici.



Autenticazione e Antifrode

Distinguere prodotti autentici da quelli adulterati.



Valutazione della Qualità

Monitorare i cambiamenti molecolari per predire la freschezza.



Nutrizione Personalizzata

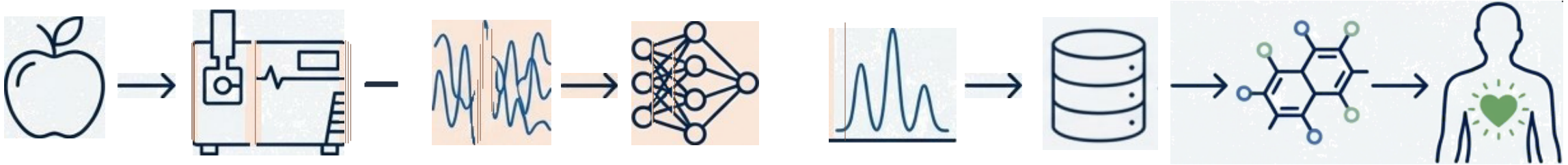
Prevedere la risposta metabolica individuale agli alimenti.



Sicurezza Alimentare

Identificare contaminanti o marker di deterioramento.

Foodomics Workflow



1. Misurazione:

Acquisizione di dati grezzi da campioni alimentari o biologici tramite piattaforme analitiche avanzate es. LC-MS

2. Pre-elaborazione e

Analisi Dati: Pulizia, normalizzazione e analisi statistica multivariata per estrarre le informazioni rilevanti. Qui intervengono i metodi di Machine Learning.

3. Identificazione dei

Metaboliti: Confronto con database consolidati per identificare i composti chimici di interesse.

4. Interpretazione Biologica:

Stabilire la connessione tra i metaboliti e gli aspetti nutrizionali e sanitari, analizzando gli effetti sulla salute umana.

Applicazioni della Foodomica e Machine Learning

Una volta decodificati, i dati alimentari permettono di affrontare sfide cruciali in tre aree principali:



Salute e Benessere

Comprendere come i composti bioattivi degli alimenti modulano i percorsi biologici per prevenire e trattare le malattie.



Autenticità e Tracciabilità

Garantire l'origine geografica e la composizione degli alimenti, combattendo le frodi e proteggendo i prodotti di qualità.



Scoperta e Innovazione

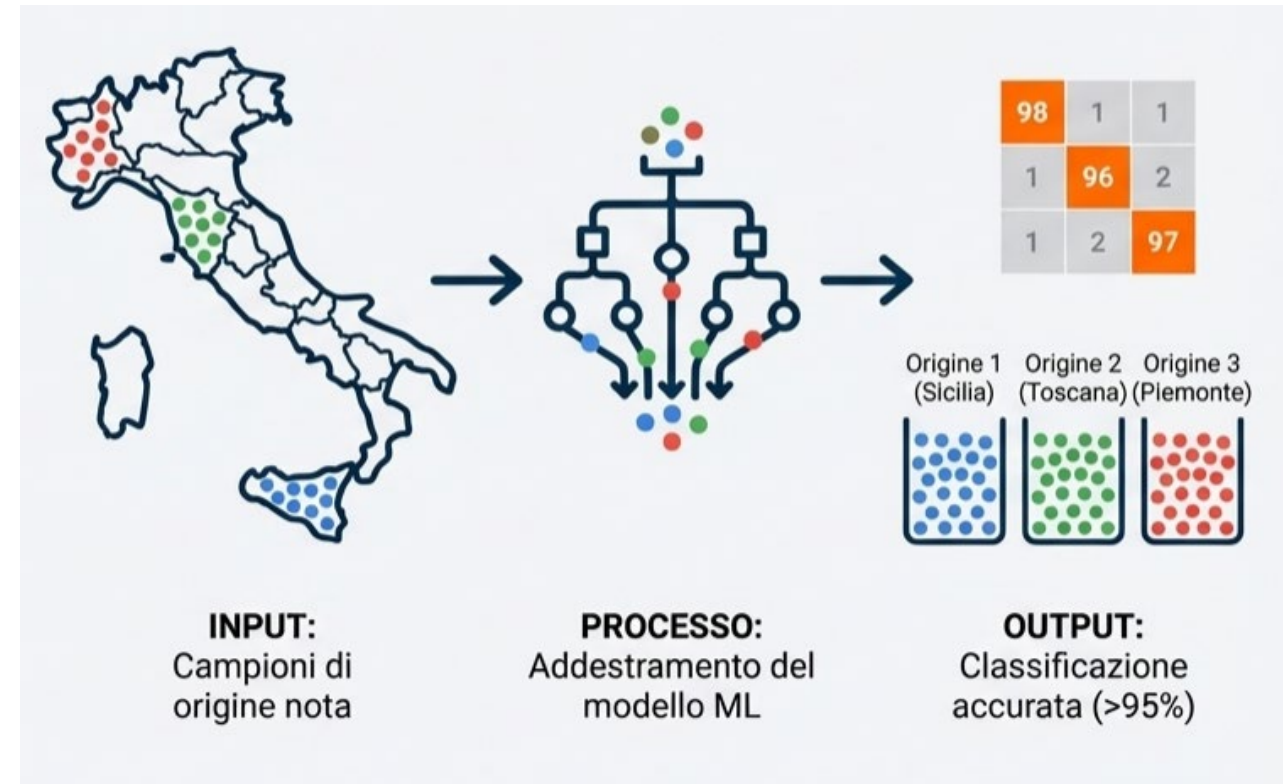
Identificare nuove molecole con potenziali benefici per la salute o nuovi marker per la qualità degli alimenti.

Autenticità: Il Machine Learning garantisce l'origine geografica degli alimenti.

La globalizzazione aumenta il rischio di frodi, con prodotti di qualità inferiore venduti come prodotti premium con indicazione geografica protetta (IGP/DOP).

Sfida: Come possiamo verificare con certezza l'origine geografica di un prodotto o l'autenticità di una cultivar? I metodi tradizionali (analisi genetiche, morfologiche) sono spesso lenti, costosi o insufficienti per catturare la complessa variabilità chimica influenzata da ambiente e pratiche agronomiche.

