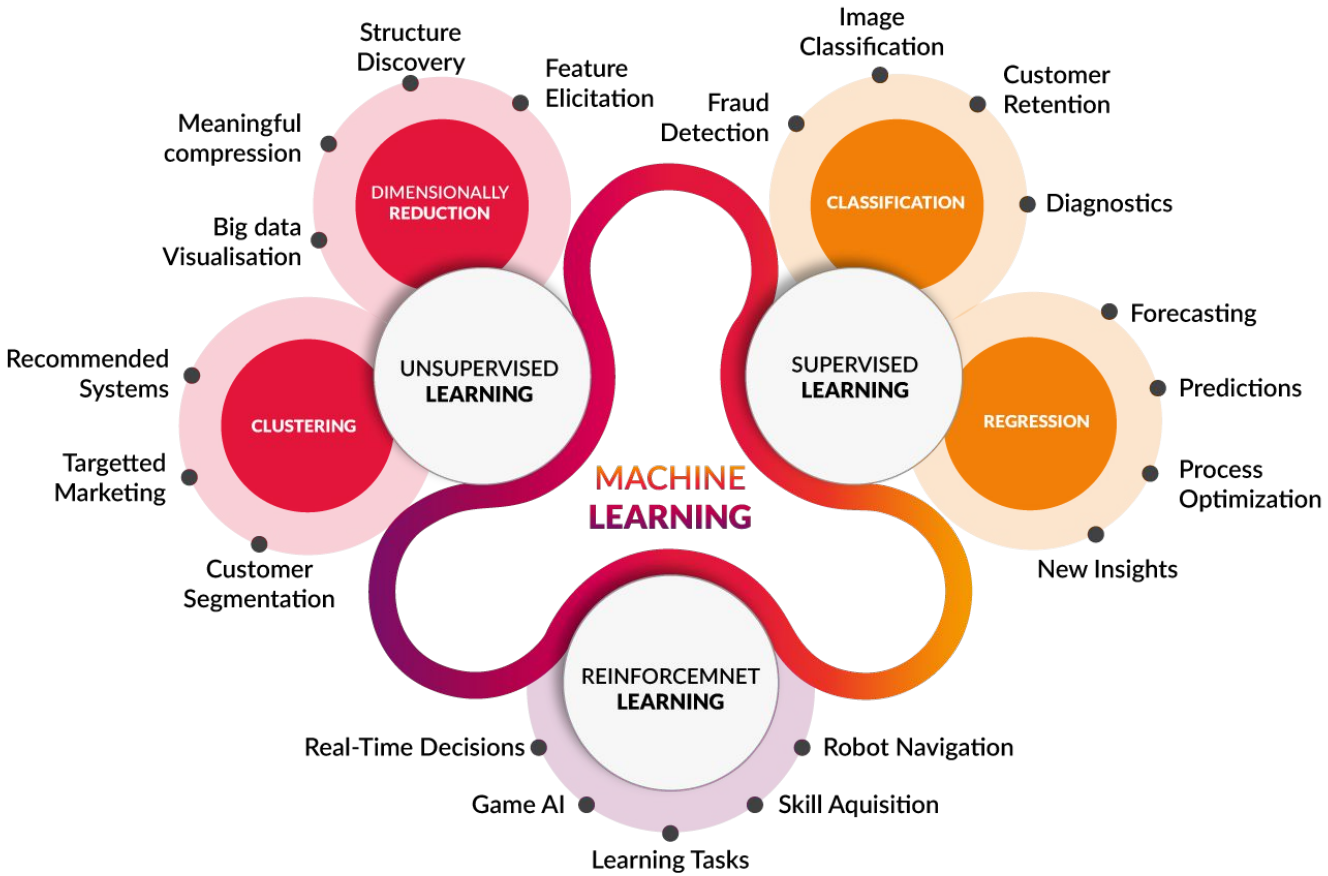
PEPI-IA

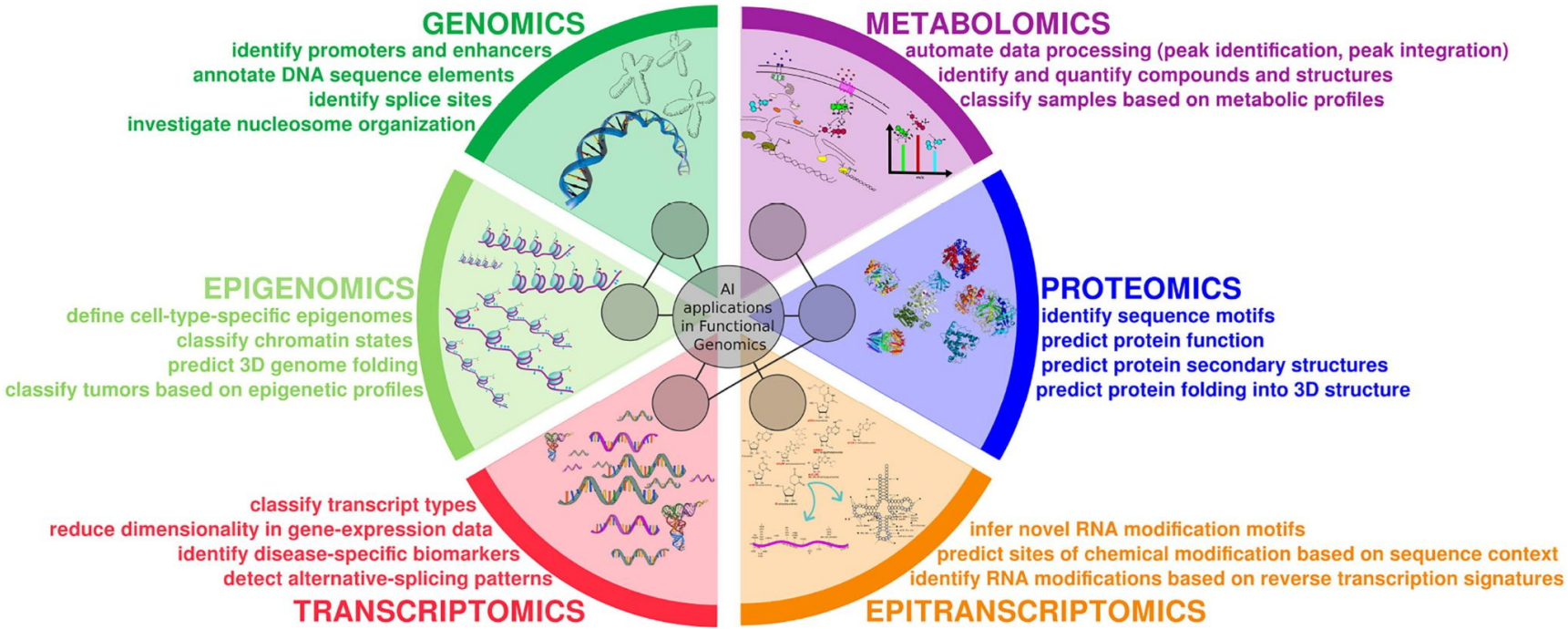
Pôle thématique sur l'IA appliquée à la génomique

Amandine Velt et Fabrice Legeai



http://www.cognub.com/index.php/cognitive-platform/

AI applications in functional genomics

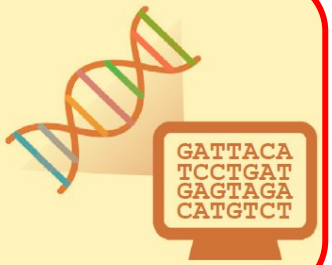


https://www.sciencedirect.com/science/article/pii/S2001037021004311

ML in Agricultural and Environment sciences

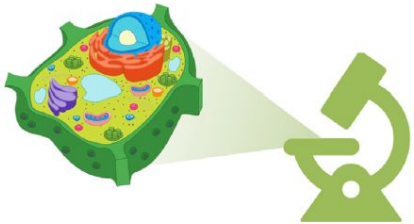
SEQUENCING

- DNA sequence
- DNA modifications (e.g. methylation)
- Transcription factor binding (e.g. ChIP-seq)
- RNA expression levels (transcriptomics)



MICROSCOPY

- Light, fluorescence, electron, flow



MASS SPECTROMETRY

- Protein / metabolite quantification
- Protein modifications
- Protein(-ligand) interactions

MACHINE LEARNING

GENOMIC REGIONS	GENE FUNCTION/REGULATION	OTHER
• Pseudogenes	• Promoter activity	• Single-cell RNA sequencing
• Crossovers	• Regulatory networks	• Metabolic pathways
• Selective sweeps	• Candidates for traits	• Multi-omics

iScience

CellPress
OPEN ACCESS

Review

Machine learning in plant science and plant breeding

Aalt Dirk Jan van Dijk,[1,2,*] Gert Kootstra,[3] Willem Kruijer,[2] and Dick de Ridder[1]

SUMMARY

Technological developments have revolutionized measurements on plant genotypes and phenotypes, leading to routine production of large, complex data sets. This has led to increased efforts to extract meaning from these measurements and to integrate various data sets. Concurrently, machine learning has rapidly evolved and is now widely applied in science in general and in plant genotyping and phenotyping in particular. Here, we review the application of machine learning in the context of plant science and plant breeding. We focus on analyses at different phenotype levels, from biochemical to yield, and in connecting genotypes to these. In this way, we illustrate how machine learning offers a suite of methods that enable researchers to find meaningful patterns in relevant plant data.

