

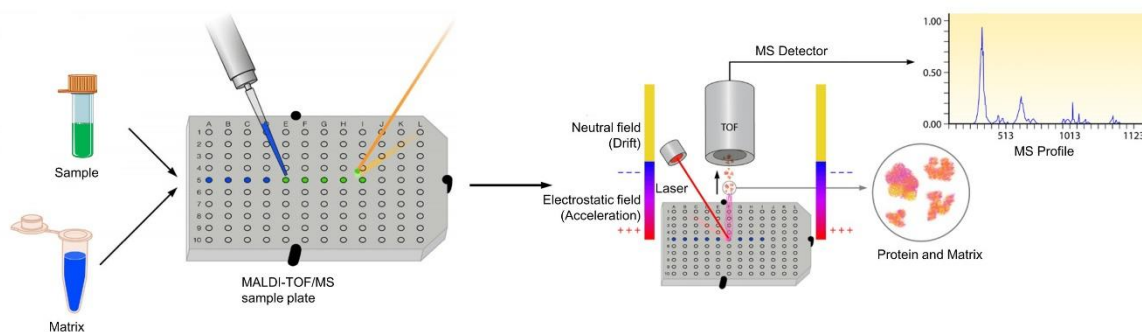
Антимикробна резистентност и видово определяне на микробни патогени посредством алгоритми за изкуствен интелект

Научен обзор

Антимикробната резистентност (AMP) е способността на микроорганизмите да устояват на антимикробното лечение, което представлява глобално предизвикателство, заплашващо ефективната превенция и лечение на инфекциозните заболявания и увеличаващо обхвата на инфекциите, причинени от бактерии и други патогенни микроорганизми. **В клиничната практика бързото и надеждно идентифициране на потенциални патогени е от изключително значение за своевременното започване на подходящо антимикробно лечение.** През последните десетилетия MALDI-TOF MS (*matrix assisted laser desorption ionization-time of flight mass spectrometry*) и краткото време на изпълнение и резултат революционизира клиничната диагностика със способността си за **бързо, надеждно и рентабилно видово идентифициране на катиони, както и с потенциала си за рационализиране на тестването за антимикробна чувствителност.** Многобройни проучвания показват, че MALDI-TOF MS надминава конвенционалните диагностични методи от гледна точка на цена, скорост и точност при определяне на микробиалните видове. Конвенционалните фенотипни диагностични методи, като антимикробен градиент и диск дифузионен тест, изискват до 48 часа и струват около 100 долара, докато MALDI-TOF MS може да предостави информация за патогена в рамките на минути от 24 часова култура на цената на няколко долара.

MALDI-TOF MS е революционизирала микробиологията чрез улесняване на процеса по прецизна и бърза видова идентификация.

Фиг. 1: Принцип на процеса на MALDI-TOF MS



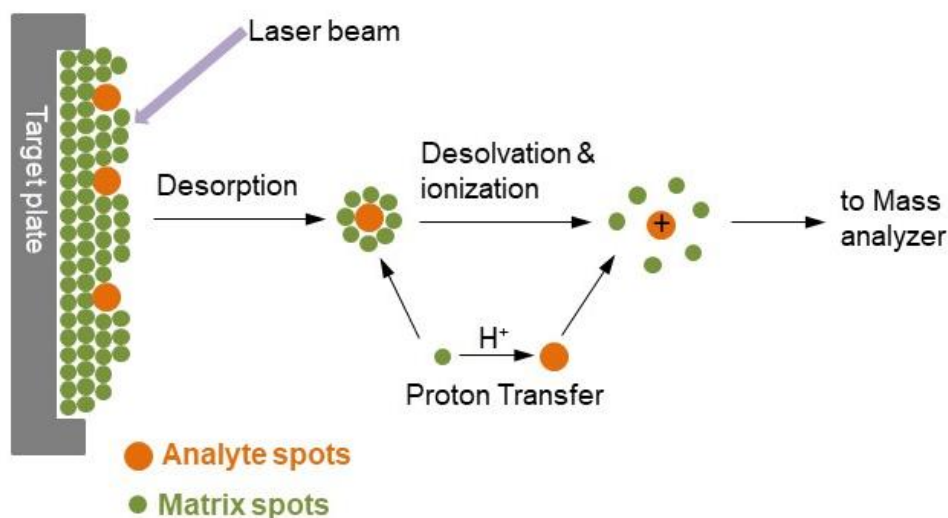
Напоследък техниките и алгоритмите за изкуствен интелект (*machine learning*) се използват все по-често, за да се използва максимално информацията, съдържаща се в MALDI-TOF мас спектрите, с крайната цел да се прецизира идентификацията на видовете и да се рационализира определянето на антимикробната резистентност.

В PubMed/Medline, Scopus и Web of Science литературата са налични доста проучвания по темата за внедряване на изкуствения интелект в молекулярно диагностичните методи и приложенията на тези иновативни компютърни науки в анализа на данните от MALDI-TOF, прилагани в микробиологията и определянето на антимикробната чувствителност.

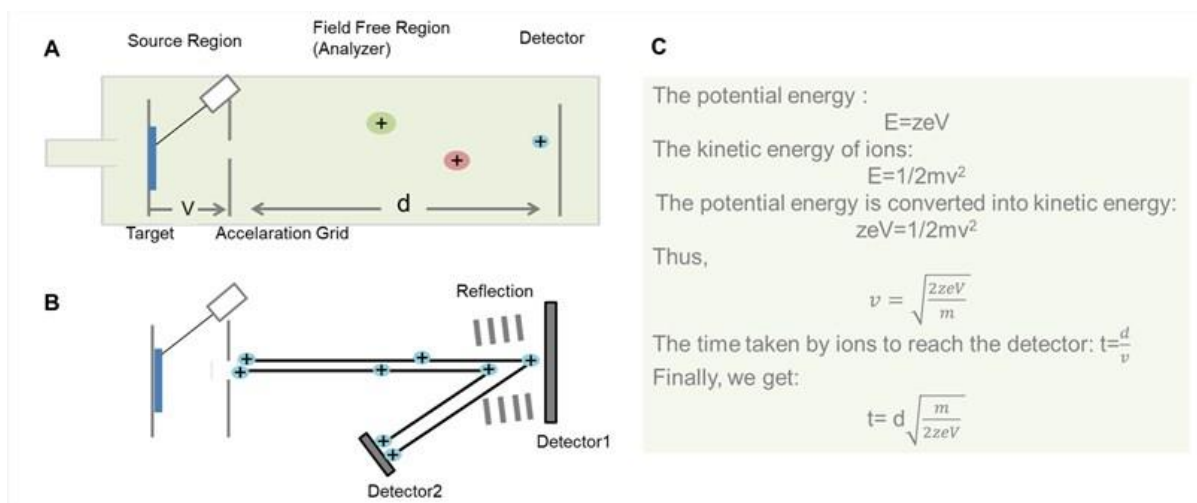
В част от проучванията е извършен систематичен преглед съгласно насоките на PRISMA (*Preferred Reporting Items for Systematic reviews and Meta-Analysis*) и е извършена оценка на качеството на моделите на изкуствен интелект. От 36-те намерени проучвания, 27 използват изкуствен интелект за видово определяне и девет за тестване за антимикробна чувствителност. *Vector Machines*, *Genetic Algorithms*, *Artificial Neural Networks* и *Quick Classifiers* са най-често използваните алгоритми на изкуствения интелект.

Все по-голям брой проучвания използват изкуствения интелект, за да оптимизират анализа на мас спектри от MALDI-TOF. Има обаче някои недостатъци на настоящите подходи, подкрепяни от изкуствения интелект, които трябва да бъдат разрешени, за да станат широко достъпни и включени в клиничната практика.

Обикновено анализът на мас спектрите от MALDI-TOF разчита на малък брой набор от показатели като височина на пика и площ на пика, които са емпирично свързани с микробните видове.



Фиг. 2: Принцип на действие на MALDI-TOF масспектрометрията



Фиг. 3: Схематично представен Time of Flight (TOF) анализатора

Приложения на MALDI-TOF маспектрометрия:

Определяне на интактна маса (*Intact Mass determination*)

Определянето на интактната маса е основно и важно за характеризиране на протеините, тъй като правилното молекулно тегло на протеина може да покаже структурата му. MALDI техниката е подходяща за протеини, които са крехки и фрагментирани, когато се подлагат на йонизация чрез други йонизационни методи. Производителността на MALDI-TOF MS се влияе по-малко от буферните компоненти, детергентите и замърсителите. В допълнение, той позволява определяне на масата на протеина с достатъчна точност (≤ 500 ppm) за валидиране на последователността. MALDI-TOF MS може също да се използва за анализ на получените пептиди за по-нататъшно потвърждаване на първичната им структура чрез *peptide mass fingerprinting*.

Peptide mass fingerprinting (PMF)

MALDI-TOF има прост работен процес, добра точност, както и висока разделителна способност и чувствителност. Следователно, той има **широко приложение в протеомиката за идентифициране на протеини от прости смеси** чрез метод, наречен PMF, който често се използва с двуизмерна гел електрофореза (2-DE). При този подход пептидите се генерират чрез разграждане на целеви протеини със специфичен за последователността ензим като трипсин. След това пептидите се анализират чрез MALDI-TOF маспектрометрия, за да се получат пептидните маси. Експерименталните резултати се сравняват с база данни, съдържаща теоретични данни за маспектри от даден организъм със същата протеаза, специфична за последователността.

