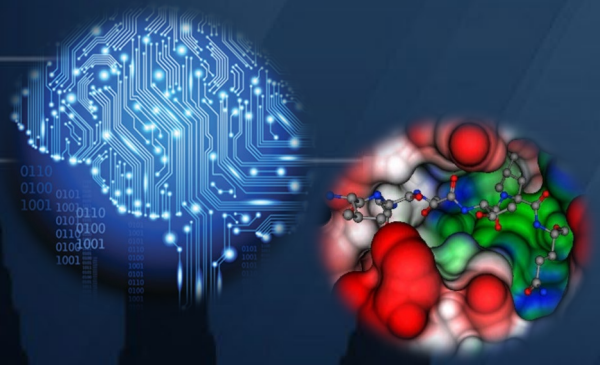
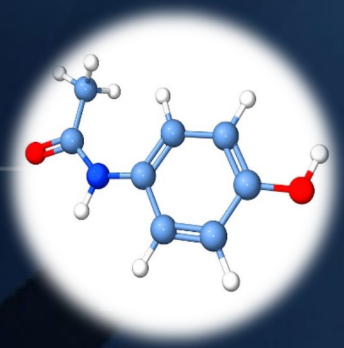
RNASA-IMEDIR
Computer Science Faculty
University of A Coruna
Spain



AI prediction of molecular properties

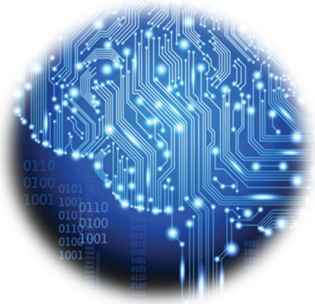
Institute of Physical Chemistry
"Ilie Murgulescu"
Bucharest, Romania, 20 July 2022



Cristian R. Munteanu
Professor, PhD
c.munteanu@udc.es
ORCID: 0000-0002-5628-2268

Research stage on "Application Artificial Intelligence in nanoparticle characterization"
(14 – 28 July 2022)

Outlines



- Short presentation: me and my group
- Models & predictions
- AI, ML & DL
- Prediction of molecular / material / system properties

RNASA-IMEDIR
Faculty of Computer Science, University of A Coruna (Galicia, Spain)

Cristian Robert Munteanu
c.munteanu@udc.es

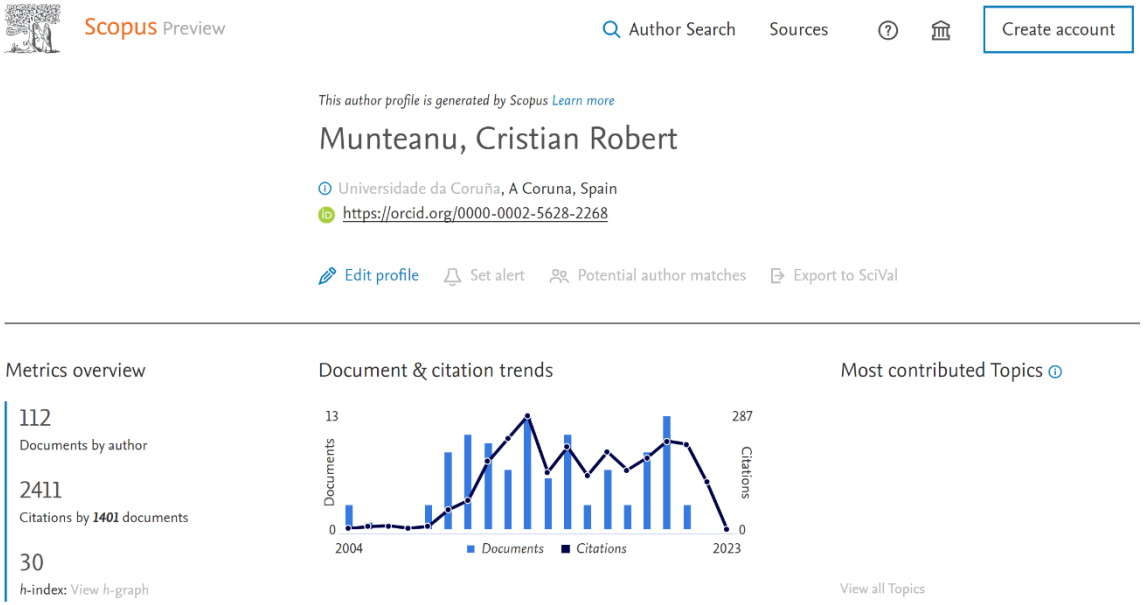
Cristian R Munteanu, PhD, Prof.

Cheminformatics Unit

Physical-Chemistry Institute
Romania
1999 – 2000

RESEARCH ASSISTANT
Chemical Kinetics Department, Romanian Academy

- Programming a Borland C/MS DOS application for graphical interpretation of Fuel/Air Mixture combustion parameters.



- Molecular information
- Drug design
- Nanoparticle, nanotubes
- AI, ML & DL
- Software development

- Participation to the Advisory Board of *The ACTION-Grid White Paper: Linking Biomedical Informatics, Grid Computing and Nanomedicine*, revised and approved by the European Commission.

AI + Chem + Bio + Pharma + Nano + Medicine

https://www.ikerdata.com
(2021)

Why models

How to obtain material / molecular properties

1) Experimental measurements
2) Direct calculation with a known formula / model
3) Training of an AI model using previous available data and use of the model to calculate properties for new data (systems)

Difficulties to obtain properties

- Experiments are too expensive or take long time (ex: properties for mixtures of millions of nanomaterials or fluids for combustion)
- Experiments are not able to measure with precision a property (ex: transition energy in rovibrational spectra of a van der Waals complex such as Ar – Cl2)
- Direct calculation that takes too much time and money (ex: interaction energies)

(1) Experiments

(ex: mixture of polymers to obtain a material with a specific electric conductance and opaque → difficile to test 1 mil combinations)

X1, X2, Y = system properties

(ex: X1, X2 = structure properties, Y=optical property

Models

Inputs			Calculate Output
X1	X2		Y1
1	1	Model Formula	2
2	2	→	4
3	3	X1 + X2 = Y	6
4	4		8
5	5		10

(2) Direct calculation

c0 + c1*X1 + c2*X2 = Y (c0, c1, c2 = coefs.)