Applications - GWAS

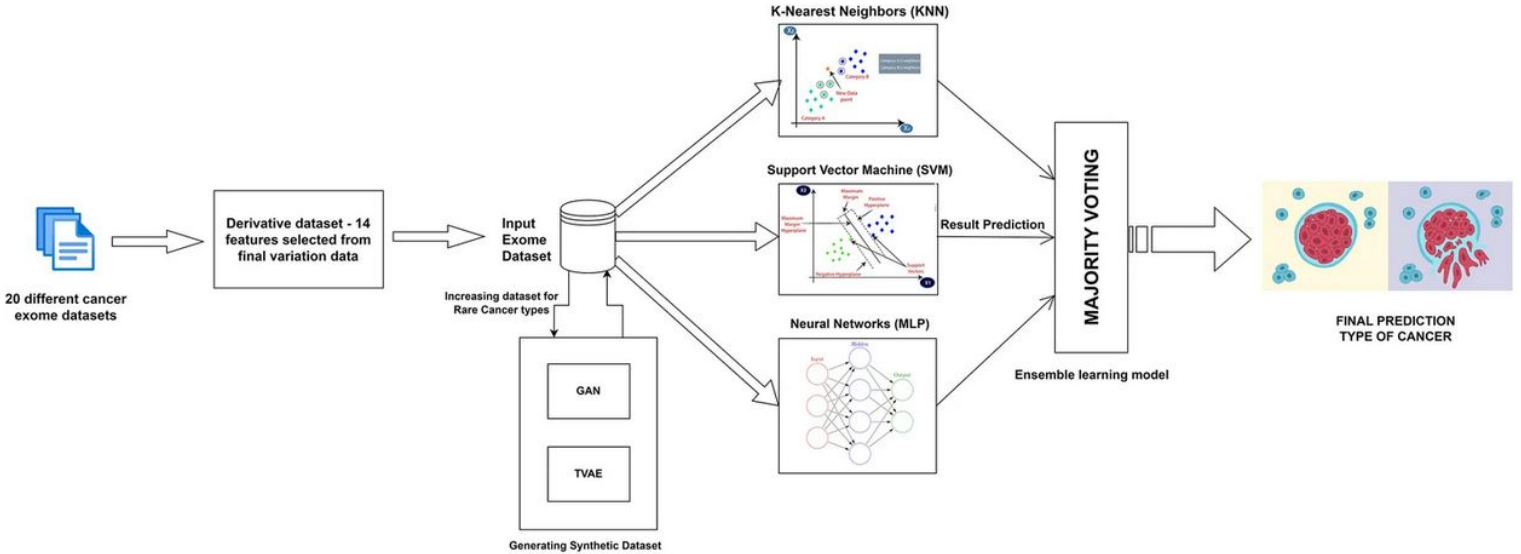
BMC Bioinformatics

Home About Articles Submission Guidelines Join The Board

Research | Open Access | Published: 18 November 2022

Implementation of ensemble machine learning algorithms on exome datasets for predicting early diagnosis of cancers

Abdu Rehaman Pasha Syed, Rahul Anbalagan, Anagha S. Setlur, Chandrashekar Karunakaran, Jyoti Shetty, Jitendra Kumar & Vidya Niranjan



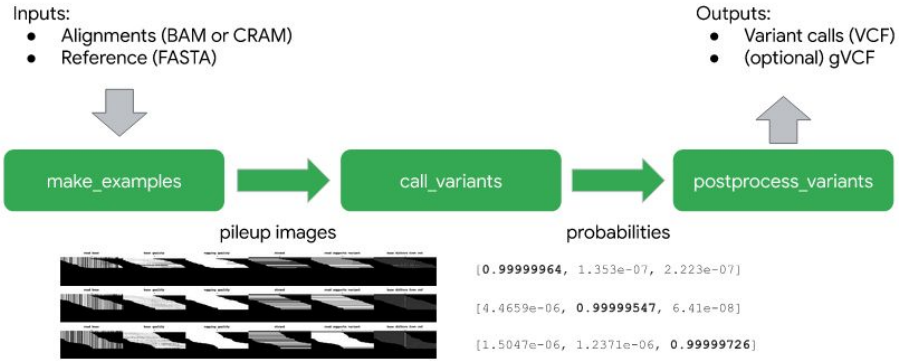
Applications - Variation prediction

LETTERS

nature biotechnology

A universal SNP and small-indel variant caller using deep neural networks

Ryan Poplin[1,2], Pi-Chuan Chang[2], David Alexander[2], Scott Schwartz[2], Thomas Colthurst[2], Alexander Ku[2], Dan Newburger[1], Jojo Dijamco[1], Nam Nguyen[1], Pegah T Afshar[1], Sam S Gross[1], Lizzie Dorfman[1,2], Cory Y McLean[1,2] & Mark A DePristo[1,2]