MALDI-TOF анализ след разпадане на източника (PSD) (*Post source decay (PSD) MALDI-TOF analysis*)

MALDI-TOF маспектрометри, оборудвани с рефлексори, **могат да анализират фрагментирани йони**, произведени от прекурсори, които спонтанно се разпадат. Такива йони обикновено се наричат метастабилни йони, а процесът на разпадане в

зоната между източника на йони и рефлектора обикновено се нарича PSD. Фрагментните йони на PSD се образуват преди да влязат в рефлектора. PSD фрагментираните йони могат да бъдат разделени, събрани и записани на детектора чрез непрекъсната промяна на напрежението на рефлектора, за да се образува PSD маспектър, който предоставя **много богата и ефективна структурна информация за първичната структура на пептидите и протеините**. Функцията за секвениране на PSD може да се приложи за **идентифициране на тези протеини**. Използвайки PSD спектроскопия, комбинирано с търсене в база данни, протеините могат да бъдат идентифицирани бързо и с висока специфичност.

Анализ на олигонуклеотиди

С развитието на техниките на молекулярната биология и технологиите за антисенс лекарства се синтезират все повече и повече олигонуклеотидни фрагменти, които да се използват като праймери, сонди и антисенс лекарства. Напълно необходимо е бързо да се открият тези фрагменти, за да се определи дали синтезът е завършен и дали синтезираната последователност е правилна. Масспектрометрията, включително MALDI-TOF-MS, е най-доброто средство за това. Олигонуклеотидният анализ с помощта на MALDI-TOF-MS е прост, бърз, точен и чувствителен, който **може да се използва за определяне на пълната олигонуклеотидна последователност**.

MALDI изображения (MALDI imaging)

MALDI-TOF може да се използва за **профилиране и изобразяване на протеини директно от тънки тъкани срезове, известни като MALDI образна масспектрометрия (MALDI-IMS)**. Този метод **предоставя специфична информация за локалния молекулен състав, относителното изобилие и пространственото разпределение на пептидите и протеините в анализирания участък**. MALDI-IMS **може да анализира множество неизвестни съединения** в биологични тъкани участъци едновременно чрез едно измерване, като така може да се получи молекулно изображение на тъканта, като същевременно се поддържа целостта на клетките и молекулите в тъканите.

MALDI-TOF мас спектрометрията **може да анализира голямо разнообразие от биомолекули, като пептиди, протеини, въглехидрати, олигонуклеотиди и т.н.** Поради факта, че образуваните йони имат ниска вътрешна енергия, голямо предимство на MALDI-TOF е, че процесът на мека йонизация **позволява наблюдение на йонизирани молекули от не- или малко фрагментирани аналити**, което позволява молекулярното идентифициране на йони на аналитите, дори в смеси. Лесен е за използване и поддръжка, с **бързо събиране на данни**. Изборът на подходящо матрично вещество е важен за успешната MALDI-TOF масспектрометрия.

Въпреки че MALDI-TOF е валидиран подход и работи доста добре на ниво определяне на видове микроорганизми, има **изобилие от информация, съдържаща се в тези спектри, която остава неизползвана**. За да използват пълноценно информацията, съдържаща се в мас спектрите от MALDI-TOF, изследователите прилагат алгоритми за изкуствен интелект в усилията си за **повторно определяне на идентичността на видовете и рационализиране на определянето на антимикробната резистентност**. Методите за изкуствен интелект са в състояние да

определят статистическите зависимости в данните, като се вземат предвид и нелинейните ефекти и ефектите на взаимодействие между характеристиките. По този начин техниките за изкуствен интелект могат да разкрият нова или непозната информация, която е вградена в мас спектрите MALDI-TOF. Тази информация се оказва полезна за определяне/охарактеризиране и диференциация на видовете, особено тези, които са филогенетично близки, както и подвидовете. Освен това наскоро бе признато, че информацията, съдържаща се в MALDI-TOF масспектрите, също може да помогне за профилиране и определяне на антибиотичната резистентност. По-специално, техниката MALDI-TOF се основава предимно на *fingerprinting* анализи на рибозомни протеини, които са подходящи за идентифициране на бактерии и систематичната взаимовръзка помежду им. Ако приемем, че характеристиките на AMP при тясно свързани щамове са сходни, MALDI-TOF спектрите на бактериите с известна антимикуробна резистентност служат като отправна точка за определяне на AMP профилите на бъдещите анализи.

От всички проучвания по темата за приложенията на изкуствения интелект при анализа на масспектрите бе извлечена следната информация: а) характеристики на публикацията (автор, време на публикуване), б) цели на проучването (диференциация на видовете, идентичност или изпитване за антимикуробна чувствителност), в) подбор (вид, род, размер на извадката), г) спецификации на използвания инструмент MALDI-TOF, д) избор на модел (приложен алгоритъм за изкуствен интелект, платформи (софтуер) и пакети, параметри), е) статистически данни за модел по форма (методи за оценка на модела, средна стойност и мярка на дисперсията) и ж) методи за избягване на *overfitting*¹, както и външни стратегии за валидиране.

Въз основа на девет критерия (адаптирани от *Qiao* и екип), бе оценено качеството на проучванията за изкуствен интелект. Оценката на качеството включва следните категории: неудовлетворени нужди (*rationale for Machine Learning algorithm*), възпроизводимост (инженеринг на функции, софтуер и хардуер, хиперпараметри), здравина (валидни методи за преодоляване на *overfitting*, стабилност на резултатите), обобщеност (заверяване на външни данни) и клинична значимост (тълкуване на предикторите и предполагаема клинична употреба). Общо 36 публикувани проучвания отговарят на критериите: 27 за идентификация на микробни видове и девет за изследване за чувствителност към антибиотици. По-голямата част от изключените проучвания (n = 924) не отговарят на един или повече критерии, като например приложения анализ на данни (напр. статистически подходи, включително линейна и логистична регресия), тема на научните изследвания отвъд този научен обзор и структура на проучването (напр. доклади за случаи и прегледи).

Изследвани видове микроорганизми и антимикуробни средства

Широк спектър от микробни родове са изследвани в прегледаните проучвания. Най-често изследваните родове са *Staphylococcus* (n = 14), *Streptococcus* (n = 6), *Escherichia* (n = 4), *Bacillus* (n = 4) и *Klebsiella* (n = 3). По отношение на тестването за чувствителност към антимикуробни средства, са включени проучвания, насочени към широкоспектърните антибиотици ванкомицин (n = 3) и карбапенеми (n = 1), тясноспектърния антибиотик метицилин (n = 3) и противогъбичното лекарство

¹ <https://elitedatascience.com/overfitting-in-machine-learning>

флуконазол ($n = 1$). Броят на анализирани изолати варира значително в различните проучвания. Wang et al. включва 787 изолата в анализа си, докато други са имали по-малко от 50 изолата.

Преглед на алгоритмите и софтуера

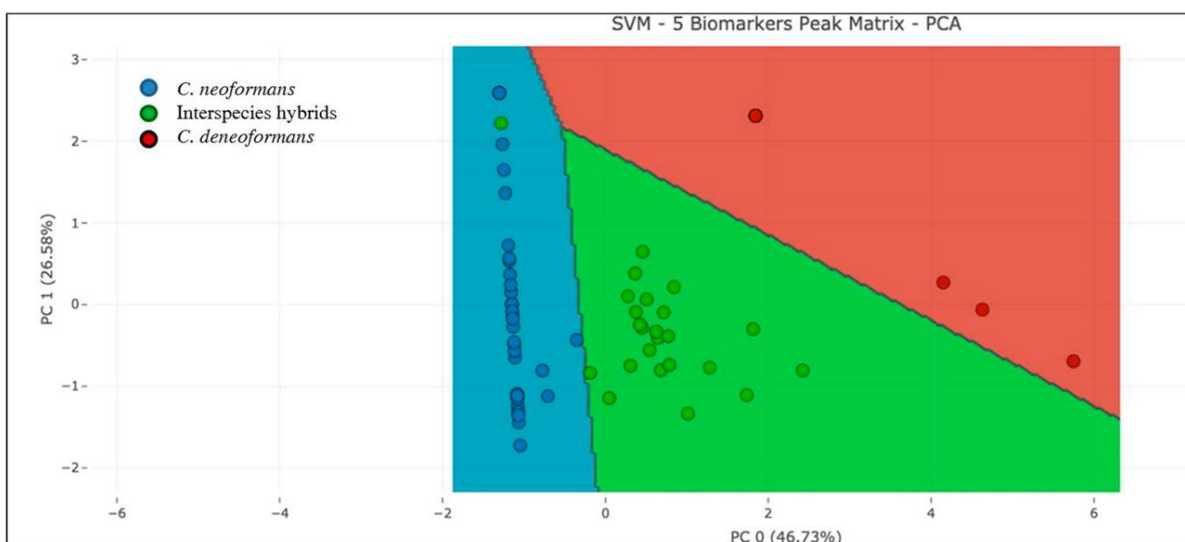
Както е подчертано в таблица 1б, най-широко използваните алгоритми за изкуствен интелект са: **поддържащите векторни машини (SVM, 18 проучвания), генетичните алгоритми (GA, 15 проучвания), невронни мрежи (Artificial/Supervised Neural Networks) (ANN, 13 проучвания) и Quick Classifiers (QC, 11 проучвания).** Подробно описание на тези четири алгоритъма може да бъде намерено в таблица 1а. По-рядко избраните алгоритми включват: *clustering*/йерархичен клъстерен анализ (*UHCA*), *Random Forest (RF)*, *дървета за вземане на решения (DT)*, *k-Nearest Neighbors (kNN)*, *Multiple Logistic Regression (MLR)*, *Naïve Bayes* и *Aristotle Classifiers*. Голям брой от прегледаните проучвания ($n = 17$) използват *ClinProTools* софтуер (т.е. *Brucker Daltonik*) за извършване на анализ на маспектрите MALDI-TOF. Останалите проучвания анализират данните си с *R* или *R Studio* ($n = 9$), *MATLAB* ($n = 7$), *Python* ($n = 1$), *MALDI Biotools 3.0* ($n = 1$), Статистическа програма за социални науки ($n = 1$), *Mathematica* ($n = 1$) или комбинация от тях.