Data + Model/formula → Results

Inputs			Output
X1	X2		Y1
1	1		2
2	2		4
3	3	→	6
4	4		8
5	5		10

(3) AI Linear Model

c0 + c1*X1 + c2*X2 = Y (c0, c1, c2 = coefs.)

↓

Data + Results → Model/Formula

X1 + X2 = Y

New data
X1, X2

Prediction
Y

Cristian Robert Munteanu
c.munteanu@udc.es

AI = Algorithm

Input & Output

Data

Numerical values

Quantification of any system: molecule, material, graph representation, network interaction, time series inputs such as spectra, sequences, images, text, etc.

Feature engineering
Feature selection

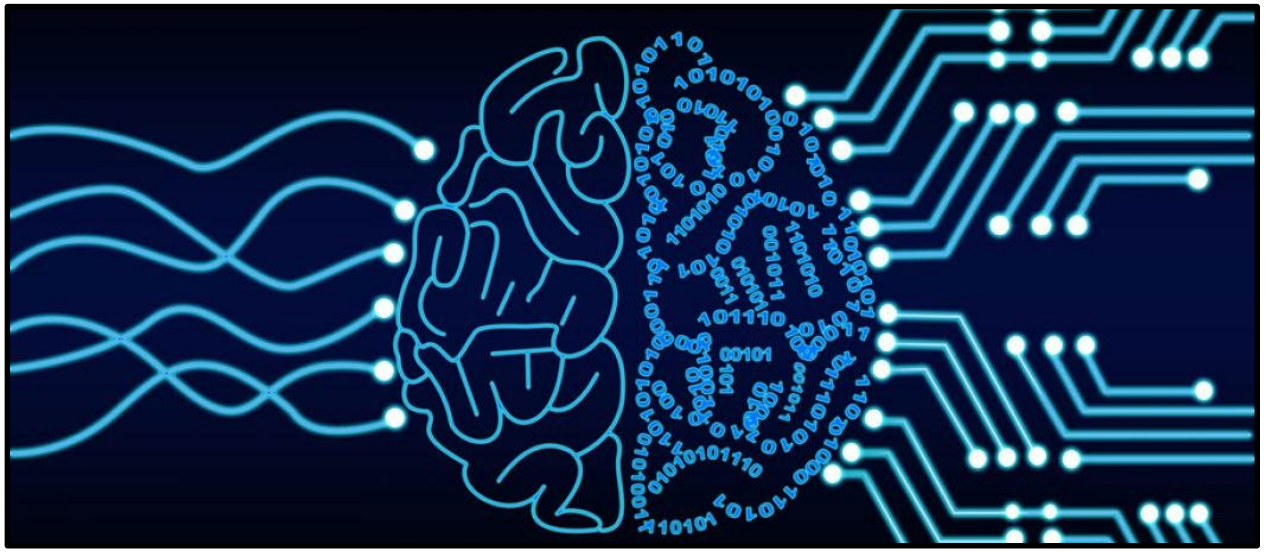
Algorithm

Model

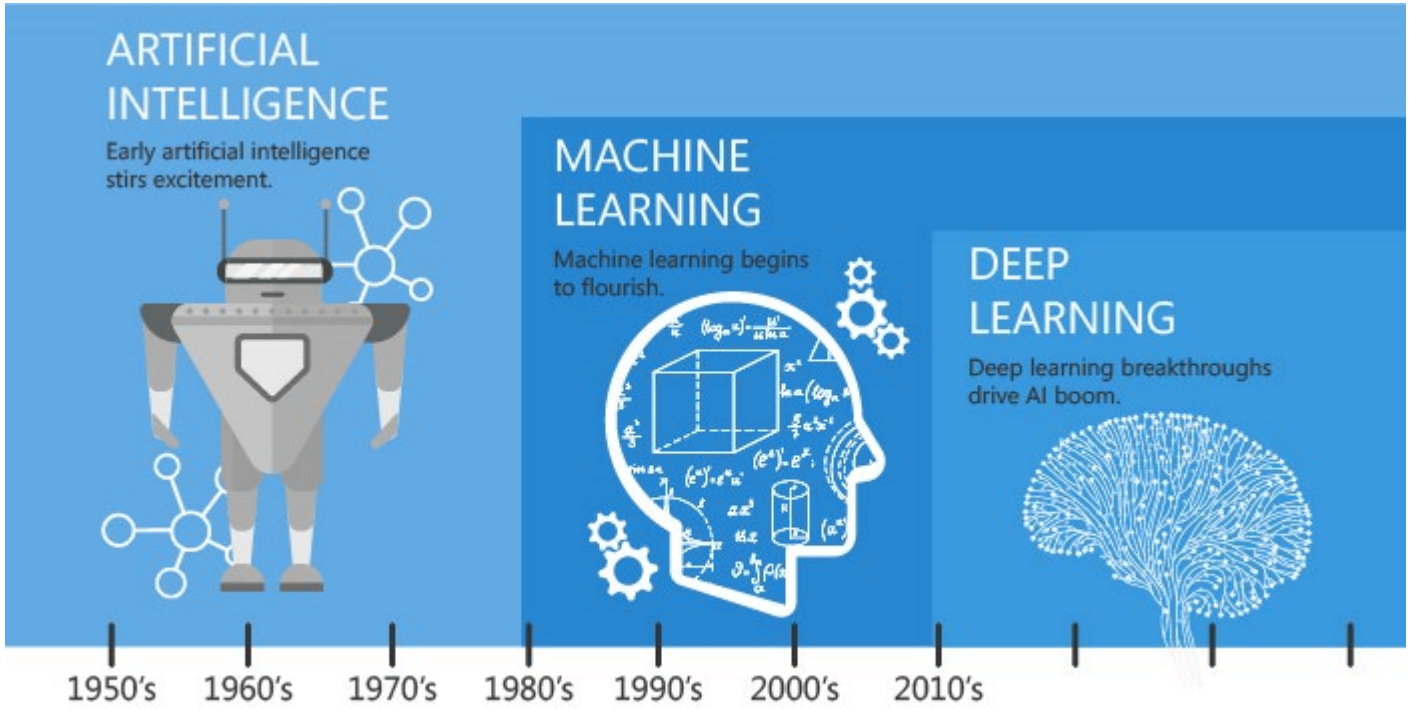
$$f(\mathbf{x})$$