nature methods

Explore content · About the journal · Publish with us · Subscribe

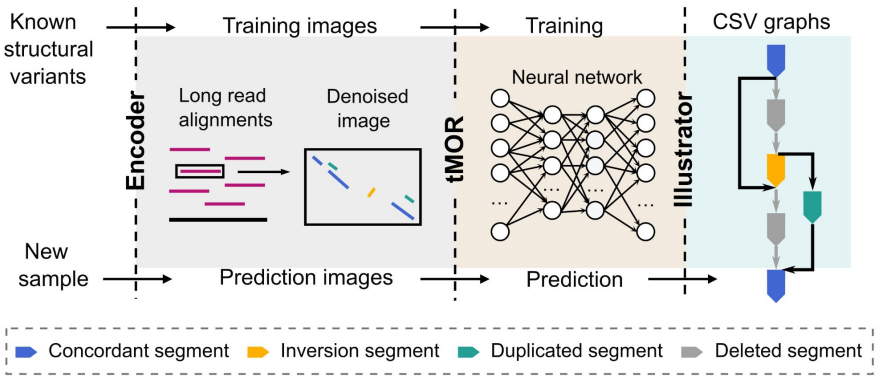
nature › nature methods › brief communications › article

Brief Communication | Published: 16 September 2022

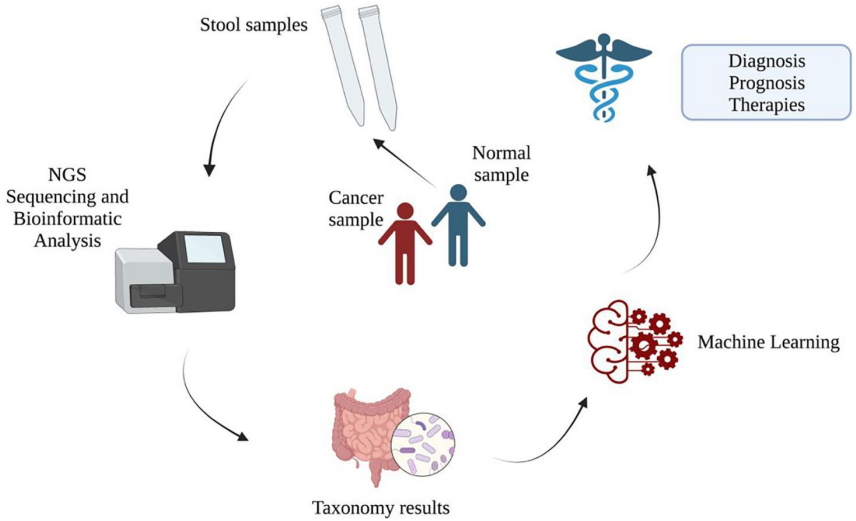
SVision: a deep learning approach to resolve complex structural variants

Jiadong Lin, Songbo Wang, Peter A. Audano, Deyu Meng, Jacob I. Flores, Walter Kosters, Xiaofei Yang, Peng Jia, Tobias Marschall, Christine R. Beck & Kai Ye

Nature Methods **19**, 1230–1233 (2022) | Cite this article



Metagenomics



frontiers in Genetics

REVIEW
published: 25 June 2019
doi: 10.3389/fgene.2019.00579

A Review and Tutorial of Machine Learning Methods for Microbiome Host Trait Prediction

Yi-Hui Zhou[1]* *and Paul Gallins*[2]

[1] *Department of Biological Sciences, North Carolina State University, Raleigh, NC, United States,* [2] *Bioinformatics Research Center, North Carolina State University, Raleigh, NC, United States*

ELSEVIER

Life Sciences
Volume 311, Part A, 15 December 2022, 121118

Review article

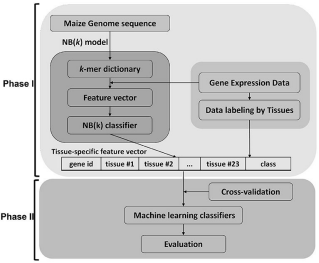
The influence of machine learning technologies in gut microbiome research and cancer studies - A review