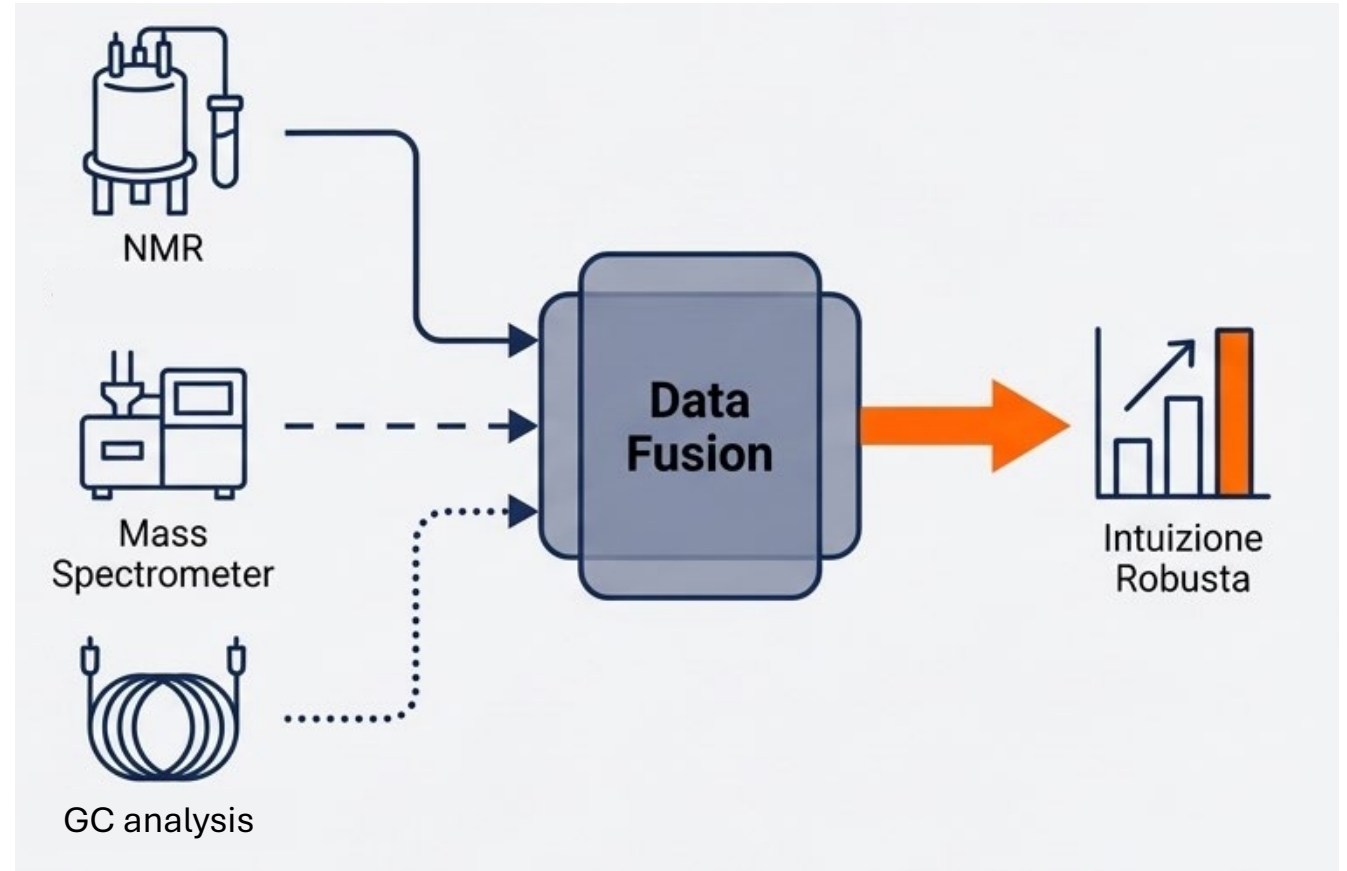
Approccio ML: Si acquisiscono 'impronte digitali' chimiche da campioni di origine nota. Algoritmi di classificazione (SVM, Random Forest) vengono addestrati per costruire modelli predittivi in grado di riconoscere l'origine di nuovi campioni.



Risultato: I modelli raggiungono accuratèzze di classificazione superiori al 95% per prodotti come miele, vino, e olio d'oliva, fornendo uno strumento potente per la tutela.

La fusione di dati multi-strumentale crea una visione più completa e accurata.

- Nessuna singola tecnica analitica può catturare l'intera complessità del Foodome. La **fusione di dati** da piattaforme complementari è una delle strategie più potenti.
- Strategie di Fusione Dati:
 - **Low-Level** Fusion: Combinazione dei dati grezzi prima della modellazione.
 - **Mid-Level** Fusion: Estrazione di feature da ogni dataset, e successiva fusione di queste feature per l'addestramento.
- Vantaggio: 1 modelli basati su dati fusi superano quasi sempre quelli basati su una singola fonte di dati, fornendo risultati **più** robusti.



Caso studio 1:

Non Tutte le Impronte Digitali Sono Uguali



L'efficacia della tracciabilità non dipende solo dalla completezza del profilo, ma anche dalla scelta dei marcatori chimici più informativi. Un'analisi ha messo a confronto due diversi profili chimici per classificare quattro cultivar di olive (*Olea europaea*).

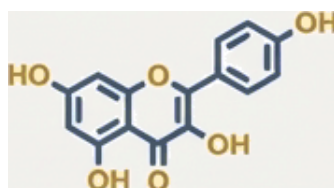
Caso di Studio: Classificazione delle cultivar di olive (Arbequina, Arbosana, Frantene, Koroneiki).

Le Due Impronte a Confronto:

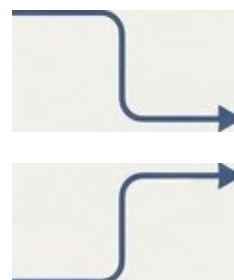
1. Profilo degli Acidi Grassi: Un marcatore tradizionale per la qualità dell'olio.
2. Profilo Fenolico: L'insieme completo di polifenoli e altri composti fenolici.