Model = Mathematical formula

Cristian Robert Munteanu
c.munteanu@udc.es

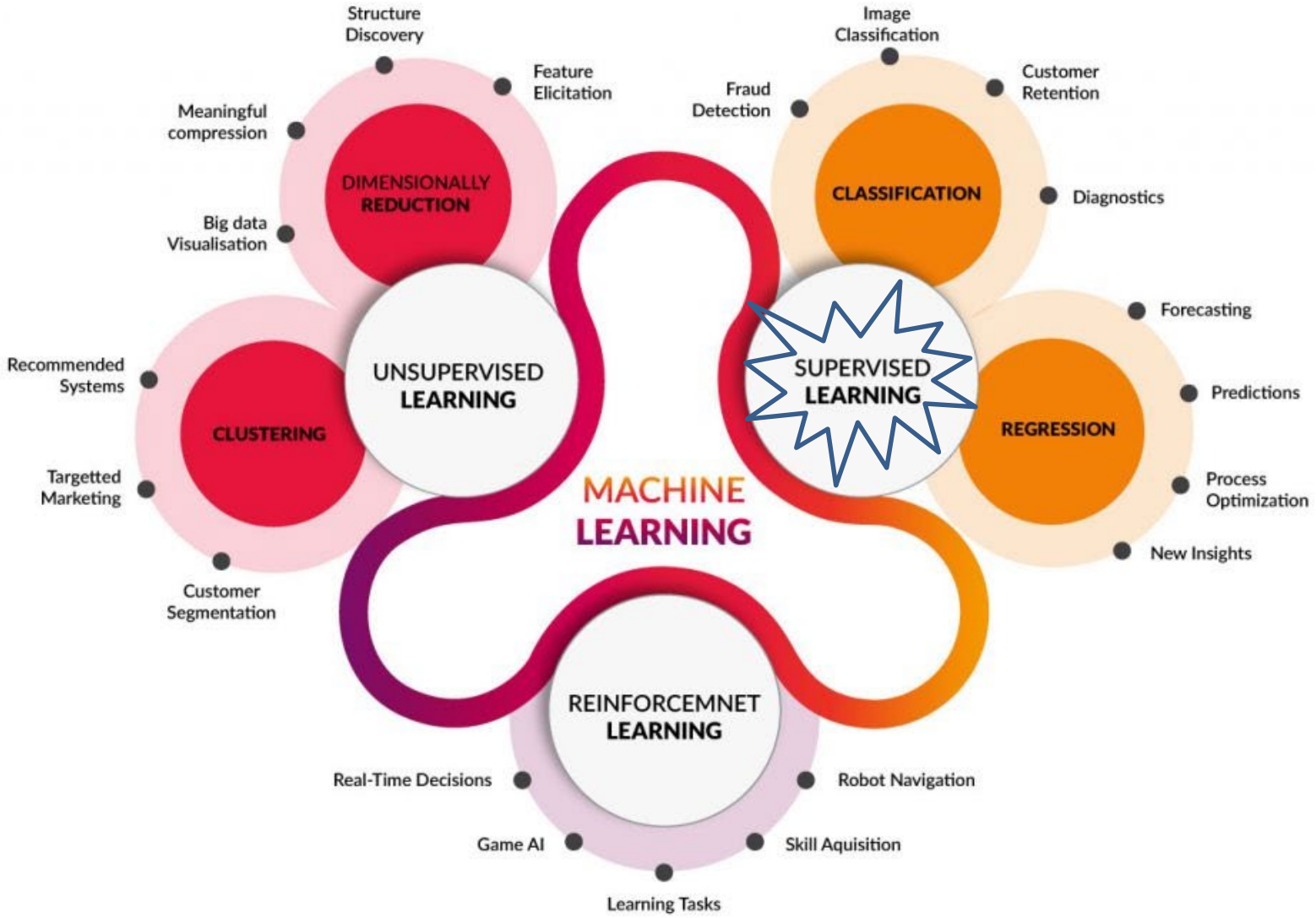


Artificial intelligence, sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and other animals.



Since an early flush of optimism in the 1950's, smaller subsets of artificial intelligence - first machine learning, then deep learning, a subset of machine learning - have created ever larger disruptions.

Machine learning is a subset of artificial intelligence in the field of computer science that often uses statistical techniques to give computers the ability to "learn" with data, without being explicitly programmed.