*SVM - Support Vector Machines*²

Support Vector Machines е нов алгоритъм за изучаване на проблеми с класификация в две групи. Алгоритъмът концептуално реализира следната идея: входните вектори са нелинейно картографираны с много голям набор от характеристики. Така от тези характеристики се изгражда линейна зависимост за вземане на решения. Специалните характеристики за вземане на решения осигуряват висока обобщаваща способност на обучаваната машина. Показана е висока способност за обобщаване на мрежите с опорни вектори, използващи полиномни входни трансформации. Също така се сравняват производителността на мрежата на опорния вектор с различни класически алгоритми за обучение, като всички са включени в сравнително проучване на оптичното разпознаване на символи.

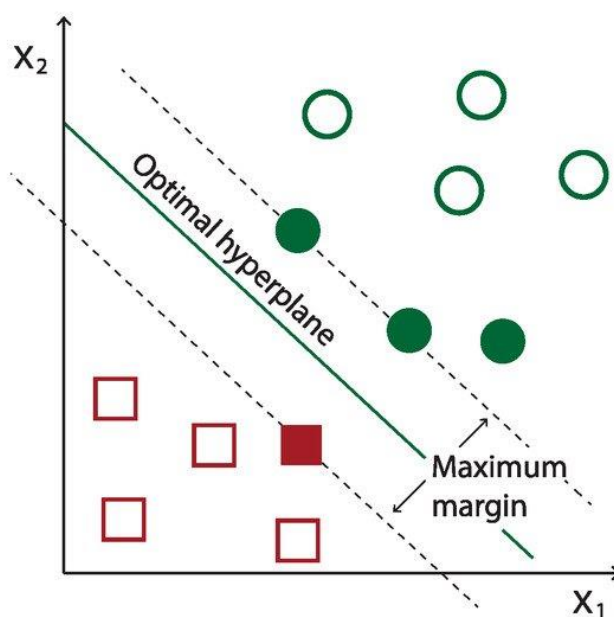
Пример за употребата на този вид алгоритми от изкуствения интелект при определяне и диференциране на бактериални видове може да бъде намерено в следното проучване: „Implementation of MALDI-TOF Mass Spectrometry and Peak Analysis: Application to the Discrimination of *Cryptococcus neoformans* Species Complex and their Interspecies Hybrids“ на Manuel J. Arroyo Pulgar - <https://www.mdpi.com/2309-608X/6/4/330/htm>

² <https://link.springer.com/content/pdf/10.1007/BF00994018.pdf>



Фиг. 4: Изглед на анализирани посредством изкуствен интелект данни от маспектроскопия и разпределяне на подвидовете микроорганизми

Фиг. 5: Принцип на действие на Support Vector Machines



Genetic Algorithms - Генетични алгоритми³

Генетичният алгоритъм е базиран на търсене алгоритъм, използван за решаване на оптимизационни проблеми в изкуствения интелект. Този алгоритъм е важен, защото решава трудни проблеми, чието разрешаване би отнело много време. Използва се в различни приложения от реалния живот като бази данни, дизайн на електронни схеми, разбиване на кодове, обработка на изображения и изкуствен интелект. Някои от основните термини, които могат да помогнат да се разберт генетичните алгоритми са:

Популация: Това е подмножество от всички вероятни решения, които могат да решат дадения проблем.

³ <https://www.section.io/engineering-education/the-basics-of-genetic-algorithms-in-ml/>

Хромозоми: Хромозомата е едно от решенията в популацията.

Ген: Това е елемент в хромозомата.

Алел: Това е стойността, дадена на ген в специфична хромозома.

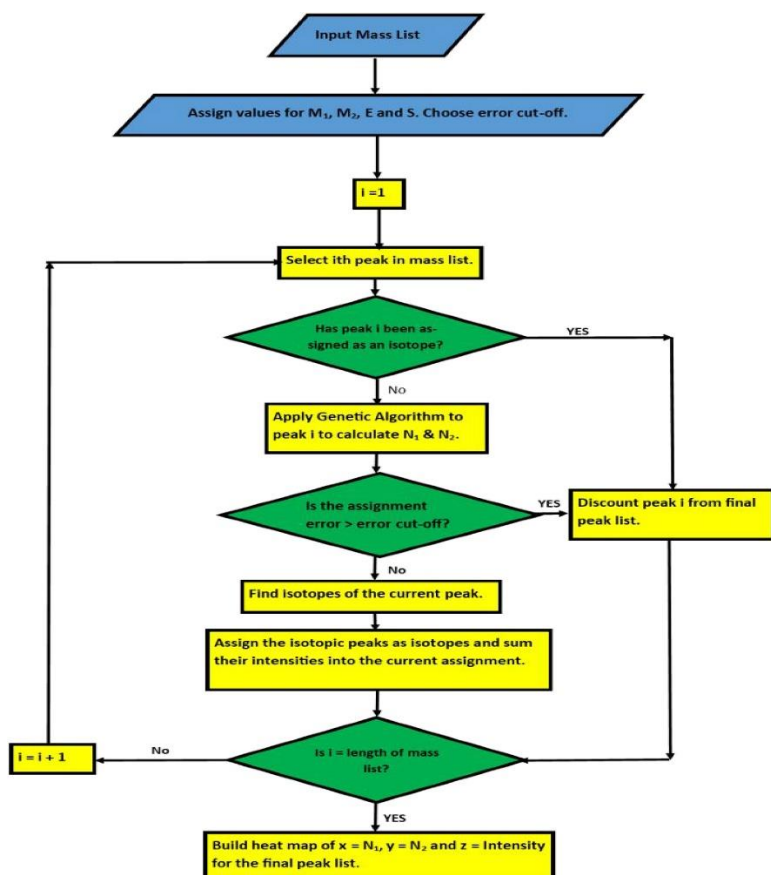
Fitness function: Това е функция, която използва специфичен вход, за да произведе подобрен резултат. Решението се използва като вход, докато изходът е под формата на пригодност на решението.

Генетични оператори: В генетичните алгоритми най-добрите индивиди се чифтосват, за да възпроизведат потомство, което е по-добро от родителското. Генетичните оператори се използват за промяна на генетичния състав на това следващо поколение.

Генетичният алгоритъм (GA) е алгоритъм за търсене, използван за решаване на проблеми с търсене и оптимизация на данни. Този алгоритъм е подмножество от еволюционни алгоритми, които се използват при изчисленията. Генетичните алгоритми използват концепцията за генетиката и естествения подбор, за да предоставят решения на проблемите.

Тези алгоритми имат **по-добра селективност** от алгоритмите за произволно търсене, тъй като използват исторически данни, за да отведат търсенето до най-ефективния изход в пространството за решения.

GA също се основават на поведението на хромозомите и тяхната генетична структура. Всяка хромозома играе ролята на осигуряване на възможно решение. *Fitness function* помага за осигуряване на характеристиките на всички индивиди в популацията. Колкото по-голяма е функцията, толкова по-добро е решението.



Фиг. 6: Принцип на действие на алгоритъма

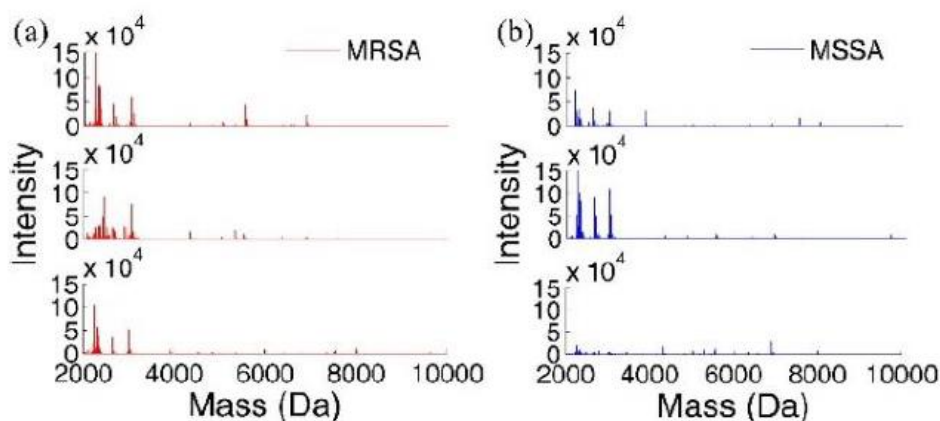


Figure 1. Replication of The Same Type Strains and Discrimination Between Two Types for The Mass Spectra of *S. aureus*.

Фиг. 7: Сравнение на спектрите, анализирани по конвенционален начин и тези, при които е използван генетичен алгоритъм при *S. Aureus*

Пример за внедряване на генетичен алгоритъм при анализа на спектрите от MALDI-TOF с цел определяне на вида бактерия е проучването „*Classification of Methicillin-Resistant and Methicillin-Susceptible Staphylococcus Aureus Using an Improved Genetic Algorithm for Feature Selection Based on Mass Spectra*“ на J. Bai, Z. C. Fan, L. P. Zhang, X. Y. Xu, Z. L. Zhang, достъпно на следния интернет адрес: <https://dl.acm.org/doi/abs/10.1145/3093293.3093299>.

