Deep learning: A Promising Approach For High-Throughput Plant Metabolomics

- **ISHA JOSHI**

High throughput sequencing and metabolomics technologies have shown a way towards storing and evaluating s big scale datasets on experimental basis to draw significant patterns [1-2]. Deep learning techniques provide an effective way to study, store and predict molecular phenotypes associated with a given plant architecture and draw insightful conclusions from it. Different plants possess different metabonomic profiles and molecular fingerprinting which are obtained by multiomic analysis like NMR spectroscopy, isotopic labelling, LC-MS and GC-MS [3]. Deep learning tools [4-5] enlisted in table 1 can pick statistical similarity matches between the metabolic profiles retrieved from databases like KNApSAcK[6] , PlantCyc[7], BioCyc[8], KEGG[9] and MetaPhen[10] to predict gene mutation[11], genotype[12], metabolic pathway[13], ecological niche[14], enzyme-metabolite interactions[15] , Biotic and Abiotic Stress associated changes[16]. MS/MS data in combination with machine learning is to map the mass spectra into molecular fingerprints [17-18]. Various medicinal plants acquire their health beneficial properties because of their differential metabolite content profile which is mainly contributed by higher rate of secondary metabolism. Deep learning studies focussing on the prediction of health beneficiary metabolites are enlisted and described in table 1. Parallel use of deep learning and MRI scanning can be used for secondary metabolome profiling of live plants [19].

Table.1- Tools and applications of High-throughput metabolomic studies.

S.No.	Secondary metabolites	Algorithm	Medicinal Effects	REFRENCES
1.	Phenolic Compounds, Flavonoids, Isocoumarins, Sesquiterpene, Lactones, Diterpenes	ANN	Anti-Inflammatory	[20]
2.	Phenol	SVM RF RP	Anti-Inflammatory, Antioxidant, Antimutagenic, Anticarcinogenic	[21-22]
3.	Alkaloids	RF	Pain-Relieving, Calming, Sleep-Inducing Properties	[23]
4.	Flavonoids, Anthocyanins, Tannins, Phenols	MLP RF ANN	Antioxidants	[24]
5.	Saponins	KNN SVM GBM RF	Toxicity Prevention, Decreased Blood Lipids, Lower Cancer Risks, Lower Blood Glucose Response	[25]
6.	Alkaloids, Glycosides, Saponins, Phenols, Tannins, Proteins, Flavonoids, Terpenoids	DL model	Antibacterial, Antioxidant, Wound Healing, Anti-Inflammatory activities	[26]
Metabolomics Tools-		Weka project, R GUI, MetNet, MetaFIND, MetaGeneAlyse, FIEmspro, XCMS, PRIMe, MeltDB, Workflow4Metabolomics, MetaboAnalyst, Metabox, MetFrag, MZmine2, MetAlign, MS-Dial, eMZed, MzMatch, IDEOM, MET-COFEA, MAVEN, iMet-Q, MarVis, TracMass, MaxQuant, OpenMS, ProtMAX, MS-Finder, Phenometer, ingerID, SIRIUS, Metaboanalyst, MeltDB 2.0, KNIME, WEKA, ORANGE, TensorFlow		[27-28].

References-

1. Sun, S., Wang, C., Ding, H. and Zou, Q., 2020. Machine learning and its applications in plant molecular studies. *Briefings in Functional Genomics*, *19*(1), pp.40-48.
2. Nayarisseri, A., Khandelwal, R., Madhavi, M., Selvaraj, C., Panwar, U., Sharma, K., Hussain, T. and Singh, S.K., 2020. Shape-based machine learning models for the potential novel COVID-19 protease inhibitors assisted by molecular dynamics simulation. *Current topics in medicinal chemistry*, *20*(24), pp.2146-2167.
3. Zampieri, G., Vijayakumar, S., Yaneske, E. and Angione, C., 2019. Machine and deep learning meet genome-scale metabolic modeling. *PLoS computational biology*, *15*(7), p.e1007084.
4. Scott, I.M., Vermeer, C.P., Liakata, M., Corol, D.I., Ward, J.L., Lin, W., Johnson, H.E., Whitehead, L., Kular, B., Baker, J.M. and Walsh, S., 2010. Enhancement of plant metabolite fingerprinting by machine learning. *Plant Physiology*, *153*(4), pp.1506-1520.
5. Cuperlovic-Culf, M., 2018. Machine learning methods for analysis of metabolic data and metabolic pathway modeling. *Metabolites*, *8*(1), p.4.
6. Nakamura, Y., Mochamad Afendi, F., Kawsar Parvin, A., Ono, N., Tanaka, K., Hirai Morita, A., Sato, T., Sugiura, T., Altaf-Ul-Amin, M. and Kanaya, S., 2014. KNApSAcK metabolite activity database for retrieving the relationships between metabolites and biological activities. *Plant and Cell Physiology*, *55*(1), pp.e7-e7.
7. Zhang, P., Dreher, K., Karthikeyan, A., Chi, A., Pujar, A., Caspi, R., Karp, P., Kirkup, V., Latendresse, M., Lee, C. and Mueller, L.A., 2010. Creation of a genome-wide metabolic pathway database for Populus trichocarpa using a new approach for reconstruction and curation of metabolic pathways for plants. *Plant Physiology*, *153*(4), pp.1479-1491.
8. Karp, P.D., Ouzounis, C.A., Moore-Kochlacs, C., Goldovsky, L., Kaipa, P., Ahrén, D., Tsoka, S., Darzentas, N., Kunin, V. and López-Bigas, N., 2005. Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic acids research*, *33*(19), pp.6083-6089.
9. Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M., 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic acids research*, *42*(D1), pp.D199-D205.
10. Carroll, A.J., Badger, M.R. and Millar, A.H., 2010. The MetabolomeExpress Project: enabling web-based processing, analysis and transparent dissemination of GC/MS metabolomics datasets. *BMC bioinformatics*, *11*(1), p.376.
11. Carroll, A.J., Zhang, P., Whitehead, L., Kaines, S., Tcherkez, G. and Badger, M.R., 2015. PhenoMeter: a metabolome database search tool using statistical similarity matching of metabolic phenotypes for high-confidence detection of functional links. *Frontiers in bioengineering and biotechnology*, *3*, p.106.
12. Taylor, J., King, R.D., Altmann, T. and Fiehn, O., 2002. Application of metabolomics to plant genotype discrimination using statistics and machine learning. *Bioinformatics*, *18*(suppl_2), pp.S241-S248.
13. Toubiana, D., Puzis, R., Wen, L., Sikron, N., Kurmanbayeva, A., Soltabayeva, A., Wilhelmi, M.D.M.R., Sade, N., Fait, A., Sagi, M. and Blumwald, E., 2019. Combined network analysis