Profilo Acidi Grassi



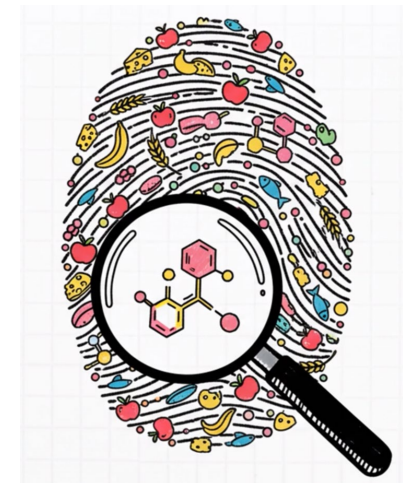
Profilo Fenolico



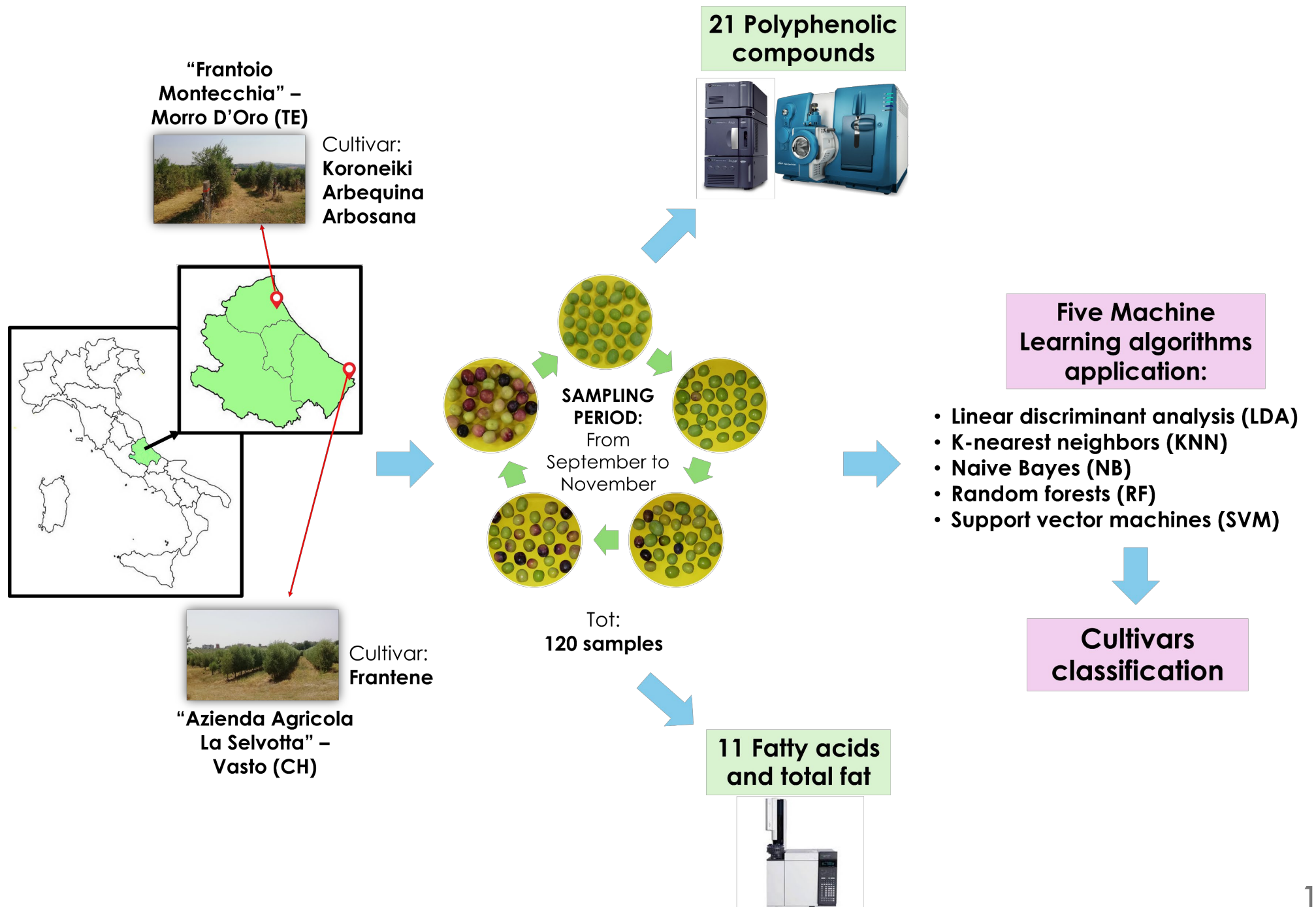
Modello di
Machine Learning



Classificazione



Workflow

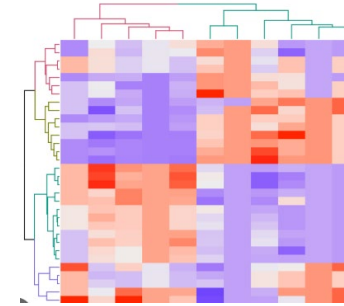


Analisi multivariata per la classificazione delle varietà di olive



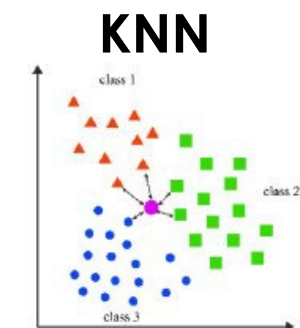
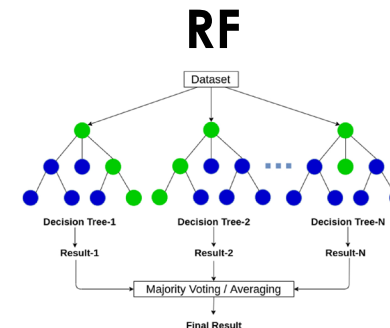
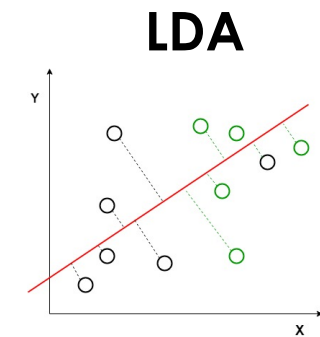
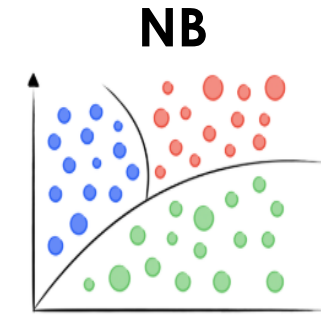
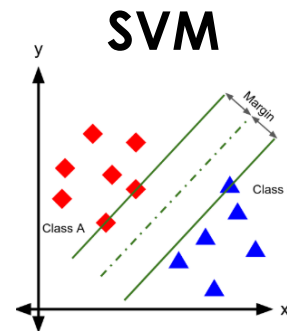
Approccio non supervisionato