Artificial/Supervised Neural Networks⁴

Изкуствените невронни мрежи са **изчислителни модели и са вдъхновени от човешкия мозък**. Много от последните постижения са направени в областта на изкуствения интелект, включително **гласово разпознаване, разпознаване на изображения, роботика, използваща изкуствени невронни мрежи**. Изкуствените невронни мрежи са биологично вдъхновени симулации, извършвани на компютър за изпълнение на определени специфични задачи като:

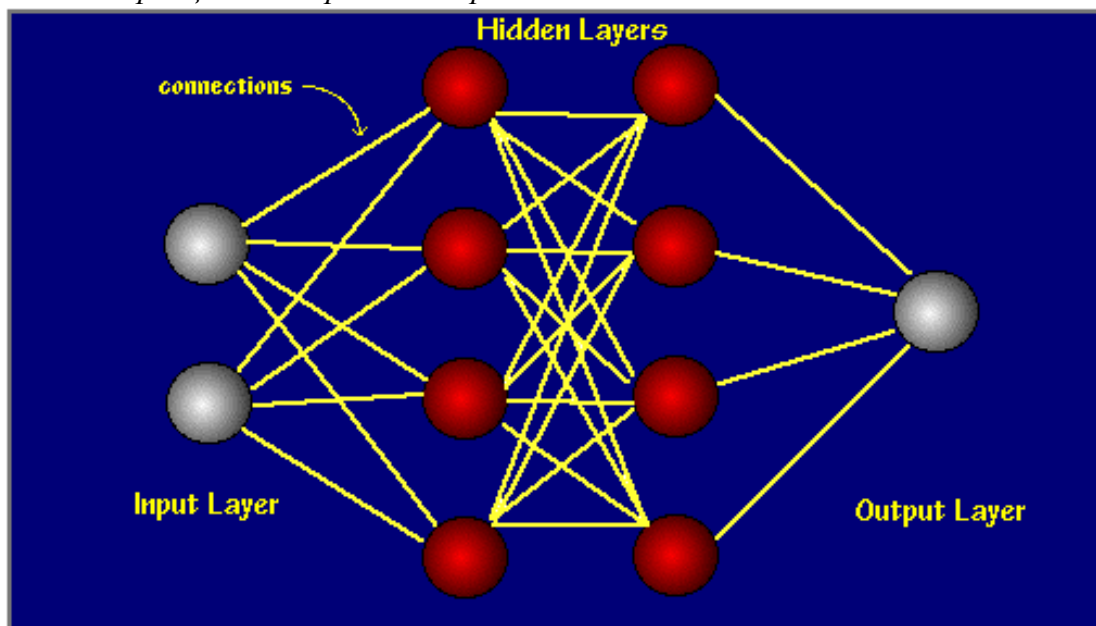
- Групиране
- Класификация
- Разпознаване на шаблон

Тези биологични методи на изчисление са известни като следващото голямо постижение в компютърната наука.

Терминът „невронна“ произлиза от основната функционална единица на човешката (животинската) нервна система „неврони“ или нервни клетки, присъстващи в мозъка и други части на човешкото (животинското) тяло. Невронната мрежа е група от алгоритми, които **удостоверяват основната връзка в набор от данни, подобни на човешкия мозък**. Невронната мрежа помага за промяна на входа, така че мрежата да дава най-добрия резултат, без да се препроектира изходната процедура.

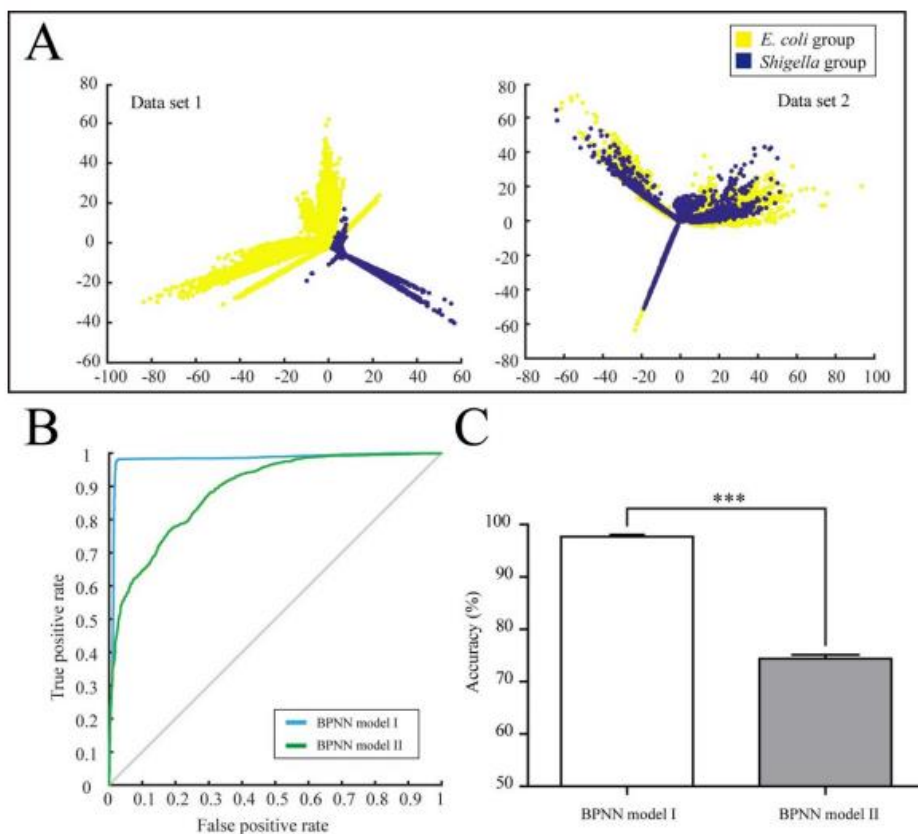
⁴ <https://www.xenonstack.com/blog/artificial-neural-network-applications>

Фиг. 8: Принцип на невронните мрежи:



Пример за внедряване на невронни мрежи като част от изкуствения интелект при анализа на маспектрите и видовото определяне на микроорганизми може да се намери в три проучвания:

1. „A novel short-term high-lactose culture approach combined with a matrix-assisted laser desorption ionization-time of flight mass spectrometry assay for differentiating *Escherichia coli* and *Shigella* species using artificial neural networks“ на Jin Ling, Hong Wang, Gaomin Li1, Zhen Feng, Yufei Song, Peng Wang, Hong Shao, Hu Zhou, Gang Chen
2. „Identification of *Bacillus anthracis* by Using Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry and Artificial Neural Networks“ на Peter Lasch, Wolfgang Beyer, Herbert Nattermann, Maren Sta`mmler, Enrico Siegbrecht, Roland Grunow, and Dieter Naumann (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2786504/pdf/0857-09.pdf>)
3. „An Overview of Multi-Task Learning in Deep Neural Networks“ - Sebastian Ruder - <https://arxiv.org/pdf/1706.05098.pdf>



Фиг. 9: Изглед на анализираните спектри чрез невронни мрежи и изкуствен интелект, проучване, което може да бъде достъпно на следния линк: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0222636>

Quick Classifiers⁵

Класификаторът в изкуствения интелект е алгоритъм, който **автоматично подрежда или категоризира данните в един или повече набори от „класове“**. Един от най-често срещаните примери е имейл класификатор, който сканира имейли, за да ги филтрира по етикет в клас: Спам или Не Спам.

Алгоритмите за изкуствен интелект са **полезни за автоматизиране на задачи**, които преди това трябваше да се извършват ръчно. Те могат да **спестят огромно количество време и пари** и да направят бизнеса по-ефективен.

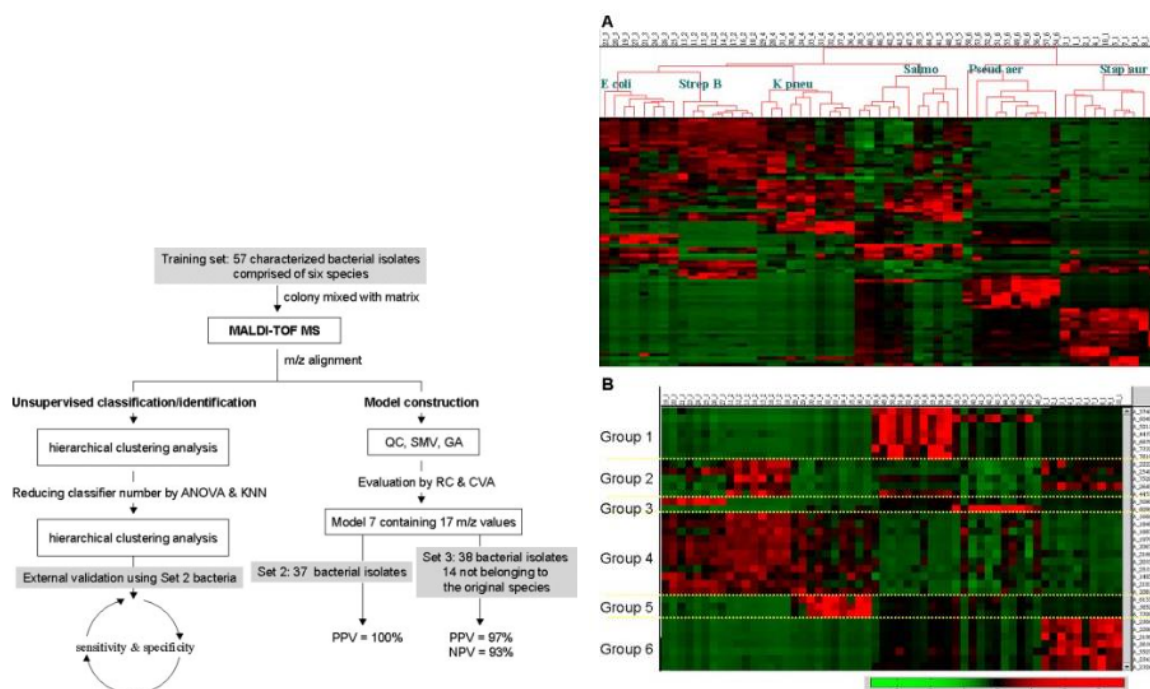
Класификаторът е самият алгоритъм – правилата, използвани от машините за класифициране на данни. Класификационният модел, от друга страна, е крайният резултат от изкуствения интелект на класификатора. Моделът се обучава с помощта на класификатора, така че моделът в крайна сметка класифицира постъпилите данни.