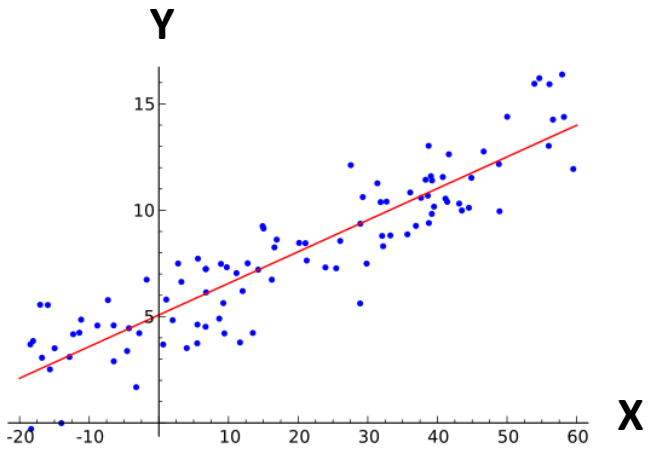

Structure Discovery
Feature Elicitation
Meaningful compression
DIMENSIONALLY REDUCTION
Big data Visualisation
Image Classification
Customer Retention
Fraud Detection
CLASSIFICATION
Diagnostics
Recommended Systems
CLUSTERING
Targetted Marketing
Customer Segmentation
UNSUPERVISED LEARNING
SUPERVISED LEARNING
Forecasting
Predictions
REGRESSION
Process Optimization
New Insights
MACHINE LEARNING
REINFORCEMNET LEARNING
Real-Time Decisions
Robot Navigation
Game AI
Skill Aquisition
Learning Tasks