**Clustering gerarchico
con Heatmap**

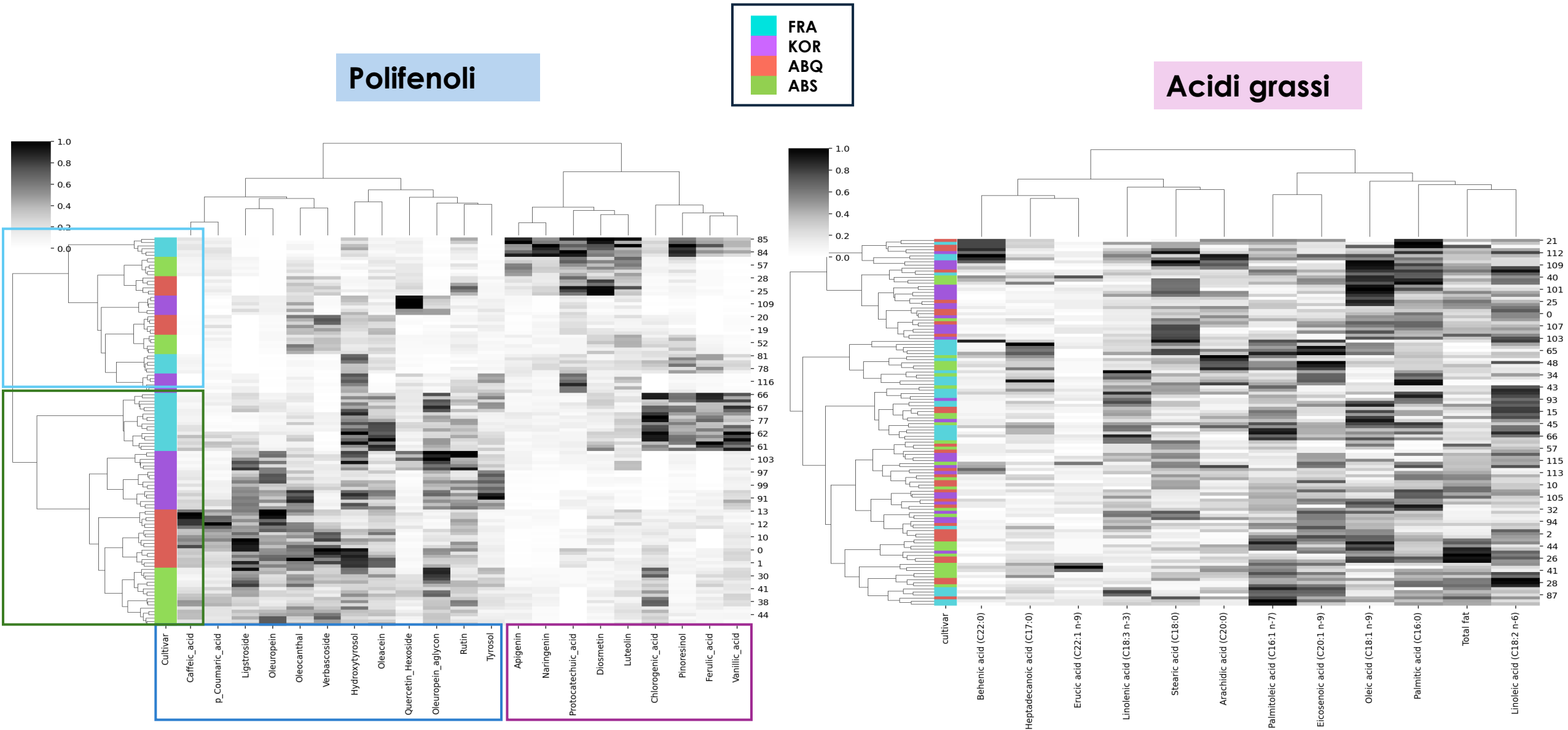


Approccio supervisionato

**Cinque modelli di
machine learning**

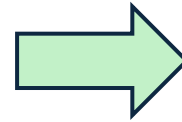
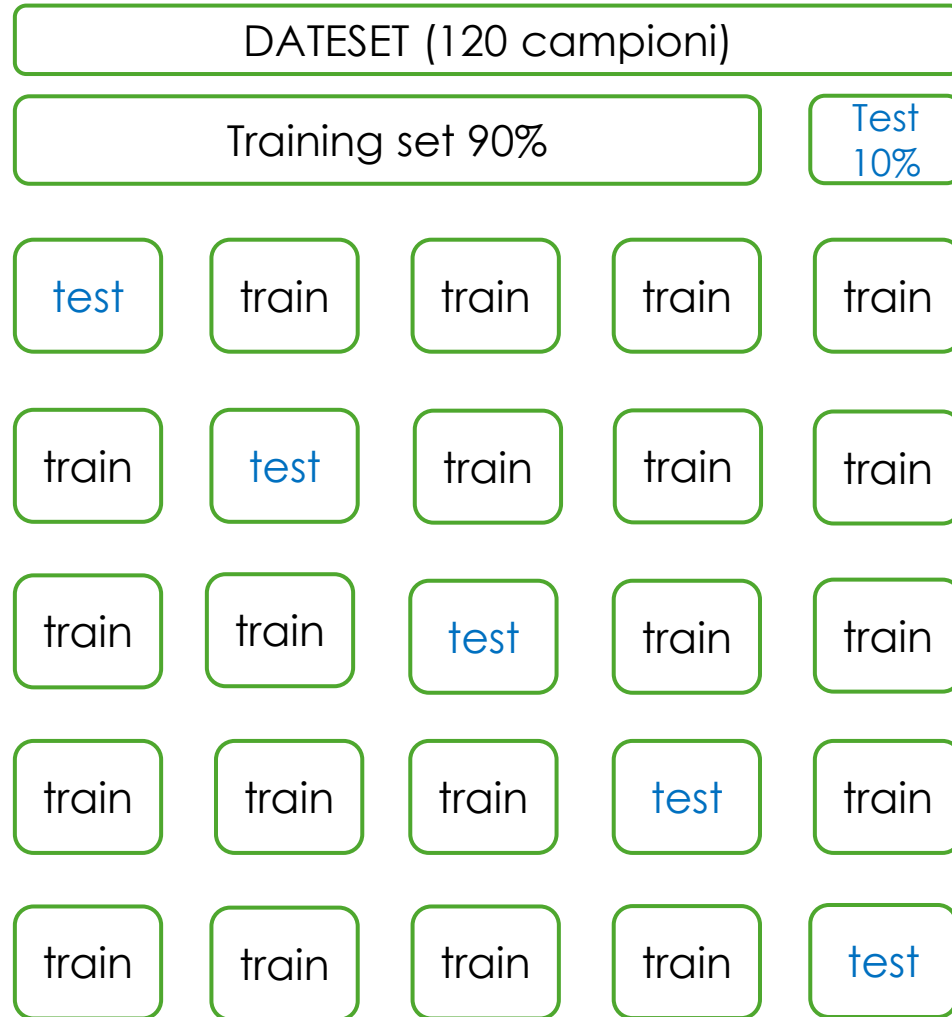


Approccio non supervisionato: clustering gerarchico con Heatmap

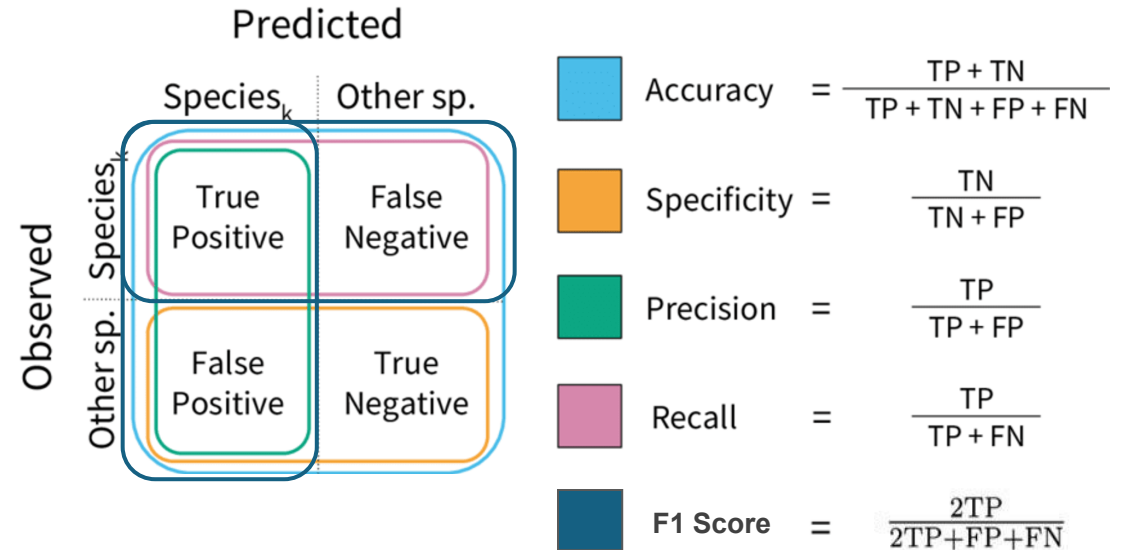


Approccio supervisionato: validazione e prestazioni dei modelli

Validazione Cross stratificata 10 volte



Matrice di confusione

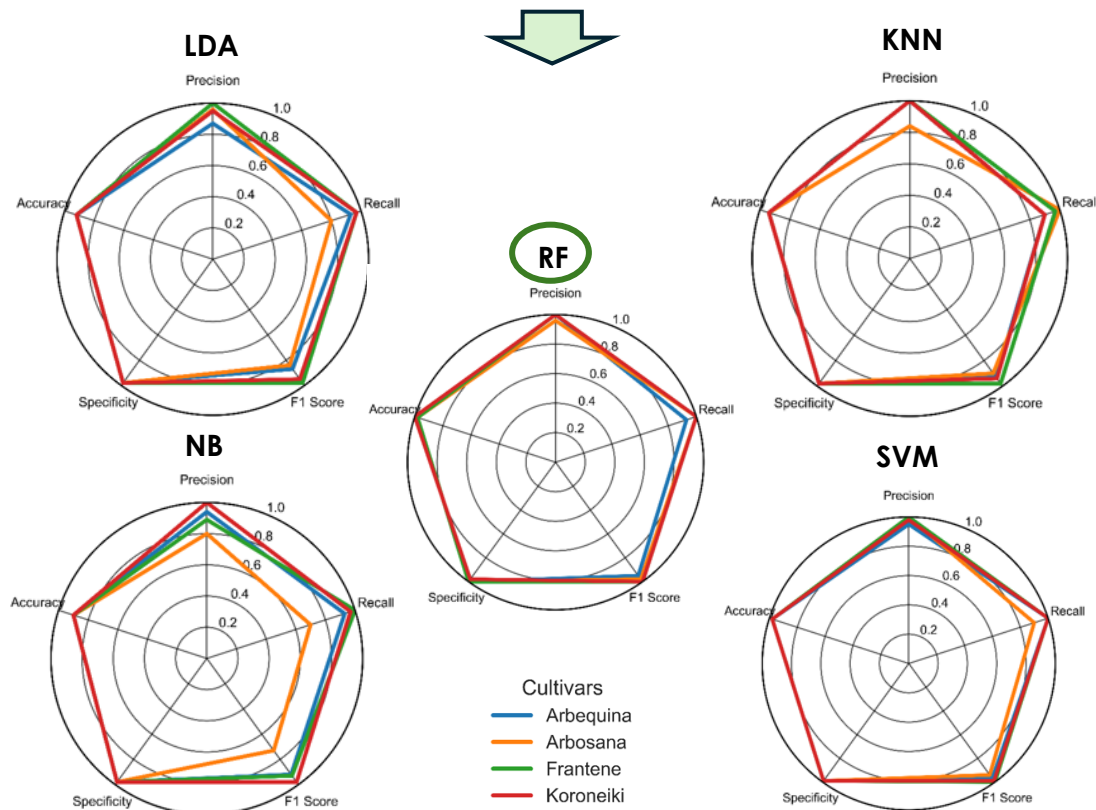


Risultati dei modelli di machine learning

	Polyphenols				
	LDA	KNN	NB	RF	SVM
Precision	0.95	0.96	0.91	0.99	0.98
Recall	0.92	0.94	0.9	0.98	0.98
F1 Score	0.91	0.94	0.89	0.98	0.97
Specificity	0.98	0.98	0.98	1	0.99
Accuracy	0.92	0.94	0.9	0.98	0.98