Има както контролирани, така и неконтролирани класификатори. Класификаторите за изкуствен интелект без надзор се подават само на немаркирани набори от данни, които те класифицират според разпознаването на модели или структури и аномалии в данните. На контролираните и полуконтролираните класификатори се подават набори от обучителни данни, от които се научават да класифицират данни според предварително определени категории.

⁵ <https://monkeylearn.com/blog/what-is-a-classifier/>

Анализът на настроената и мненията при хората при провеждане на социологически проучвания е пример за контролирано машинно обучение, при което класификаторите се обучават да анализират текст за полярност на мнението и да извеждат текста в класа: Положителен, Неутрален или Отрицателен.

Пример за приложение на бързите класификатори от изкуствения интелект за класифициране на микроорганизмите в клъстери е проучването на тема „*Highly Efficient Classification and Identification of Human Pathogenic Bacteria by MALDI-TOF MS*“ на *Sen-Yung Hsieh, Chiao-Li Tseng, Yun-Shien Lee, An-Jing Kuo* и *ekun* - <https://www.sciencedirect.com/science/article/pii/S1535947620312433#fig2>

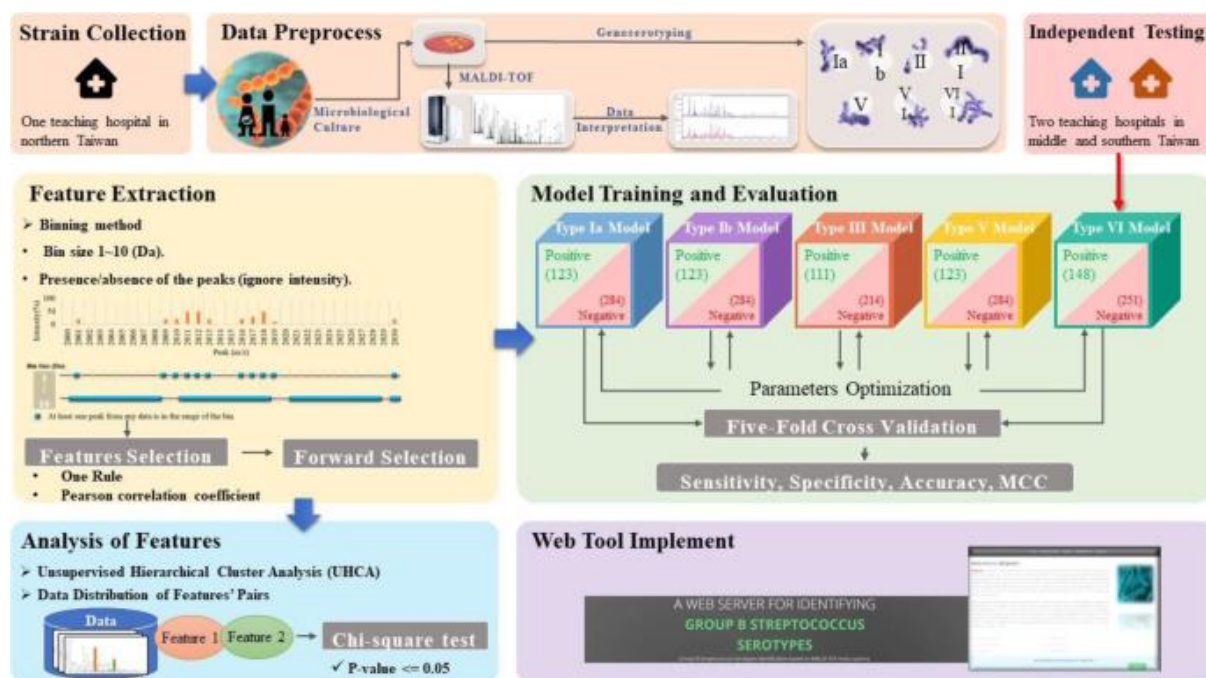


Фиг. 10: Изглед на самия алгоритъм (отляво) и резултатите от анализа на спектрите (отдясно)

Валидиране на модела

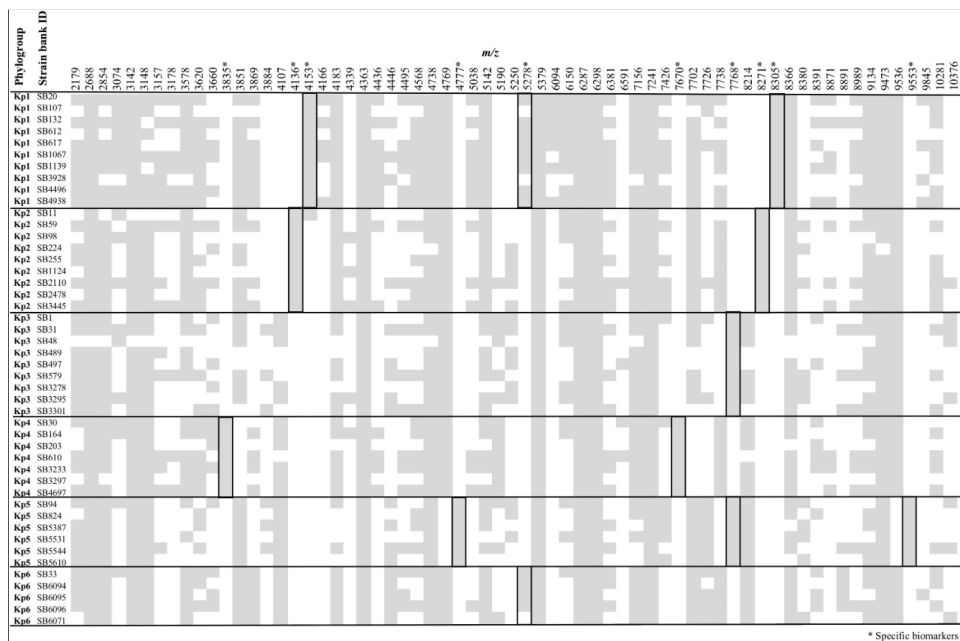
При всички прегледани проучвания се използва вид **кръстосано валидиране** (напр. двойно, десетократно или безусловно кръстосано валидиране), за да се избегнат грешни алгоритми или неточни данни и резултати. Едно проучване може да използва множество алгоритми за *overfitting*. Допълнително валидиране на моделите за данни от MALDI-TOF извън стандартното (т.е. външно валидиране) е извършено само в четири проучвания. Наборът от данни за външно валидиране се състои от данни, които са били получени от бактериална биобанка или данни от официален контрол или референтни такива. По-специално, *Wang* и колеги са събрали данни за обучението и тестовите от бактериалната банка на учебната болница в Северен Тайван. С цел да се разграничат заразните от екологичните щамове на *Streptococcus uberis* в млекодайните стада, *Esener* и екип са събрали данни от 29 стопанства, от които данните на 19 стопанства са използвани за обучение и тестване на алгоритмите, а останалите десет са използвани за външно валидиране. Повече за това проучване на тема: „*Discrimination of contagious and environmental strains of Streptococcus uberis in dairy herds by means of mass spectrometry and machine-learning*“ на *Necati Esener, Martin J. Green, Richard D. Emes, Benjamin Jowett,*

Peers L. Davies, Andrew J. Bradley & Tania Dottorini може да бъде намерено на следния интернет адрес: <https://www.nature.com/articles/s41598-018-35867-6.pdf>



Фиг. 11: Схематично представен процеса по избиране на алгоритъм, обучението му и респективно прилагането му при *Streptococcus*

В проучване на тема: „*Identification of Klebsiella pneumoniae, Klebsiella quasipneumoniae, Klebsiella variicola and Related Phylogroups by MALDI-TOF Mass Spectrometry*“ на *Carla Rodrigues, Virginie Passet, Andrianiaina Rakotondrasoana and Sylvain Brisse* алгоритми за изкуствен интелект на 40 изолата *Mycobacterium abscessus*, събрани във Франция, са обучени и тествани. Последващото външно валидиране е извършено на 40 *M. abscessus* изолати от френския Национален референтен център за микобактерии и е изследвана устойчивостта на *Mycobacteria* на антитуберкулозни препарати. *Rodrigues et al.* са насочили проучванията си към точно идентифициране на различни видове от рода *Klebsiella* чрез изкуствен интелект, въз основа на 46 щама, събрани от различни източници по целия свят (напр. от хора, околна среда, води, растения) като обучителни. За да валидират своите модели, базирани на SVM, *Rodriguez et al.* са анализирали изолати, принадлежащи към филогенетичните групи *K. pneumoniae*, получени от 49 фекални проби от хора в Мадагаскар. Чувствителността, която е постигната при тези алгоритми е между 60-100%, а специфичността е между 85-100%. Фигури 12 и 13 са представителни за това проучване за *K. Pneumoniae*.