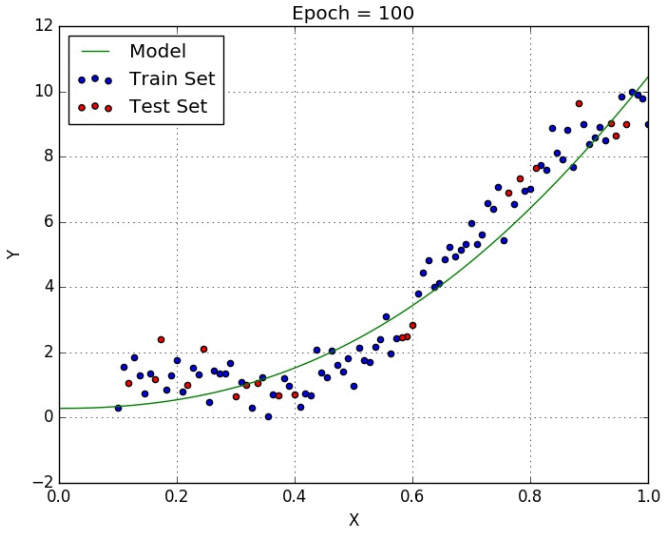
Univariate vs. multivariate

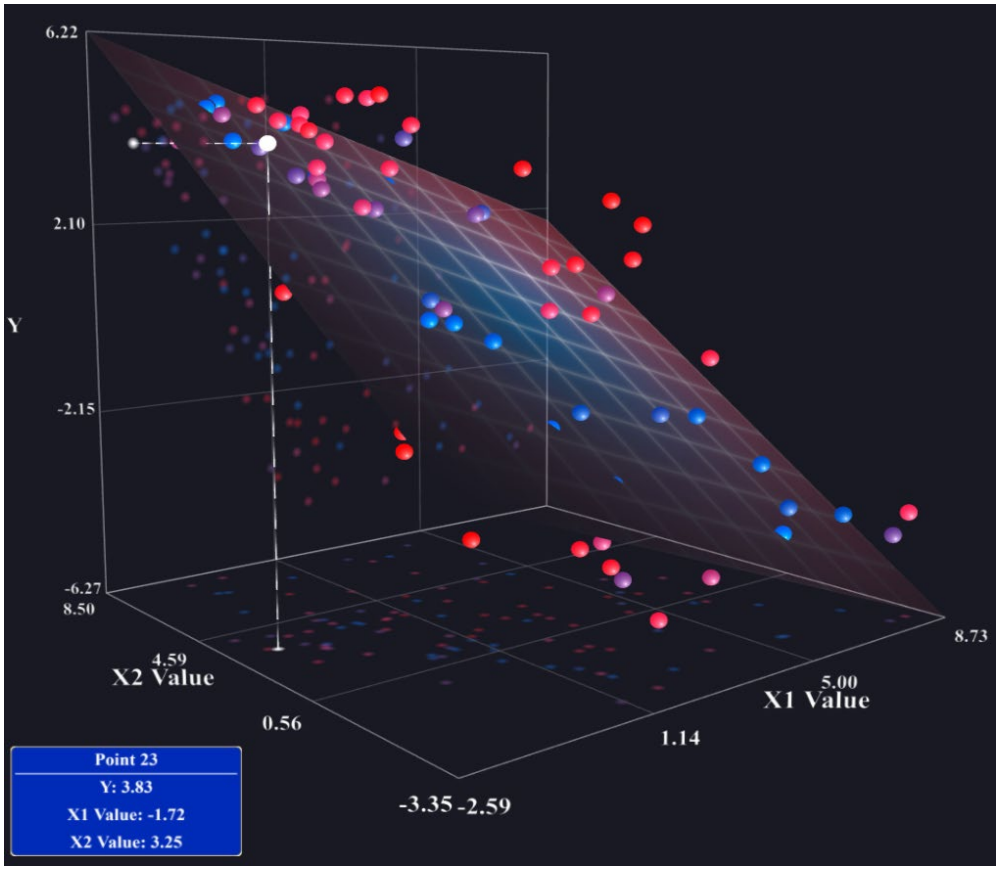
$$Y = a + bx \quad (1)$$

$$y = a + bx + \epsilon \quad (2)$$

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + \epsilon \quad (3)$$





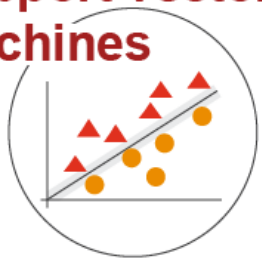


Cristian Robert Munteanu
c.munteanu@udc.es

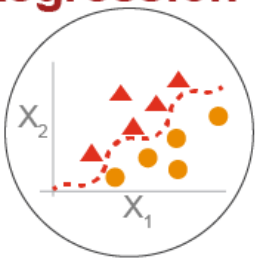
Machine learning methods

Decision trees

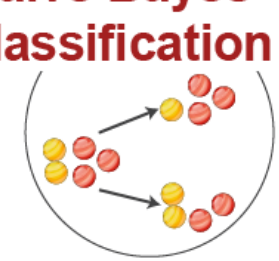
Support vector machines



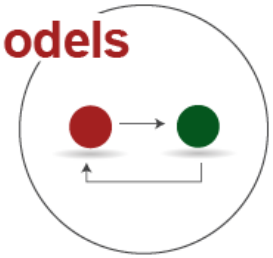
Regression



Naive Bayes classification



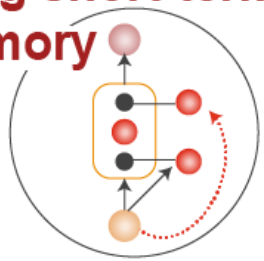
Hidden Markov models



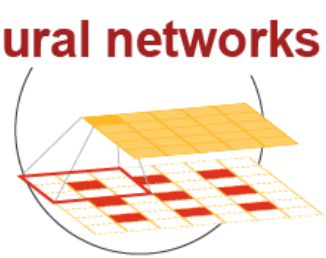
Random forest

Recurrent neural networks

Long short-term memory

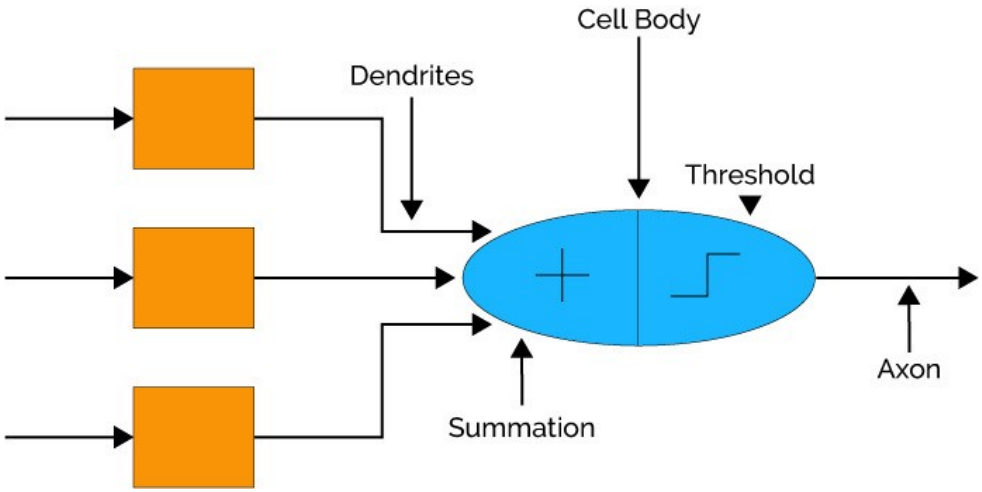


Convolutional neural networks

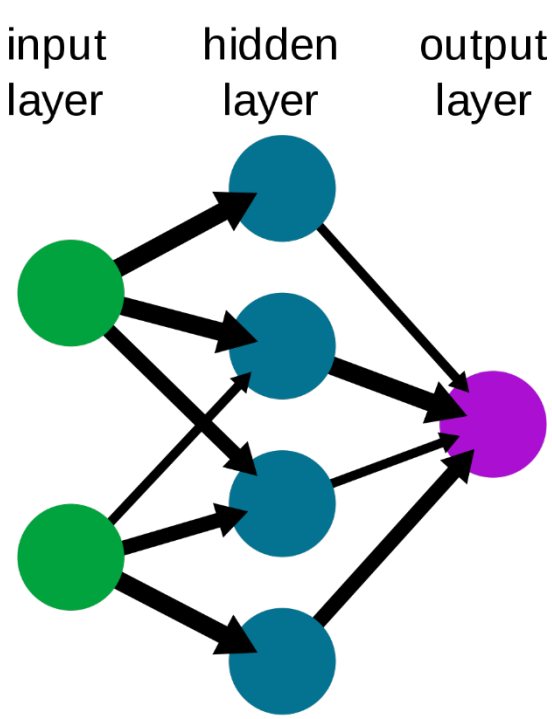


Source: pwc.com/nextintech

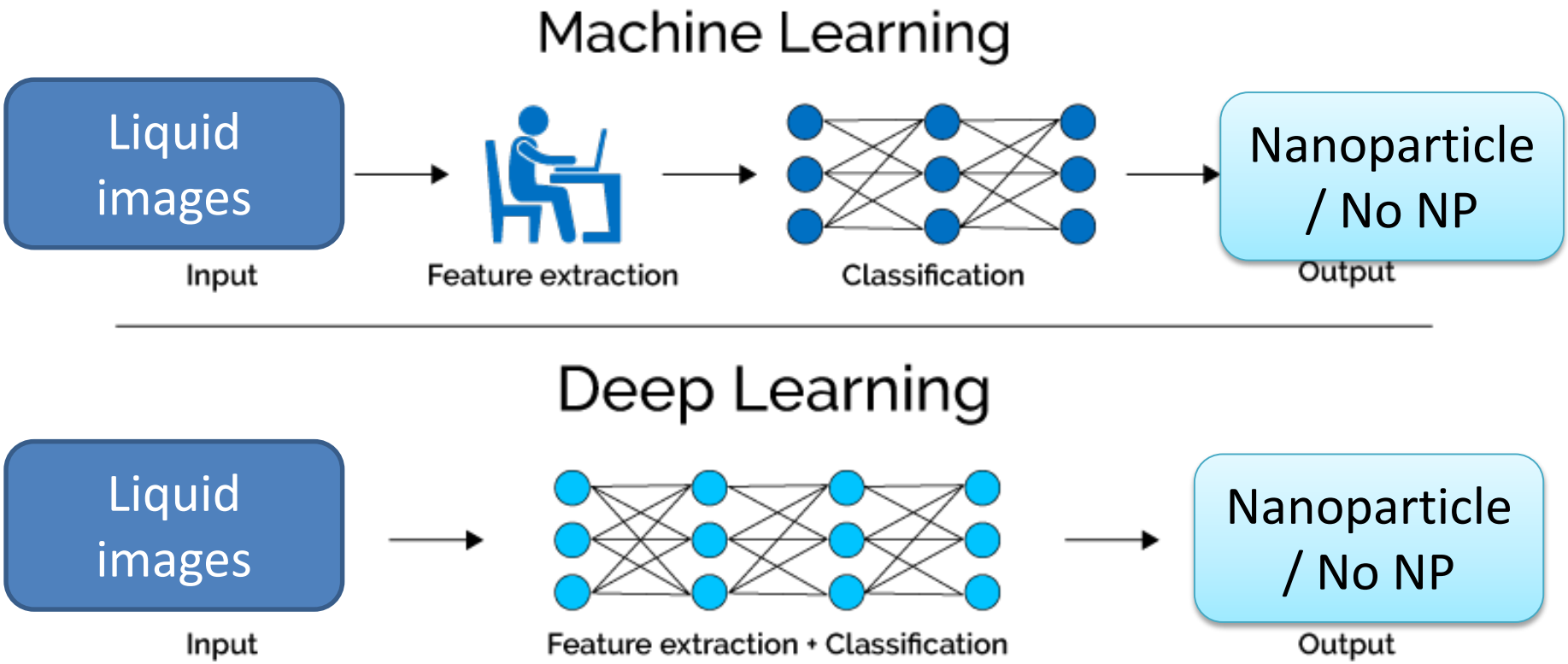
Artificial Neural Networks



A simple neural network



Cristian Robert Munteanu
c.munteanu@udc.es