Tamizhini Loganathan, George Priya Doss C

Expression data

frontiers | Frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 26 May 2022
doi: 10.3389/frai.2022.830170



Predicting Tissue-Specific mRNA and Protein Abundance in Maize: A Machine Learning Approach

*Kyoung Tak Cho[1], Taner Z. Sen[2] and Carson M. Andorf[3]**

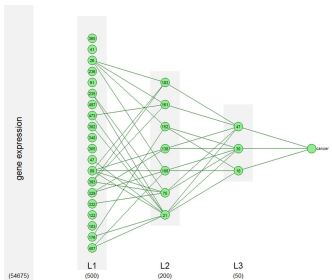
Hanczar *et al. BMC Bioinformatics* (2020) 21:501
https://doi.org/10.1186/s12859-020-03836-4

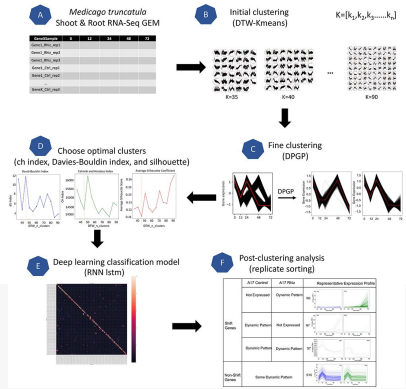
BMC Bioinformatics

RESEARCH ARTICLE Open Access

Biological interpretation of deep neural network for phenotype prediction based on gene expression

Blaise Hanczar*†, Farida Zehraoui†, Tina Issa and Mathieu Arles





ORIGINAL RESEARCH article
Front. Plant Sci., 07 April 2022
Sec. Plant Bioinformatics
https://doi.org/10.3389/fpls.2022.861639

Time Series Transcriptome Analysis in *Medicago truncatula* Shoot and Root Tissue During Early Nodulation

Yueyao Gao[1], Bradley Selee[2], Elise L. Schnabel[1], William L. Poehlman[1,3], Suchitra A. Chavan[1], Julia A. Frugoli[1] and Frank Alex Feltus[1,4,5*]

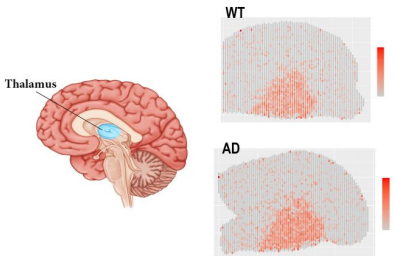
CSH PRESS GENOME RESEARCH

HOME | ABOUT | ARCHIVE | SUBMIT | SUBSCRIBE | ADVERTISE | AUTHOR INFO | CON

Institution: INRAE Inst Natl Recherche pour l'Agriculture, l'Alimentation et l'Environnemer

A model-based constrained deep learning approach for clustering spatial-resolved single-cell data

Xiang Lin[1], Le Gao[1], Nathan Whitener[2], Asraa Ahmed[3] and Zhi Wei[1,4]



Annotation

BMC Bioinformatics

Home About Articles Submission Guidelines Join The Board

Research | Open Access | Published: 06 October 2022

EnsembleSplice: ensemble deep learning model for splice site prediction

Victor Akpokiro, Trevor Martin & Oluwatosin Oluwadare

BMC Bioinformatics **23**, Article number: 413 (2022) | C

716 Accesses | **5** Altmetric | Metrics

PNAS

Identification of the expressome by machine learning on omics data

Ryan C. Sartor[a], Jaclyn Noshay[b], Nathan M. Springer[b], and Steven P. Briggs[a,1]

[a]Division of Biology, University of California San Diego, La Jolla, CA 92093; and [b]Department of Plant Biology, University of Minnesota, St. Paul, MN 55108

Contributed by Steven P. Briggs, July 11, 2019 (sent for review August 14, 2018; reviewed by James A. Birchler and Virginia Walbot)

Accurate annotation of plant genomes remains complex due to the presence of many pseudogenes arising from whole-genome duplication-generated redundancy or the capture and movement

Most researchers study the predicted genes that are derived from whole-genome annotations. These annotation approaches can be complicated by the presence of sequences with homology