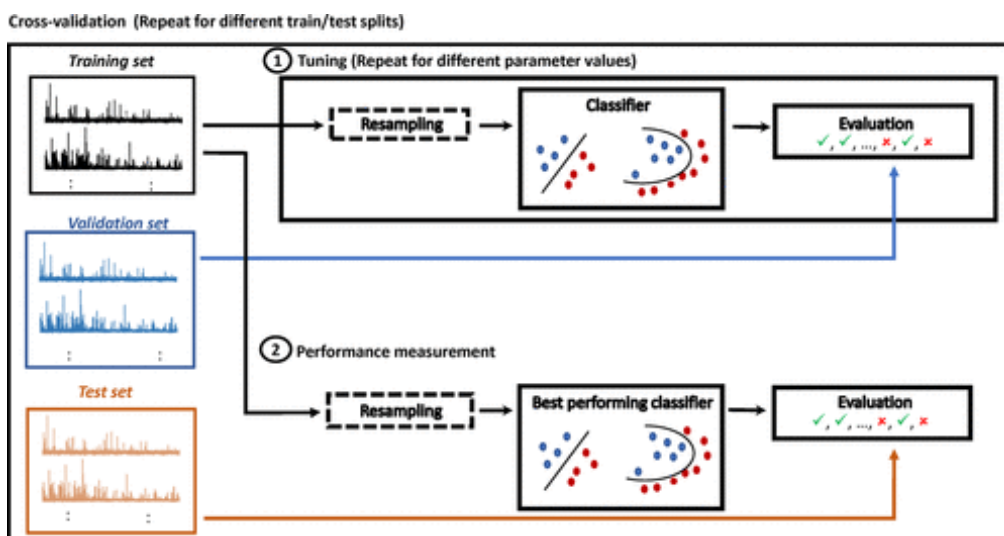


фиг. 12

Present in Kp phylogroup(s)	Peak position (m/z) ¹	Sensitivity ² [95% CI]	Specificity ³ [95% CI]	Possible proteins ⁴
Kp1	4153	10/10, 100% [69.2–100.0%]	35/36, 97.2% [85.5–99.9%]	Y]bJ (double charged ion)
Kp2	8305	10/10, 100% [69.2–100.0%]	36/36, 100% [90.3–100.0%]	Y]bJ
	4136	9/9, 100% [66.4–100.0%]	37/37, 100% [90.5–100.0%]	Y]bJ (double charged ion)
Kp3 and Kp5	8271	9/9, 100% [66.4–100.0%]	37/37, 100% [90.5–100.0%]	Y]bJ
	7768	15/15, 100% [78.2–100.0%]	31/31, 100% [88.8–100.0%]	Ribosomal protein L31
Kp4	3835	7/7, 100% [59.0–100.0%]	39/39, 100% [91.0–100.0%]	YdgH (chain, double charged ion)
	7670	7/7, 100% [59.0–100.0%]	39/39, 100% [91.0–100.0%]	YdgH (chain)
Kp5	4777	6/6, 100% [54.1–100.0%]	40/40, 100% [91.2–100.0%]	Ribosomal protein S20 (demethionated form, double charged ion)
	9553	6/6, 100% [54.1–100.0%]	40/40, 100% [91.2–100.0%]	Ribosomal protein S20 (demethionated form)
Kp1 and Kp6	5278	12/15, 80% [51.9–95.7%]	41/41, 100% [88.8–100.0%]	Ribosomal protein S22

CI, confidence interval. ¹Position in the spectra using a tolerance of $\pm 0.03\%$. ²Proportion of true positives that are correctly identified as such. ³Proportion of true negatives that are correctly identified as such. ⁴As determined using TagIdent (<https://web.expasy.org/tagident/tagident.html>).

фиг. 13



Фиг. 14: Принцип на обучение на алгоритъм и кросвалидиране

Таблица 1: а) Алгоритми за изкуствен интелект и кратка характеристика за тях, б) видове алгоритми и брой проучвания, в които са приложени:

А)

Algorithm	Description
Genetic Algorithm (GA)	Genetic algorithms are optimization algorithms belonging to the larger class for evolutionary algorithms. They rely on operations inspired by biological operations, such as mutations, crossover and selection by evolving a collection of candidate solutions towards a better performing solution. The genetic algorithm is used to select a combination of peaks that separate the classes best, using a cost function that measures the variance between classes. When the genetic algorithm is selected in ClinProTools, the genetic algorithm is only used as a peak (feature) selection algorithm. The classification on unseen instances is performed using a kNN algorithm based on the selected peaks.
Artificial/Supervised Neural Network (ANN)	Artificial neural networks are algorithms structurally modelled after mammalian brain neural networks. The network consists of several stacked layers of relatively simple mathematical units, which take as input information of several neurons of the previous layers and pass on the output to several neurons in the next layer. Most commonly, the learning takes place by taking the gradient of the loss function and adjusting the model weights to decrease the loss, leading those networks also to be called 'back propagation neural networks' (BPNNs). While unsupervised versions of artificial neural networks exist, the more common supervised setting is used for classification with neural networks.
Support Vector Machine (SVM)	Support vector machines are a supervised learning algorithm that finds the best separating maximum margin hyperplane between the classes in a higher dimensional representation of the instances. Hyperplane is the general term for planes with one dimension less than the dimension they are in. During optimization the hyperplane maximizing the gap between the plane and the instances is determined. The data representation in higher dimensional space is performed by kernel functions. Commonly used kernel functions include the radial basis function kernel and the polynomial kernel. While SVM is a classification algorithm, when selected in ClinProTools, the SVM algorithm is only used as a peak (feature) selection algorithm. The classification on unseen instances is performed using a kNN algorithm based on the selected peaks.
Quick Classifier (QC)	The Quick Classifier calculates the average area of each peak together and provides a p value per class. During classification, the peak areas are sorted by the univariate sorting algorithm and an average over all peaks is calculated which indicates class membership.
K Nearest Neighbor (kNN)	K-nearest neighbor algorithm is a classification based on the similarity between the instance to classify and known training instances. For classification of an unseen instance, the similarity between the instance and each training data point. The assigned class is determined as the majority class of the closest k training data points. A commonly used similarity measure is the Euclidean distance.

Б)

Type of algorithm	n (%)
Support Vector Machine (SVM)	18 (50)
Genetic Algorithm (GA)	15 (41.7)
Artificial/Supervised Neural Network (ANN)	13 (36.1)
Quick Classifier (QC)	11 (30.5)
Random Forest (RF)	9 (5)
Clustering/hierarchical cluster analysis (UHCA)	8 (22.2)
k-Nearest Neighbors (kNN)	5 (13.9)
Decision Tree (DT)	4 (11.1)
Logistic regression (single and multi)	3 (8.3)
Aristotle Classifier	1 (2.8)
Linear Discriminant Analysis	1 (2.8)
Naïve Bayes	1 (2.8)

Качество на включените проучвания

Резултатите от оценката на качеството на проучванията варира от лошо (и 60%) до много добро (100%). Само едно проучване изпълнява всичките девет критерия за качество и по този начин достигна 100% в рейтинга. Четири от деветте критерия са изпълнени от повече от 97% от проучванията (35 от 36). А именно, **всички проучвания подчертават ограниченията на настоящите подходи, не включващи изкуствен интелект** и предоставят информация за хардуера и софтуера, използвани в проучването, използват валидни методи за избягване на *overfitting* на модела и предоставят информация за клиничната значимост. Само 11% от проучванията валидират моделите за изкуствен интелект върху набор от външни данни.

От общо 36 проучвания е налице значително припокриване между проучванията по отношение на видовете патогени (напр. *Staphylococcus*, *Streptococcus* и *Escherichia*) и видовете приложени алгоритми за изкуствен интелект (напр. SVM, RF и GA). Това е така, защото голяма част от анализите са извършени чрез софтуер, предоставен от производителя, като *flexAnalysis* и *Clin-ProTools* от *Bruker Daltonics*, с предварително програмирани алгоритми за изкуствен интелект. Четири алгоритъма за изкуствен интелект са имплементирани в *ClinProTools*, а именно GA, SVM, SNN и QC. Като втора стъпка е извършена систематична оценка на проучванията, за да се оцени качеството на използваните модели за изкуствен интелект. Тази оценка на качеството недвусмислено посочва **няколко недостатъка на реализациите на изкуствения интелект, които трябва да бъдат преодолени, за да бъдат включени в рутинната диагностика.**

От 36 проучвания, 27 използват изкуствен интелект за подобряване на MALDI-TOF-базирана видове идентификация. От гледна точка на анализа на данните, идентификацията на видът може да бъде определен като мултикласова (или двоична) задача, включваща определени **предизвикателства, включително висока разделителна способност, огромен брой класове и небалансирани класови съотношения.**

Видове идентификация и профилиране за антибиотична резистентност

Ранното използване на ефективни антимикробни лечения е от решаващо значение за изхода от инфекциите и предотвратяването на резистентност към лечението. **Тестването на антимикробна резистентност дава възможност за избор на оптимални антибиотични лечения, но настоящите клетъчно културални фенотипни техники, могат да отнемат до 72 часа, за да генерират резултати.** В проучване на тема: *Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning* - *Caroline Weis, Aline Cuénod, Bastian Rieck, Olivier Dubuis, Susanne Graf, Claudia Lang, Michael Oberle, Maximilian Brackmann, Kirstine K. Søgaard, Michael Osthoff, Karsten Borgwardt & Adrian Egli* са разработени **нови подходи за машинно обучение за прогнозиране на антимикробна резистентност директно от профилите на маспектрите на клиничните изолати от MALDI-TOF.** Обучени са калибрирани класификатори с помощта на новосъздадена публично достъпна база данни с профили на мас спектри от клинично най-значимите изолати със свързани фенотипове на антимикробна чувствителност. Този набор от данни комбинира повече от 300 000 маспектри с повече от 750 000 фенотипа на антимикробна резистентност от четири медицински институции. Валидирането е осъществено с панел от клинично важни патогени, **включително *Staphylococcus aureus*, *Escherichia coli* и *Klebsiella pneumoniae*,** което демонстрира потенциала за използване на изкуствения интелект за прогнозиране на антимикробната резистентност и за промяна в клиничната и терапевтичната практика. Освен това в клинично проучване на 63 пациенти е установено, че **прилагането на този подход би променило терапията и лечението в девет случая, което би било от полза за клиничната практика.** По този начин базираните на маспектри MALDI-TOF изкуствен интелект и алгоритмите към него може да бъдат важен нов инструмент за оптимизиране на лечението и управление на потреблението на антибиотици.