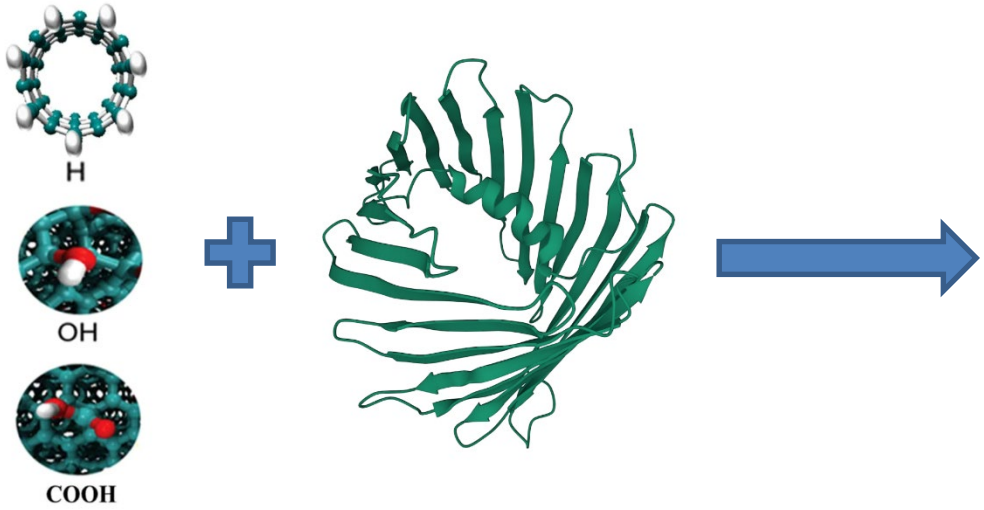


Cristian Robert Munteanu
c.munteanu@udc.es

Decrypting Strong and Weak Single-Walled Carbon Nanotubes Interactions with Mitochondrial Voltage-Dependent Anion Channels (VDAC) Using Machine Learning

Nanomaterials 2017, 7(11), 386; https://doi.org/10.3390/nano7110386

VDAC is the most abundant and highly conserved channel-protein in outer mitochondrial membrane of cells from all eukaryotic kingdoms including human (*Homo sapiens*). VDAC is the main communication route between this organelle and cytosol, generating the mitochondrial membrane potential, sucrose exchange, mitochondrial [Ca^{2+}] dynamics and ATP-efflux between these two cellular compartments



Free energy binding (FEB) of mitochondrial voltage-dependent anion channel (VDAC) with pristine, hydroxylated and carboxylated single-walled carbon nanotubes (SWCNT, SWCNT-OH and SWCNT-COOH, respectively)

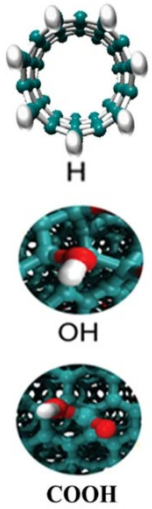
Weak vs Strong

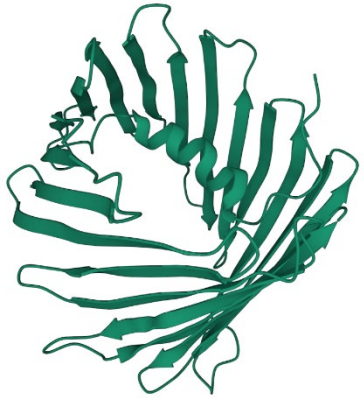
We have: Limited number of SWCNTs with the FEB energies