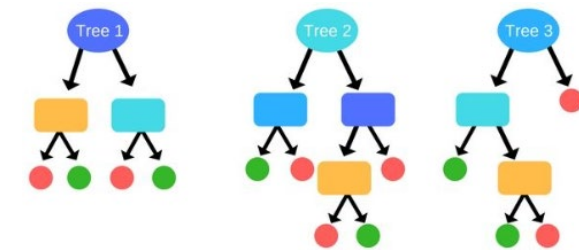
VS

	Fatty Acids				
	LDA	KNN	NB	RF	SVM
Precision	0.65	0.53	0.66	0.62	0.65
Recall	0.62	0.5	0.65	0.63	0.63
F1 Score	0.59	0.48	0.61	0.6	0.61
Specificity	0.91	0.89	0.91	0.91	0.91
Accuracy	0.62	0.5	0.65	0.63	0.63



Polyphenols	LDA	RF	SVM
Chlorogenic acid	12.8	63.7	31.8
Ferulic acid	15.6	9.9	10.4
Apigenin	9.2	7.0	3.7
Verbascoside	9.2	5.2	17.4
p-Coumaric acid	5.4	4.4	3.2
Pinoresinol	8.4	3.5	8.7
Ligstroside	0.3	2.0	0.5
Naringenin	1.3	1.5	0.7
Oleuropein	0.0	1.2	0.0
Protocatechuic acid	4.1	0.9	1.5
Tyrosol	8.1	0.6	2.5
Luteolin	1.3	0.3	1.2
Caffeic acid	1.0	0.0	3.0
Diosmetin	3.6	0.0	3.5
Hydroxytyrosol	1.0	0.0	1.5
Oleacein	1.0	0.0	0.7
Oleocanthal	1.6	0.0	1.5
Oleuropein aglycon	2.1	0.0	1.0
Quercetin-Hexoside	10.4	0.0	2.0
Rutin	0.7	0.0	4.2
Vanillic acid	2.8	0.0	1.0

L'analisi del modello di Machine Learning ha rivelato che il potere di classificazione non derivava dai composti fenolici più abbondanti, ma da quelli minori, spesso considerati secondari



Caso studio 2:

Food Chemistry: X 23 (2024) 101769



ELSEVIER

Contents lists available at ScienceDirect

Food Chemistry: X

journal homepage: www.sciencedirect.com/journal/food-chemistry-x



SCAN ME



Fingerprinting alkaloids for traceability: Semi-untargeted UHPLC-MS/MS approach in raw lupins as a case study

Fabiola Eugelio^a, Sara Palmieri^a, Marcello Mascini^a, Francesco Della Valle^a, Federico Fanti^{a,*}, Eleonora Oliva^a, Michele Del Carlo^a, Dario Compagnone^a, Manuel Sergi^b

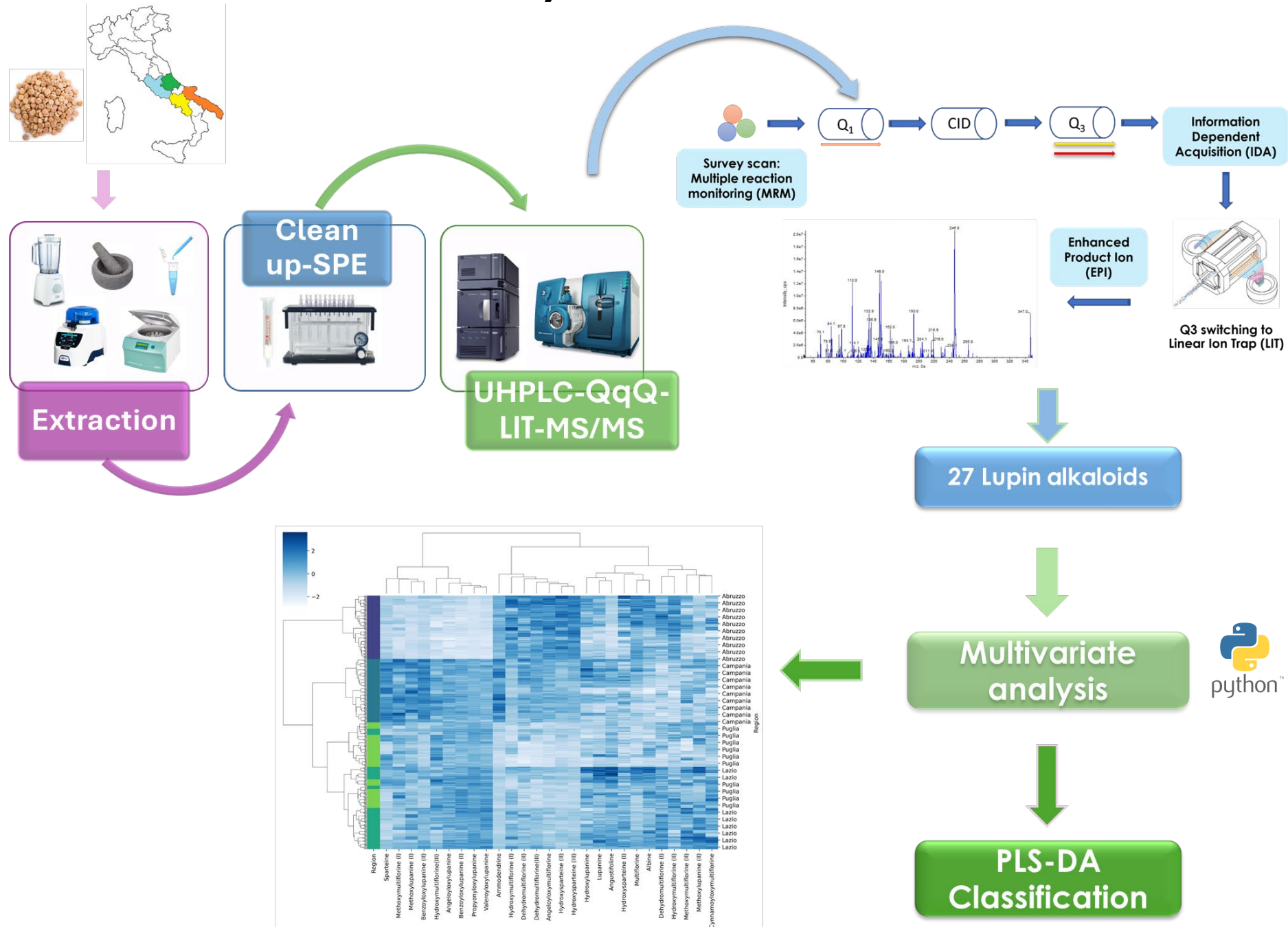
^a Department of Bioscience and Technologies for Food, Agriculture and Environment, University of Teramo, Via R. Balzarini, 1, 64100 Teramo, TE, Italy

^b Department of Chemistry, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy

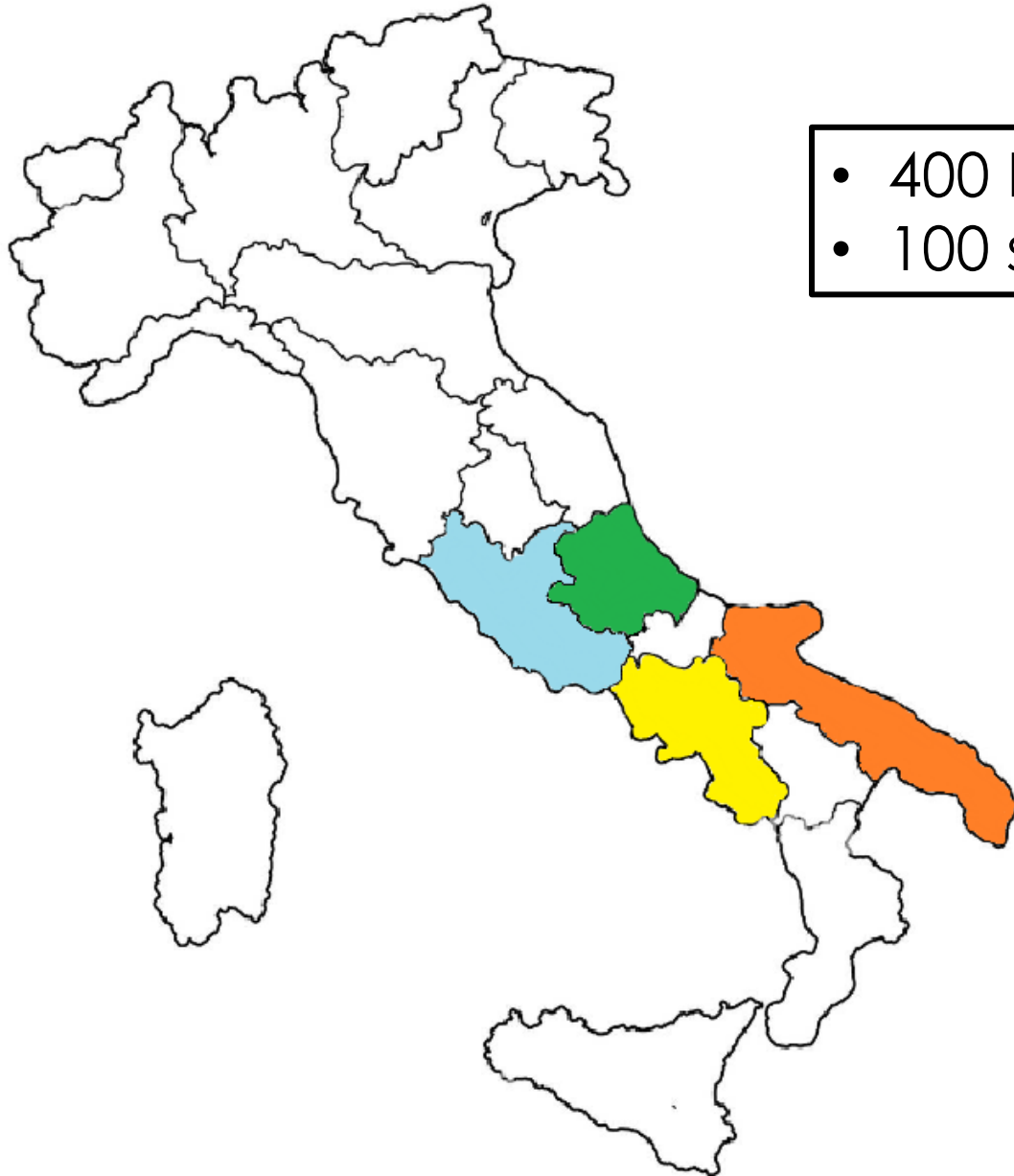
Confronto tra due approcci per classificare i lupini da quattro diverse regioni italiane basandosi sul loro profilo di **alcaloidi chinolizidinici** (conferiscono sapore amaro al lupino e sono tossici per la salute)



Analysis workflow



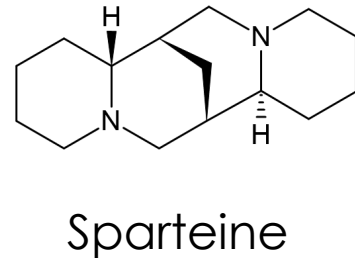
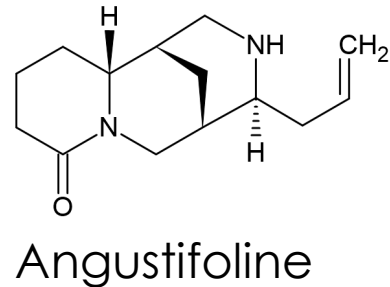
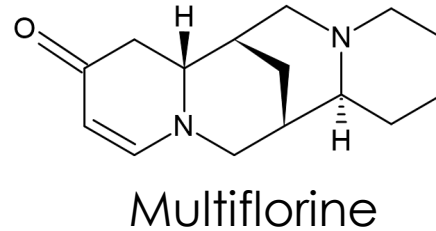
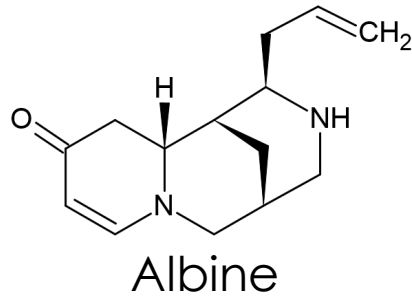
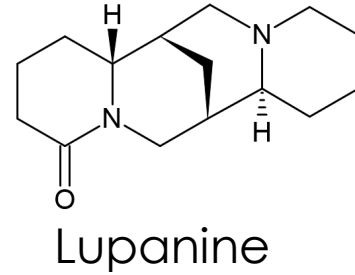
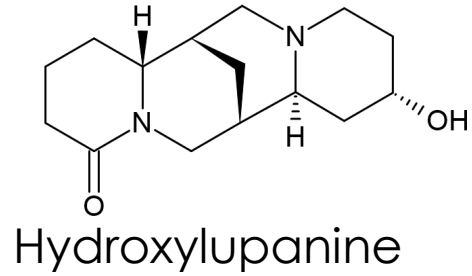
Application on *Lupinus albus* L. samples



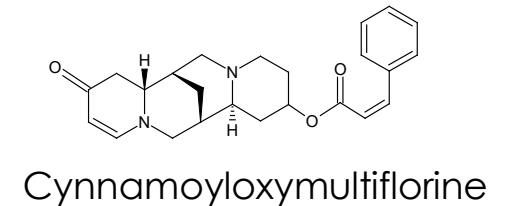
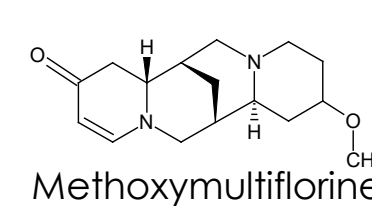
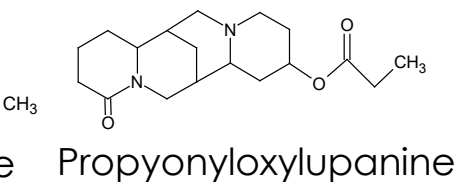
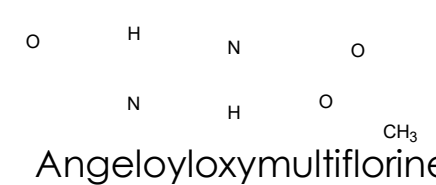
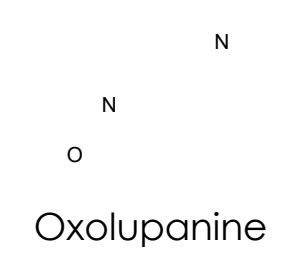
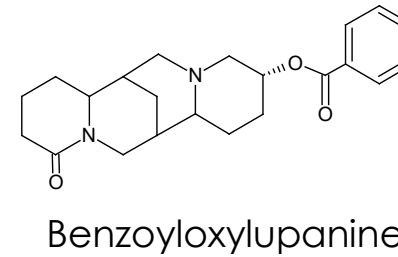
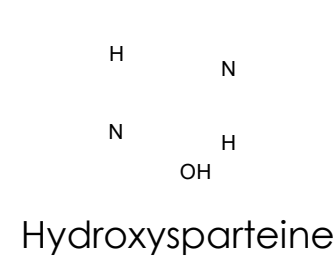
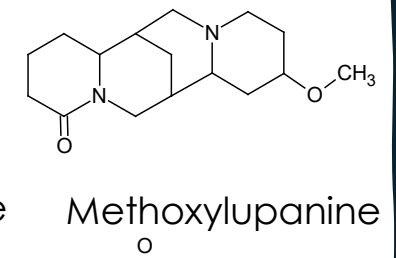
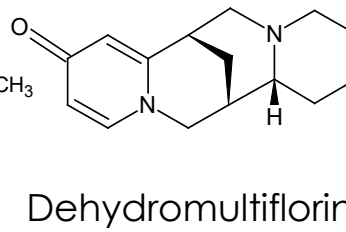
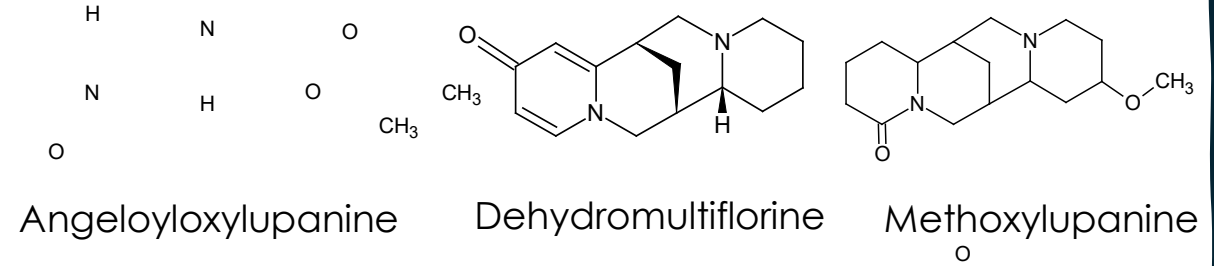
- 400 Raw *Lupinus albus* L. samples
- 100 samples from each highlighted region