Пълния набор от данни в проучването може да бъде достъпен на: <https://datadryad.org/stash/dataset/doi:10.5061/dryad.bzkh1899q>, а абстракт на обстойното проучване - <https://www.nature.com/articles/s41591-021-01619-9#article-info>.

Методите на изкуствения интелект може да се използват за модифициране на представените от MALDI-TOF спектрите данни чрез премахване на неинформационни променливи и да се определи коя специфична част от (неизвестна) информацията ще бъде най-ценна за получаване и оптимизиране на идентификация на бактериални видове. Чрез разпознаване на конкретни шаблони (*patterns*) в данните, техниките за изкуствен интелект използват максимално информацията, която е вградена в MALDI-TOF спектрите. Повечето проучвания съобщават за по-добра производителност на алгоритмите за изкуствен интелект в задачата за класификация на бактериите в сравнение с настоящите подходи. Например, използването на ANN на MALDI-TOF спектрите позволява бързото и точно идентифициране на *Bacillus fragilis* и някои подвидове от него. По подобен начин *De Bruyne* и екип демонстрират колко добре техниките за изкуствен интелект могат да разграничат MALDI-TOF спектрите от различни видове или класове. По-специално, те прилагат SVM и RF към бинарни MALDI-TOF спектри на видове в родовете *Leuconostoc*, *Fructobacillus* и *Lactococcus* и дискриминационното представяне на тези методи е отлично. В допълнение към идентификацията на вида, алгоритмите на изкуствения интелект биха могли да се прилагат за прогнозиране на AMP. Шест проучвания са използвали алгоритми на изкуствен интелект за тестване на антимикробната чувствителност на *Staphylococcus aureus*, патоген, който причинява висока смъртност както в болничните заведения, така и в общността поради недостига на ефективни лечения. *Sogawa* и екип използват SVM, за да диференцират метицилин-резистентни (MRSA) от метицилин-чувствителни *S. aureus* (MSSA) въз основа на характеристики, получени от маспектрите MALDI-TOF. Техният модел достига **точност на прогнозиране от над 85%** и **значително намалява времето до започване на целево антибиотично лечение** в сравнение с профилиране/определяне на фенотипната резистентност. На фона на резистентността ванкомицин се появява като лечение на първа линия за MRSA инфекции. За съжаление, щамовете на MRSA също са станали резистентни към ванкомицин в резултат на прекомерната употреба/злоупотребата, която е причинила натрупване на мутации в гените, участващи в синтеза на клетъчна стена. Получените щамове на ванкомицин-междинен *S. aureus* (VISA) са свързани с упорити персистиращи инфекции и недостатъчност на лечението. Както SVM, така и RF са приложени към VISA и ванкомицин-чувствителен *S. aureus* (VSSA) MALDI-TOF маспектри за диференциране на щамове и фенотипно прогнозиране, с **докладвани точности от над 97%**. Сравними успехи на прогнозата за AMP, базирана на изкуствен интелект, също са докладвани за β -лактам-резистентен *B. fragilis*. **SVM-базираните алгоритми са в състояние да предскажат наличието на *cfi* ген в *B. fragilis* с точност, близо 100%**. Предполага се, че генът *cfi* кодира производството на метало- β -лактамаза, която хидролизира карбапенемите и другите β -лактамни антибиотици и по този начин прави тези лечения неефективни срещу тези щамове *B. fragilis*. *Huang* и колегите му са използвали различни алгоритми за изкуствен интелект (*KNN*, *RF*, *SVM*, *Naïve Bayes* и логистична регресия), от които RF превъзхожда останалите четири алгоритъма при разграничаване на карбапенем-устойчиви от податливи на карбапенем *Klebsiella pneumoniae* с точност над 93%. Не на последно място, едно скорошно проучване на

принципа на доказване на хипотеза в областта на микологията, показва **потенциала на използването на подходи за изкуствен интелект за прогнозиране на резистентността към противогъбични средства (напр. флуконазол).**

Оценката на модела е от ключово значение

Успехът на клиничната приложимост на изкуствения интелект зависи от тяхната здравина, надеждност, валидност и специфичност. Оценката на качеството показва, че по-голямата част от проведените проучвания са предприели валидни стъпки за оценка на устойчивостта на техните модели чрез стандартно k-кратно кръстосано валидиране и посочване на стабилността на резултатите (напр. стандартно отклонение, доверителни интервали). При кръстосаното валидиране наборът от данни е разделен на k подгрупи, наречени криви (*folds*). В последствие алгоритъмът е итеративно обучен на k-1 криви, като се използват останалите криви като тестов комплект (наречен „крива на задържане“). По време на процедурата за кръстосано валидиране, хиперпараметрите могат да бъдат настроени само с помощта на оригиналния набор от инструменти за обучение. След това подборът на модела и оценката на ефективността се формират, като се използва тестовият набор (т.е. данни, които никога не са били виждани преди това). Само около 60,71% от прегледаните проучвания показват подробности за хиперпараметрите, които са незаменими при повторни изследвания. **За да се прецени клиничната приложимост на моделите, са необходими допълнителни видове валидиране.** Първо, **общата или оперативната съвместимост на моделите е от съществено значение**, за да се даде възможност за сравнение между клиничните и теоретичните приложения. Масспектрите MALDI-TOF, измерени на различни места, крият негативния ефект, произтичащ от различия в лабораторните рутинни процедури или машинните настройки. Следователно моделите, обучени за характеризирание и анализ на спектри, получени от вътреболнични изолати, вероятно имат по-лоши прогнози от данните, събрани от извънболнична среда (наричани още *out-of-distribution data*). Обобщаемостта или оперативната съвместимост на моделите се оценява чрез т.нар. външно валидиране, по време на което моделът се тества с непознати данни, които не са му представяни до момента (напр. данни за изолати от различни болници или лечебни заведения). Като се има предвид, че **само ~ 11% от всички проучвания са извършили външно валидиране, това поставя голямо ограничение и несигурност в настоящите рамки на употреба на изкуствен интелект за анализ на данни MALDI-TOF MS.**

Второ, **тълкуваемостта на моделите за изкуствен интелект е от решаващо значение за прилагането на предиктори при грижите за пациентите.** Доверителният интервал в модела е изграден чрез анализ и охарактеризиране на пиковете от MALDI-TOF масспектрите, от които се извлича информация. Почти всички прегледани изследвания дават обяснение (биологично или количествено) за значението на прогнозните пикове. *Esener et al.*, например, прави кръстосана проверка на идентифицираните пикове, които биват сравнявани с базата данни за протеини на NCBI, за да се определи съответният протеин. Въз основа на този анализ тези учени откриват, че техните пикове съответстват на бактериоцин и рибозомни протеини. В друго проучване *Ho et al.* използват *CabaNP* анализ, за да характеризират щамовете *Bacteroides fragilis*, които са класифицирани като *sfIA*-положителни и по този начин е вероятно да бъдат резистентни към антибиотици от групата на карбапенемите. Този

алгоритъм разкрива, че над една трета от тези щамове съдържат последователността на ген *cfiA*, за която се подозира, че причинява резистентност към карбапенем на *Bacteroides fragilis*. Освен това пиковите на MALDI-TOF MS могат да дадат представа дали известните свойства на резистентността или несвързаните филогенетични разлики могат да бъдат прогнозирани ефективно. По този начин, *Sogawa et al.* сравняват предварителните пикове, избрани от техния алгоритъм, с известни фрагменти от пеницилин-свързващия протеин РВР, причиняващ резистентност към метицилин в MRSA. Въпреки това биологичната интерпретация на пиковите на характеристиките е рядкост в резултат на ограничени предварителни познания, тъй като повечето механизми на резистентност не са били изследвани преди това посредством MALDI-TOF маспектрите.