We need: FEB energies for a big number of new SWCNTs (time consuming, costs)

Cristian Robert Munteanu
c.munteanu@udc.es

Decrypting Strong and Weak Single-Walled Carbon Nanotubes Interactions with Mitochondrial Voltage-Dependent Anion Channels (VDAC) Using Machine Learning





Free energy binding (FEB) energies (Weak vs Strong)

SWCNT nanodescriptors

Protein target

Perturbation Operators

Perturbations of descriptors with protein target

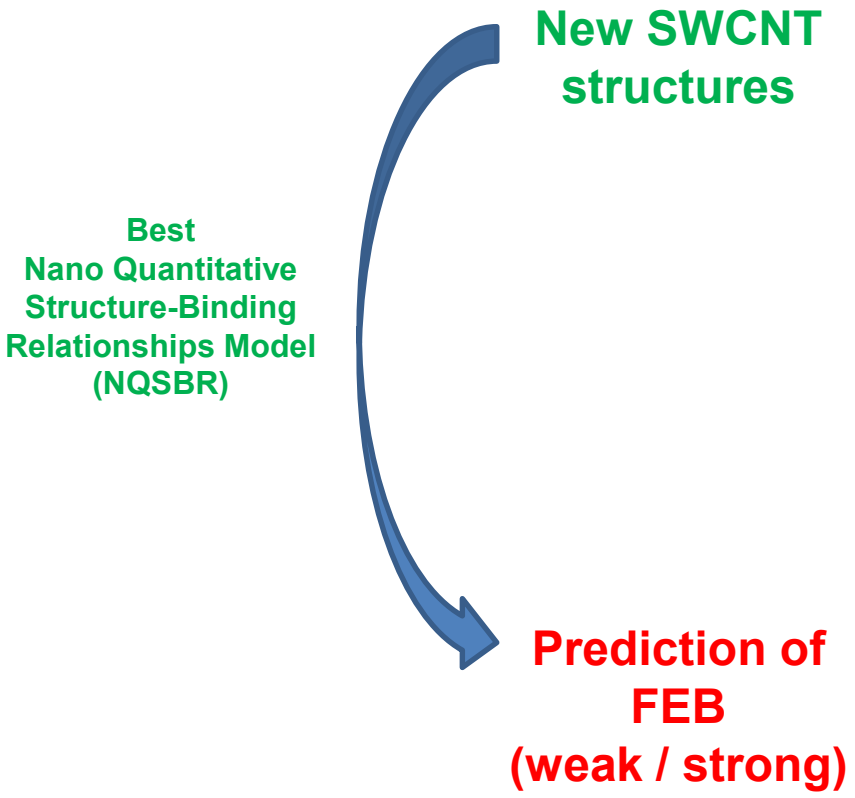
Nano Quantitative Structure-Binding Relationships Model (NQSBR)

FEB energy

Cristian Robert Munteanu
c.munteanu@udc.es

Decrypting Strong and Weak Single-Walled Carbon Nanotubes Interactions with Mitochondrial Voltage-Dependent Anion Channels (VDAC) Using Machine Learning

SWCNT nanodescriptors (V_k)
FEB (ATP) (Kcal/mol)
Tube Length (Å)
m-Hamada index
n-Hamada index
Molecular Weight (g/mol)
No. Bonds
No. Atoms
Radio Rt (nm)
Nanotube diameter: $d_{SWCNT\ (n;\ m)}$
Highest common divisor of (n, m)
Highest common divisor of (2n + m, 2 m + n)
Translation vector: T(Å)
Semi-empirical Homo-Lumo Bandgap E_g(eV)
Chiral vector: C_h (nm)
Chiral angle (θ)
mod (n−m, 3)
Semi-empirical V_{RBM} (cm^{-1})
First van Hove Singularity Optical Transitions peak-[E_{11}] (eV)
Second van Hove Singularity Optical Transitions peak-[E_{22}] (eV)
No. hexagon/unit cell (N/2)