Campioni di *Lupinus albus* L.: 27 alcaloidi identificati



6 QA identificati in **targeted**



Alcuni dei 21 QA identificati in **semi-untargeted**

Results: MRM-IDA-EPI

MRM



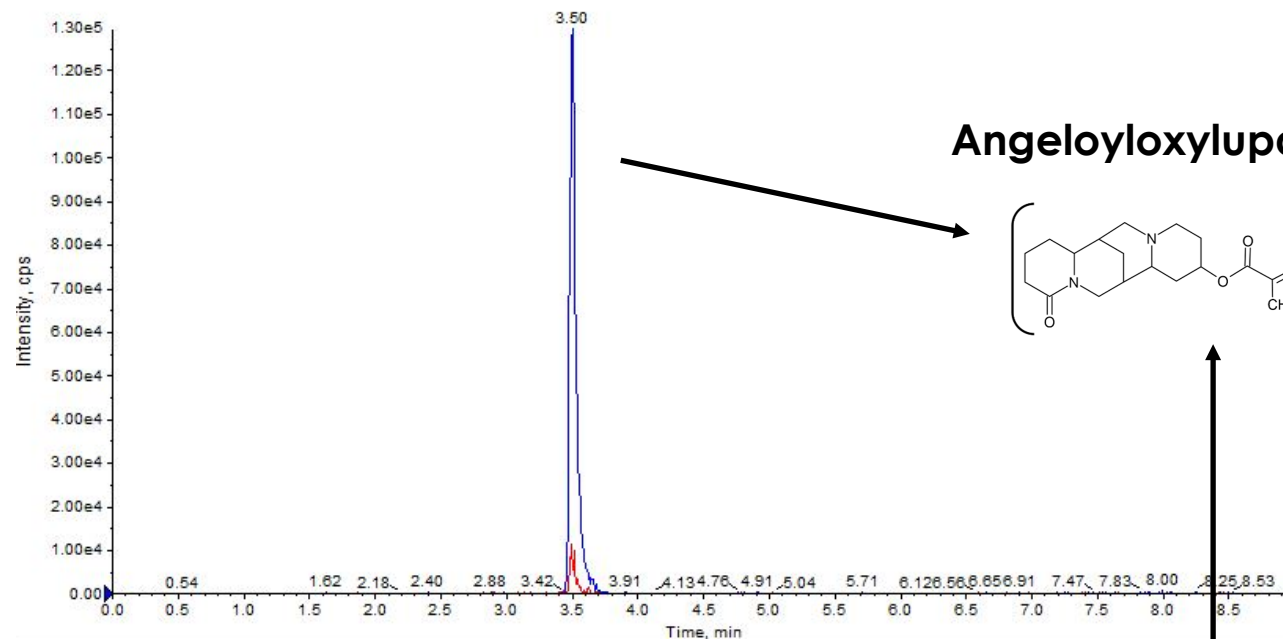
IDA



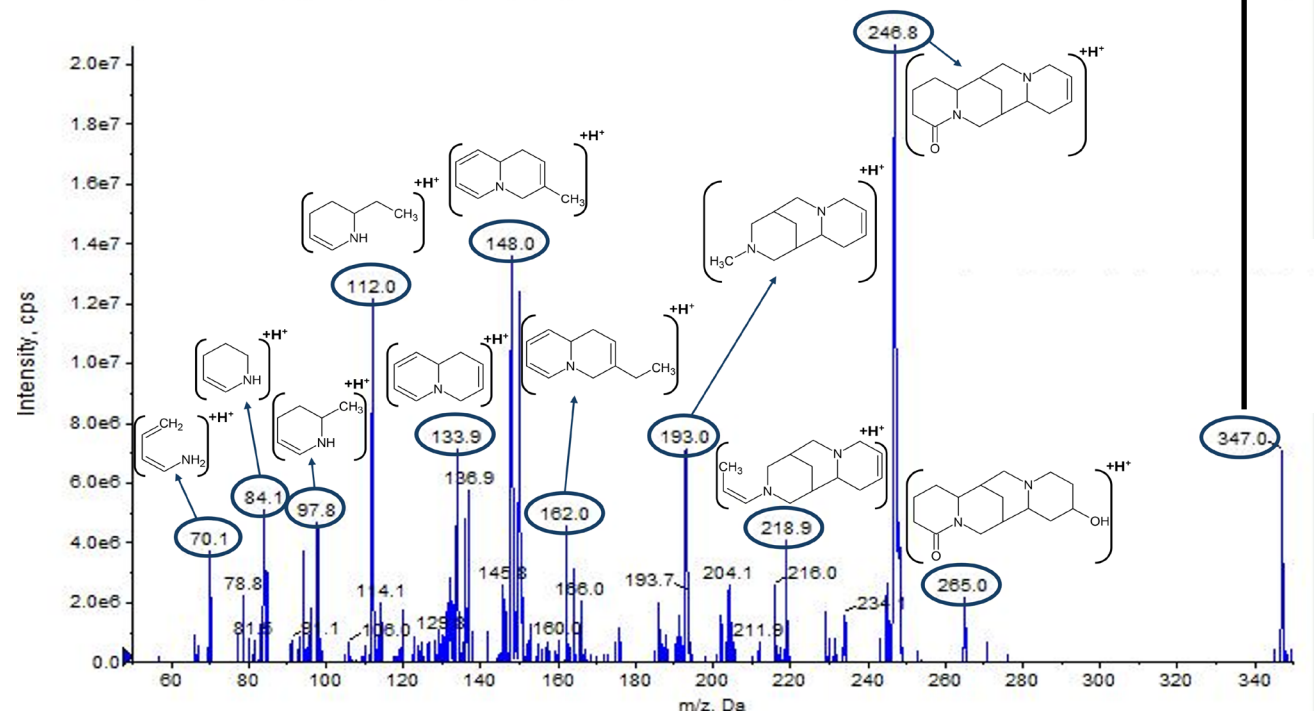
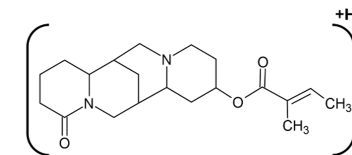
EPI



Angeloyloxylupanine
fragmentation pattern



Angeloyloxylupanine



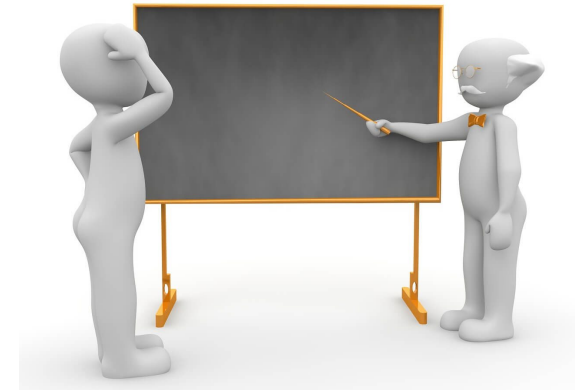
Analisi multivariata per la classificazione dell'origine dei campioni di lupino