Ограничения на приложенията за изкуствен интелект

Приложенията за изкуствен интелект за анализ на данни от MALDI-TOF маспектри все още са в ранен етап на разработка, което се потвърждава и от малкия брой налични публикации. Идентифицирани са три основни пречки пред бързия напредък: малък размер на извадката, липса на външно валидиране и лоша възпроизводимост. Всички прегледани проучвания са ограничени до малки размери на пробите, вариращи между десетки и стотици изолати за обучение и тестване, като най-голямото проучване включва 787 изолати. Изолираните колекции с такъв размер едва ли обхващат микробното разнообразие, което е налично в клиничната практика. Тренировъчните програми за изкуствен интелект са възпрепятствани от малки и микробни неотдиференцирани проби, които е вероятно да предизвикат фалшиво определяне и слабо категоризиране. Широкото използване на моделите е допълнително възпрепятствано от липсата на външно валидиране. И двете ограничения биха могли да бъдат преодоляни чрез големи публично достъпни масиви от данни. Например, *MassIVE* хранилището е общодостъпна база данни за маспектрометрия, фокусирана главно върху протеомиката и съдържаща масиви от данни от MALDI-TOF. По-голям брой спектри е осигурен от Института Робърт Кох, включително 6264 MALDI-TOF маспектри на високопатогенни микроорганизми. В допълнение към голям брой проби, тези хранилища на данни в идеалния случай следва да предоставят AMP профили заедно със спектрите и ключовата информация, за да се създаде добре дефинирана формулировка на проблема (напр. анотация на подлинните и AMP профили). И накрая, възпроизводимостта на прегледаните проучвания е ниска. Масивите от данни от MALDI-TOF MS и алгоритмите за изкуствен интелект рядко се оповестяват публично. Освен това информацията за избора на хиперпараметри за модели често е недостатъчна. Липсата на тази информация прави практически невъзможно да се разбере, да се приложи отново и най-важното да се възпроизведат описаните оперативни процедури за анализ. Проблемът с приложението на изкуствения интелект се задълбочава от факта, че по-голямата част от анализите се извършват конвенционално на софтуер, предоставен от производителя, като *flexAnalysis* и *ClinProTools* от *Bruker Daltonics*. В резултат на това, приложените работни процедури не са достъпни за моделиране и настройка според нуждите на конкретния анализ, като по този начин се възпрепятстват всякакви външни опити за повторно използване на тези анализи или за внедряване на нови. За справяне с тези предизвикателства и разгръщане на пълния потенциал на приложенията за

изкуствен интелект са **необходими съвместни усилия на клиниките, лабораториите за изпитване, статистиците и анализаторите, както и на експертите, занимаващи се със създаване на тези алгоритми за изкуствен интелект.** В допълнение към публично достъпните набори от данни от MALDI- TOF в държавите членки, които са подходящи за анализ посредством изкуствен интелект, **следва аналитичните алгоритми и информатичния код да се оповестяват публично в бази данни като *GitHub*.** Това ще насърчи нови експериментални проучвания, ще осигури основа за нови и ще позволи прилагането на усъвършенствани алгоритми за изкуствен интелект.

Липсата на проучвания, говори и за лошо представяне на алгоритмите за изкуствен интелект по отношение на идентификация на видовете и тестване на микробната чувствителност, говори за голяма вероятност от пристрастия.

За разлика от други биомедицински области, **микробиологията изостава по отношение на извличането на ползата от внедряване на изкуствен интелект за анализ на големи масиви от данни.** Въпреки че цялостното качество на по-голямата част от проучванията е добро, **три основни ограничения, а именно малкият размер на извадката, лошата възпроизводимост и липсата на външно валидиране, възпрепятстват бързия напредък.** За да се разгърне пълният потенциал на ръководените от изкуствен интелект идентификация на видове и АМР профилиране, **е важно прилагането на мултидисциплинарен подход и постоянен обмен на информация и опит, за да могат да бъдат преодолените горепосочените предизвикателства.**

Изготвил:

Красимира Захариева,

Главен експерт в Дирекция «ОРХВ», ЦОРХВ

Използвана литература:

- *Machine learning for microbial identification and antimicrobial susceptibility testing on MALDI-TOF mass spectra: a systematic review* - C.V. Weis, y, C.R.Jutzeler, K.Borgwardt
- *Highly Efficient Classification and Identification of Human Pathogenic Bacteria by MALDI-TOF MS* - Sen-Yung Hsieh, Chiao-Li Tseng, Yun-Shien Lee, An-Jing Kuo, Chien-Feng Sun, Yen-Hsiu Lin, and Jen-Kun Chen
- *Identification of Bacillus anthracis by Using Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry and Artificial Neural Networks* - Peter Lasch, Wolfgang Beyer, Herbert Nattermann, Maren Staßmüller, Enrico Siegbrecht, Roland Grunow, and Dieter Naumann
- *An Overview of Multi-Task Learning in Deep Neural Networks* - Sebastian Ruder
- *Classification of Methicillin-Resistant and Methicillin-Susceptible Staphylococcus Aureus Using an Improved Genetic Algorithm for Feature Selection Based on Mass Spectra* - J. Bai, Z. C. Fan, L. P. Zhang, X. Y. Xu, Z. L. Zhang

- *COMPUTATIONAL PREDICTIONS, DYNAMIC TRACKING, AND EVOLUTIONARY ANALYSIS OF ANTIBIOTIC RESISTANCE THROUGH THE MINING OF MICROBIAL GENOMES AND METAGENOMIC DATA - Qi Zhao, Chin Yen Tay, Liang Wang, Jian Li*
- *Fast Pathogen Identification Using Single-Cell Matrix-Assisted Laser Desorption/Ionization-Aerosol Time-of-Flight Mass Spectrometry Data and Deep Learning Methods - Christina Papagiannopoulou, René Parchen, Peter Rubbens, and Willem Waegeman*
- *Support-Vector Networks - CORINNA CORTES, VLADIMIR VAPNIK*
- *Incorporating Statistical Test and Machine Intelligence Into Strain Typing of Staphylococcus haemolyticus Based on Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry - Chia-Ru Chung, Hsin-Yao Wang, Frank Lien, Yi-Ju Tseng, Chun-Hsien Chen, Tzong-Yi Lee, Tsui-Ping Liu, Jorng-Tzong Horng and Jang-Jih Lu*
- *Combination of MALDI-TOF Mass Spectrometry and Machine Learning for Rapid Antimicrobial Resistance Screening: The Case of Campylobacter spp. - Maureen Feucherolles, Morgane Nennig, Sören L. Becker, Delphine Martiny, Serge Losch, Christian Penny, Henry-Michel Cauchie and Catherine Ragimbeau*
- *A novel short-term high-lactose culture approach combined with a matrix-assisted laser desorption ionization-time of flight mass spectrometry assay for differentiating Escherichia coli and Shigella species using artificial neural networks - Jin Ling, Hong Wang, Gaomin Li, Zhen Feng, Yufei Song, Peng Wang, Hong Shao, Hu Zhou, Gang Chen*
- *MALDI- TOF mass spectrometry: an emerging technology for microbial identification and diagnosis – Neelja Singhal, Manish Kumar, Pawan K. Kanaujia and Jugsharan S. Viridi*
- *Identification of Klebsiella pneumoniae, Klebsiella quasipneumoniae, Klebsiella variicola and Related Phylogroups by MALDI-TOF Mass Spectrometry - Carla Rodrigues, Virginie Passet, Andrianiaina Rakotondrasoa and Sylvain Brisse*
- *Incorporating Statistical Test and Machine Intelligence Into Strain Typing of Staphylococcus haemolyticus Based on Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry - Chia-Ru Chung, Hsin-Yao Wang, Frank Lien, Yi-Ju Tseng, Chun-Hsien Chen, Tzong-Yi Lee, Tsui-Ping Liu, Jorng-Tzong Horng and Jang-Jih Lu*
- *Rapid classification of group B Streptococcus serotypes based on matrix assisted laser desorption ionization-time of flight mass spectrometry and machine learning techniques - Hsin-Yao Wang, Wen-Chi Li, Kai-Yao Huang, Chia-Ru Chung, Jorng-Tzong Horng, Jen-Fu Hsu, Jang-Jih Lu and Tzong-Yi Lee*
- *Classification Algorithm for Subspecies Identification within the Mycobacterium abscessus Species, Based on Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry - Marie-Sarah Fangous, Faiza Mougari, Stéphanie Gouriou, Elodie Calvez, Laurent Raskine, Emmanuelle Cambau, Christopher Payan, Geneviève Héry-Arnaud*
- *Implementation of MALDI-TOF Mass Spectrometry and Peak Analysis: Application to the Discrimination of Cryptococcus neoformans Species Complex and Their Interspecies Hybrids - Margarita E. Zvezdanova, Manuel J. Arroyo, Gema Méndez, Jesús Guinea, Luis Mancera, Patricia Muñoz, Belén Rodríguez-Sánchez and Pilar Escribano*
- *MALDI-TOF bacterial subtyping to detect antibiotic resistance - Miriam Cordovana, Arthur Boniface Pranada, Simone Ambretti, Markus Kostrzewa*
- *Metabolic Phenotyping and Strain Characterisation of Pseudomonas aeruginosa Isolates from Cystic Fibrosis Patients Using Rapid Evaporative Ionisation Mass Spectrometry -*

Emmanuelle E. Bardin, Simon J. S. Cameron, Alvaro Perdones-Montero, Kate Hardiman, Frances Bolt, Eric W. F. W. Alton, Andrew Bush, Jane C. Davies & Zoltan Takáts

- *Discrimination of contagious and environmental strains of Streptococcus uberis in dairy herds by means of mass spectrometry and machine-learning - Necati Esener, Martin J. Green, Richard D. Emes, Benjamin Jowett, Peers L. Davies, Andrew J. Bradley & Tania Dottorini*
- https://elitedatascience.com/overfitting-in-machine-learning?_ga=2.250250627.460119621.1653553086-291318358.1653553086#examples
- <https://www.researchgate.net/publication/356176525> *Paving the Way to Overcome Antifungal Drug Resistance Current Practices and Novel Developments for Rapid and Reliable Antifungal Susceptibility Testing*
- <https://www.creative-proteomics.com/technology/maldi-tof-mass-spectrometry.htm>
- <https://monkeylearn.com/blog/what-is-a-classifier/>
- <https://training.galaxyproject.org/training-material/topics/variant-analysis/tutorials/tb-variant-analysis/tutorial.html>
- <https://datadryad.org/stash/dataset/doi:10.5061/dryad.bzkh1899q>
- <https://www.nature.com/articles/s41591-021-01619-9>
- <https://www.xenonstack.com/blog/artificial-neural-network-applications>
- <https://pages.cs.wisc.edu/~bolo/shipyard/neural/local.html>
- <https://www.nature.com/articles/ncomms12965.pdf>