and machine learning allows the prediction of metabolic pathways from tomato metabolomics data. *Communications biology*, *2*(1), pp.1-13.
14. Larsen, P.E., Collart, F.R. and Dai, Y., 2015. Predicting ecological roles in the rhizosphere using metabolome and transportome modeling. *PloS one*, *10*(9), p.e0132837.
15. Girbig, D., Grimbs, S. and Selbig, J., 2012. Systematic analysis of stability patterns in plant primary metabolism. *PLoS One*, *7*(4), p.e34686.
16. Piasecka, A., Kachlicki, P. and Stobiecki, M., 2019. Analytical methods for detection of plant metabolomes changes in response to biotic and abiotic stresses. *International journal of molecular sciences*, *20*(2), p.379.
17. Dührkop, K., Shen, H., Meusel, M., Rousu, J. and Böcker, S., 2015. Searching molecular structure databases with tandem mass spectra using CSI: FingerID. *Proceedings of the National Academy of Sciences*, *112*(41), pp.12580-12585.
18. Shen, H., Dührkop, K., Böcker, S. and Rousu, J., 2014. Metabolite identification through multiple kernel learning on fragmentation trees. *Bioinformatics*, *30*(12), pp.i157-i164.
19. Borisjuk, L., Rolletschek, H. and Neuberger, T., 2012. Surveying the plant's world by magnetic resonance imaging. *The Plant Journal*, *70*(1), pp.129-146.
20. Chagas-Paula, D.A., Oliveira, T.B., Zhang, T., Edrada-Ebel, R. and Da Costa, F.B., 2015. Prediction of anti-inflammatory plants and discovery of their biomarkers by machine learning algorithms and metabolomic studies. *Planta Medica*, *81*(06), pp.450-458.
21. Lemma, B., Grehl, C., Zech, M., Mekonnen, B., Zech, W., Nemomissa, S., Bekele, T. and Glaser, B., 2019. Phenolic compounds as unambiguous chemical markers for the identification of keystone plant species in the bale mountains, Ethiopia. *Plants*, *8*(7), p.228.
22. Iqbal, F., Lucieer, A. and Barry, K., 2020. Prediction of poppy thebaine alkaloid concentration using UAS remote sensing. *Precision Agriculture*, pp.1-12.
23. Guardado Yordi, E., Koelig, R., Matos, M.J., Pérez Martínez, A., Caballero, Y., Santana, L., Pérez Quintana, M., Molina, E. and Uriarte, E., 2019. Artificial intelligence applied to flavonoid data in food matrices. *Foods*, *8*(11), p.573.
24. Hosu, A., Cristea, V.M. and Cimpoiu, C., 2014. Analysis of total phenolic, flavonoids, anthocyanins and tannins content in Romanian red wines: Prediction of antioxidant activities and classification of wines using artificial neural networks. *Food chemistry*, *150*, pp.113-118.
25. Zheng, S., Wang, Y., Liu, H., Chang, W., Xu, Y. and Lin, F., 2019. Prediction of Hemolytic Toxicity for Saponins by Machine-Learning Methods. *Chemical research in toxicology*, *32*(6), pp.1014-1026.
26. Senthil Kumar Janahan, Arun Sahayadhas (2020) "A deep learning approach for predicting the active phytochemical constituent's presence in Premna Latifolia Leaf Extracts Data set.", *International Journal of Advanced Science and Technology*, 29(05), pp. 1075 - 1080.
27. Cuperlovic-Culf, M., 2018. Machine learning methods for analysis of metabolic data and metabolic pathway modeling. *Metabolites*, *8*(1), p.4.
28. Piasecka, A., Kachlicki, P. and Stobiecki, M., 2019. Analytical methods for detection of plant metabolomes changes in response to biotic and abiotic stresses. *International journal of molecular sciences*, *20*(2), p.379.