Approccio non supervisionato



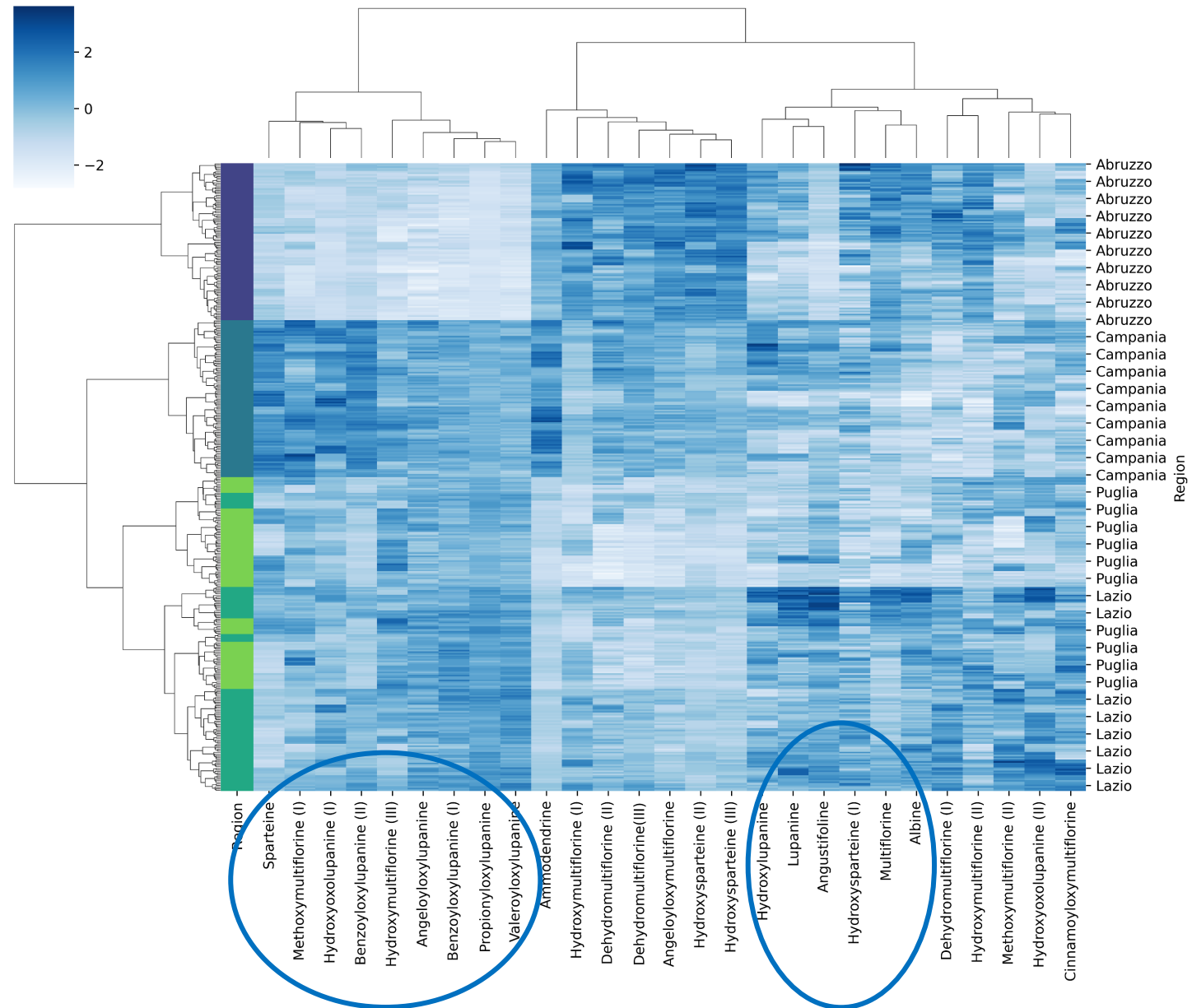
Clustering gerarchico con Heatmap

Approccio supervisionato



PLS-DA Cross-Validazione

Approccio non supervisionato: clustering gerarchico con Heatmap



Supervised approach: PLS-DA Cross-Validation (10-Fold Summary)

Target: 6 alkaloids

	Abruzzo	Campania	Lazio	Puglia
Sensitivity	1.00	0.88	0.65	0.41
Specificity	0.91	0.91	0.90	0.93
Pos Pred Value	0.78	0.77	0.68	0.66
Neg Pred Value	1.00	0.96	0.88	0.83
Balanced Accuracy	0.95	0.90	0.77	0.67

Semi-Untargeted: 27 alkaloids

	Abruzzo	Campania	Lazio	Puglia
Sensitivity	1.00	0.96	0.91	0.97
Specificity	1.00	1.00	0.99	0.96
Pos Pred Value	1.00	1.00	0.96	0.90
Neg Pred Value	1.00	0.99	0.97	0.99
Balanced Accuracy	1.00	0.98	0.95	0.97

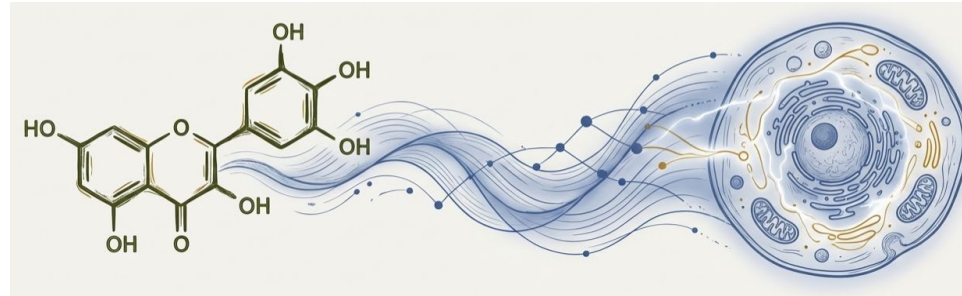
Contributo sinergico degli alcaloidi nella classificazione

	Abruzzo	Campania	Lazio	Puglia
Benzoyloxylupanine (II)	100	100	100	96.66
Ammodendrine	97.56	100	100	100
Dehydromultiflorine (III)	87.16	100	74.92	100
Valeroyloxylupanine	100	100	100	100
Angeloyloxymultiflorine	99.26	100	87.4	100
Propyonyloxylupanine	100	100	100	100
Hydroxy-Oxolupanine (I)	100	100	100	98.61
Hydroxysparteine (III)	100	100	99.42	100
Benzoyloxylupanine (I)	100	100	100	100
Sparteine	99.98	99.98	99.98	23.16
Angeloyloxylupanine	99.96	99.82	99.82	99.96
Hydroxysparteine (II)	98.03	99.73	88.65	99.73
Hydroxymultiflorine (III)	98.14	98.34	98.14	98.34
Methoxymultiflorine (I)	98.09	98.09	98.09	84.92
Hydroxymultiflorine (II)	97.56	97.56	97.56	72.47
Angustifoline	96.93	85.68	77.92	96.93
Dehydromultiflorine (I)	85.14	85.14	94.71	32.82
Hydroxymultiflorine (I)	89.02	92.74	89.02	92.74
Dehydromultiflorine (II)	60.08	86.73	42.17	86.73
Multiflorine	71.62	85.36	71.62	85.36

L'impronta digitale più completa ha fornito un potere di classificazione drasticamente superiore.

Altre prospettive: Nutrizione e salute

Comprendere l'impronta molecolare del cibo è estremamente potente, ma la vera rivoluzione nasce quando questa impronta viene collegata ai suoi effetti sull'organismo umano. Analizzare come i diversi composti alimentari interagiscono con la nostra biologia a livello molecolare permette di chiarire i meccanismi con cui il cibo contribuisce alla prevenzione delle malattie.



Obiettivo finale: Nutrizione di Precisione, ovvero creare modelli predittivi per gestire e prevenire le malattie, in cui la dieta diventa efficace quanto una terapia medica, personalizzata sulle caratteristiche biologiche, cliniche e di stile di vita dell'individuo.