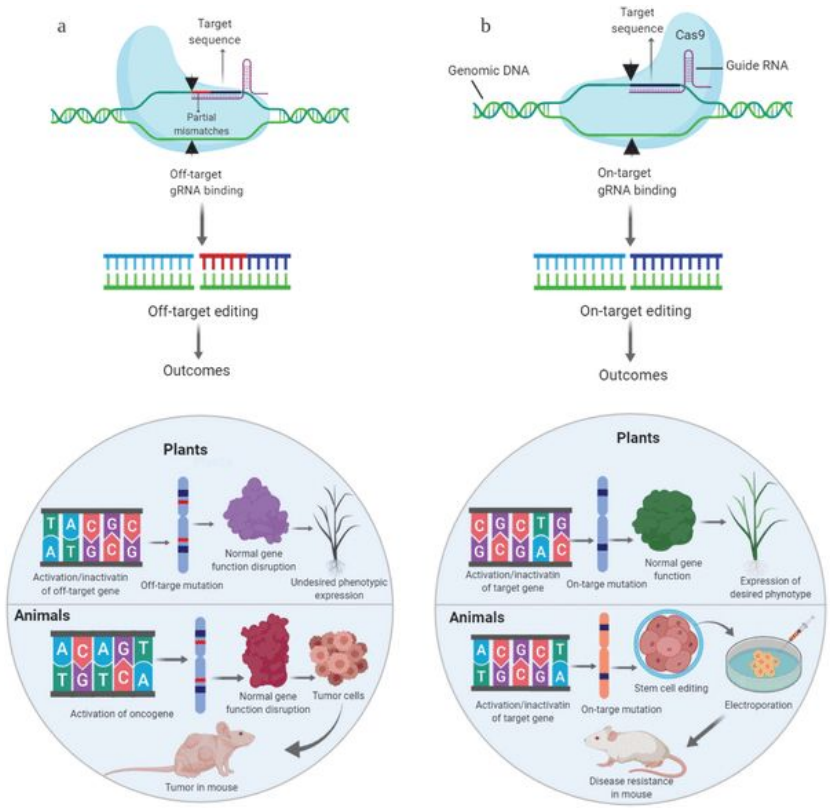
The Plant Genome OPEN ACCESS

ORIGINAL ARTICLE | Open Access

ASRpro: A machine-learning computational model for identifyi proteins associated with multiple abiotic stress in plants

Prabina Kumar Meher, Tanmaya Kumar Sahu, Ajit Gupta, Anuj Kumar, Sachin Rustgi

Genome editing



https://www.mdpi.com/2073-4409/9/7/1608

Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs

Jennifer Listgarten, Michael Weinstein, Benjamin P. Kleinstiver, Alexander A. Sousa, J. Keith Joung, Jake Crawford, Kevin Gao, Luong Hoang, Melih Elibol, John G. Doench & Nicolo Fusi

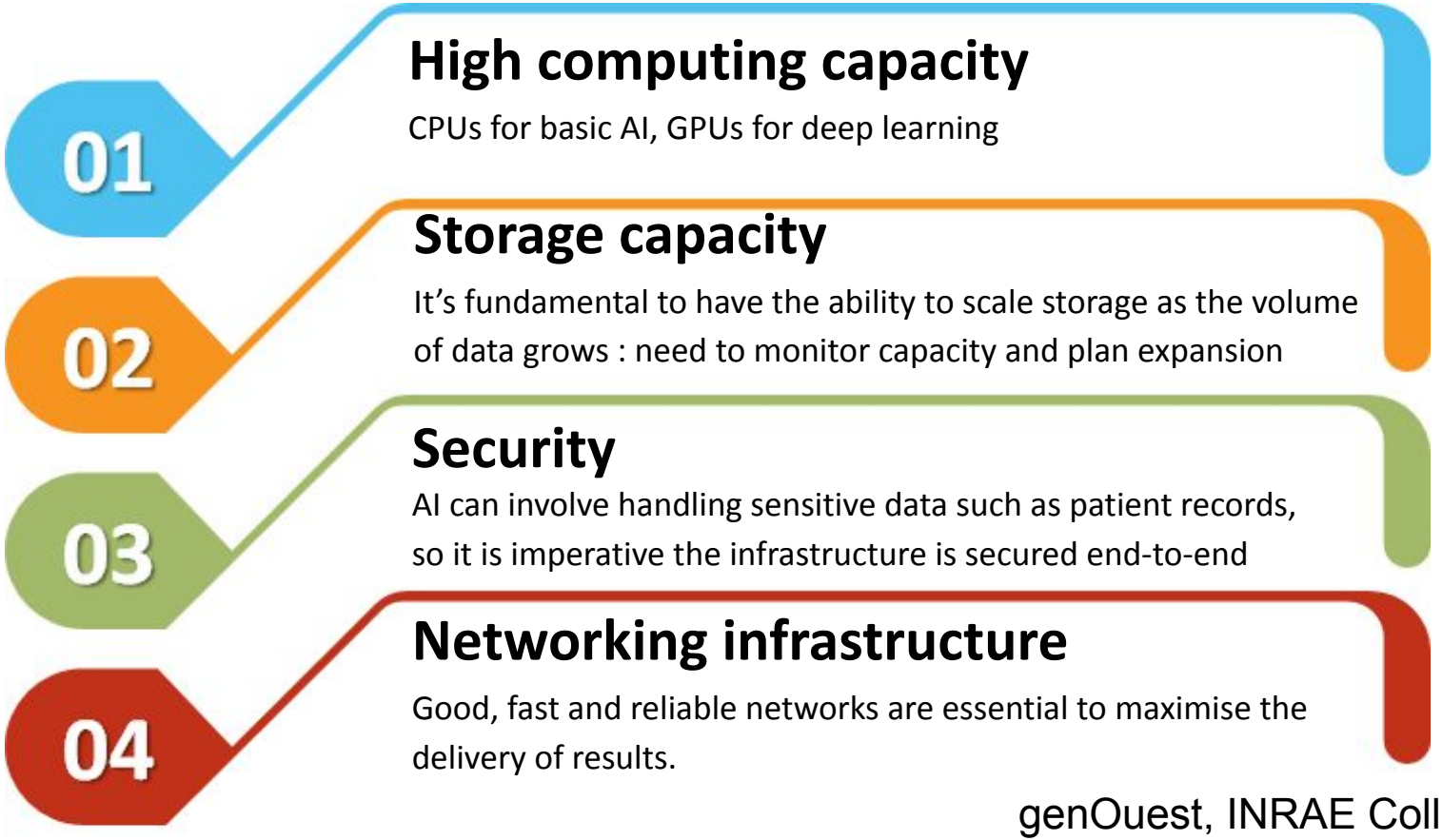
Nature Biomedical Engineering **2**, 38–47 (2018) | Cite this article

6319 Accesses | **131** Citations | **128** Altmetric | Metrics

Abstract

Off-target effects of the CRISPR–Cas9 system can lead to suboptimal gene-editing outcomes and are a bottleneck in its development. Here, we introduce two interdependent machine-learning models for the prediction of off-target effects of CRISPR–Cas9. The approach, which we named Elevation, scores individual guide–target pairs, and also aggregates them into a single, overall summary guide score. We demonstrate that Elevation consistently outperforms competing approaches on both tasks. We also introduce an evaluation method that balances errors between active and inactive guides, thereby encapsulating a range of practical use cases. Because of the large-scale and computational demands of the prediction of off-target activities, we have developed a fast cloud-based service (https://crispr.ml) for end-to-end guide-RNA design. The service makes use of pre-computed on-target and off-target activity prediction for every genic region in the human genome.

Infrastructure requirements for AI

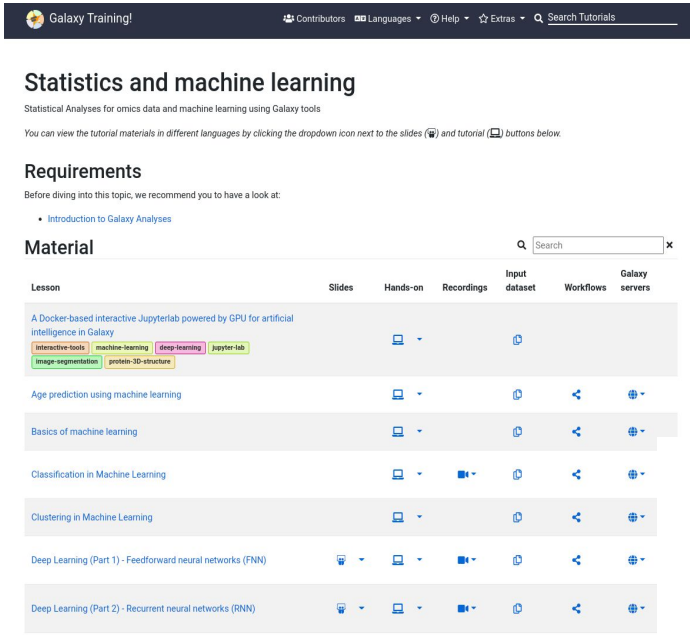


genOuest, INRAE CollabIA

https://imotep.inrae.fr/collabIA

Formations

https://training.galaxyproject.org/training-material/topics/statistics/



L'IA en sciences du vivant

Vous ne savez pas ce qu'est l'IA et vous vous demandez si elle peut vous aider à répondre à votre question biologique ? Vous développez des approches d'IA et êtes à la recherche de cas concrets pour les éprouver ? DIGIT-BIO vous invite à suivre son cycle d'animations l'IA en Sciences du Vivant.

Séance introductive - Introduction au Machine Learning - 31/03/2022

- **Digit-BIO & IA -** Christèle Robert-Granié, Marie-Laure Martin, Julien Chiquet
- **Introduction à l'IA et au Machine Learning -** Liva Ralaivola

Machine learning pour la classification supervisée - 07/04/2022

- **Comment faire des prédictions à partir de vos données biologiques grâce au Machine Learning ?** De la régression logistique aux réseaux de neurones - Blaise Hanczar
- ***Explorer le paysage cellulaire avec le Deep Learning -*** Emmanuel Moebel, Fadwa Fatmaoui (binôme biologiste - méthodologiste)

Identification de structure par classification non-supervisée - 23 /05/2022

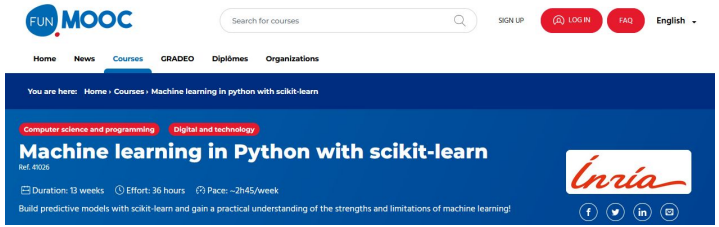
- **Comment trouver des structures dans vos données biologiques avec la classification non supervisée ?** Méthodes de distance, modèles de mélange ou réseaux de neurones - Cathy Maugis Rabusseau
- **Comprendre les interactions et l'organisation des réseaux écologiques grâce à l'analyse de réseaux -** Apprentissage non supervisé de structures de graphes - Sophie Donnet, François Massol (binôme biologiste - méthodologiste)

Identification de structure par réduction de dimension - 27/ 06/ 2022

- **De l'analyse en composantes principales aux auto-encodeurs variationnels -** Stéphanie Allassonnière
- **Quelle est la dimension pertinente de l'espace d'expression des gènes? -** Olivier Gandrillon, Franck Picard (binôme biologiste - méthodologiste)

https://www6.inrae.fr/digitbio/Animations/L-IA-en-sciences-du-vivant

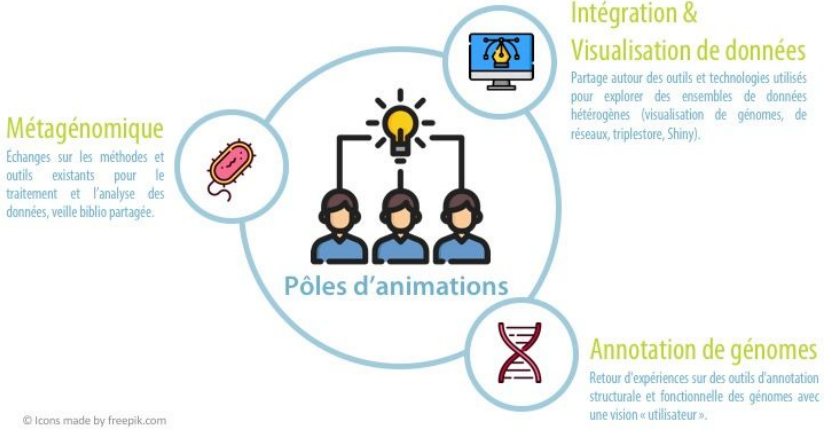
https://www.fun-mooc.fr/en/courses/machine-learning-python-scikit-learn/



PEPI-IBIS Pôle IA

Un nouveau pôle du PEPI-IBIS qui prend la suite du pôle Intégration et Visualisation de données et qui complémente

- PEPI-Annot
- PEPI-métagénomique
- PEPI-Text-mining



Les activités du pôle

Réunions régulières (1h / 2 mois)

1 exposé

- présentations de résultats, exposé de problématiques, besoins et conseils
- retours d'expérience, tests en cours (outils, méthodes)
- synthèse de publications d'intérêts
- infrastructure
- retour sur des formations
- des discussions

Journée tous les 2 ans sur un site (si financement SAPI)

- présentations générales
- ateliers
- formations

Retour lors des AG du PEPI (tous les 2 ans)

Contact

Liste de diffusion : pepi-bioinfostats-ia@inrae.fr

Site web : https://pepi-ibis.inrae.fr/ia-genomique

Orateurs

Raphaël Mourad (CBI, Université Toulouse 3, en délégation à INRAE MIAT-Mathnum) - Deep learning pour la génomique

Camille Kergal & Christophe Hitte (Equipe génétique du chien, Institute Genetics & Development, Rennes) - Méthode d'apprentissage profond pour l'analyse génomique des cancers canins comme modèles des cancers humains