Fast and cheap theoretical screening of any number of SWCNTs

Cristian Robert Munteanu
c.munteanu@udc.es

Other applications

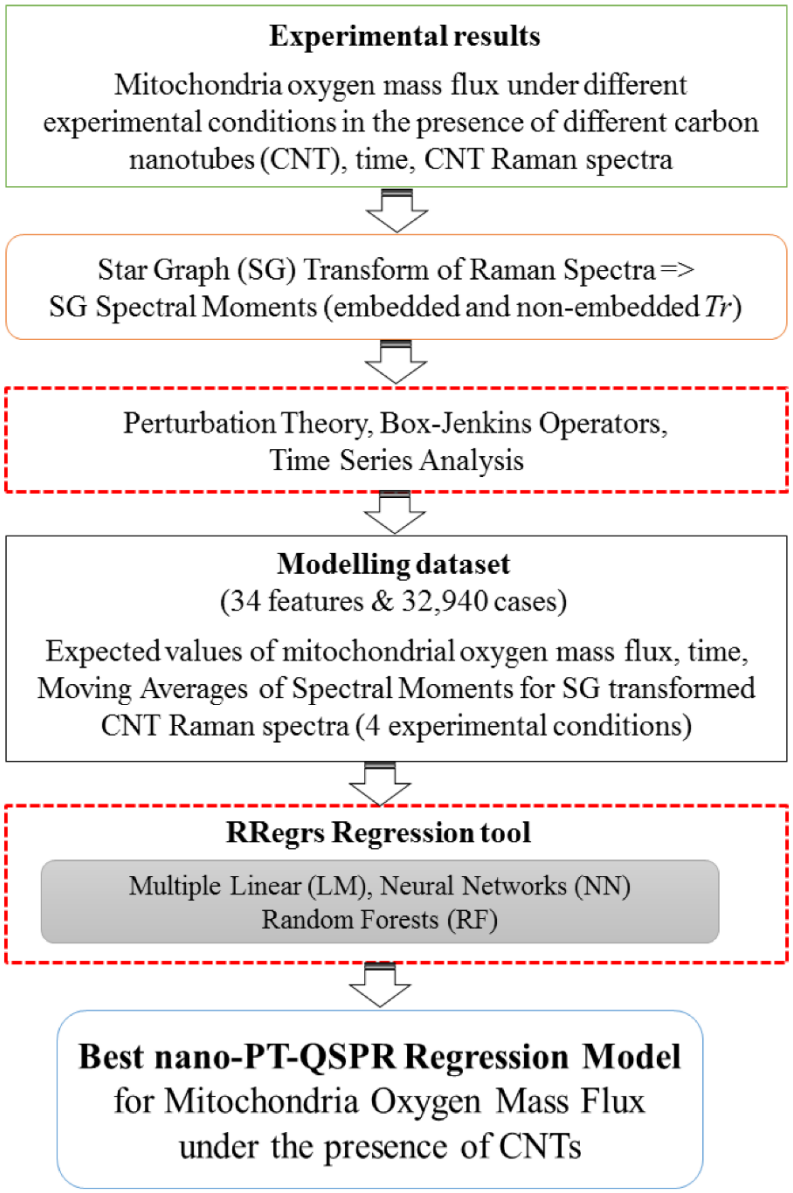
Carbon Nanotubes' Effect on Mitochondrial Oxygen Flux Dynamics: Polarography Experimental Study and Machine Learning Models using Star Graph Trace Invariants of Raman Spectra

Nanomaterials 2017, 7(11), 386; https://doi.org/10.3390/nano7110386

The impact of carbon nanotubes (CNTs) on mitochondrial oxygen mass flux (J_m) under three experimental conditions.

Model features: Nanodescriptors = Star Graph Trace Invariants of Raman Spectra

Model output: Mitochondrial Oxygen Mass Flux



RF material risk assessments

Cristian Robert Munteanu
c.munteanu@udc.es

Other applications

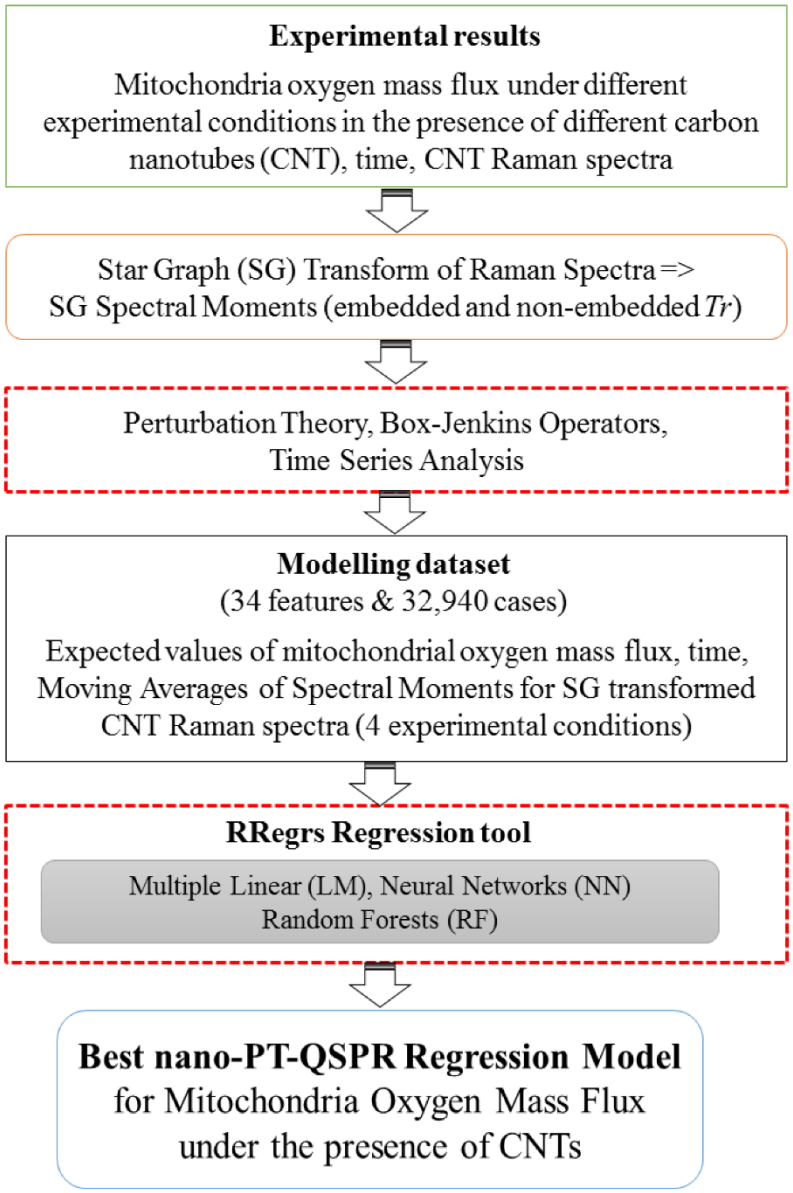
Carbon Nanotubes' Effect on Mitochondrial Oxygen Flux Dynamics: Polarography Experimental Study and Machine Learning Models using Star Graph Trace Invariants of Raman Spectra

Nanomaterials 2017, 7(11), 386; https://doi.org/10.3390/nano7110386

The impact of carbon nanotubes (CNTs) on mitochondrial oxygen mass flux (J_m) under three experimental conditions.

Model features: Nanodescriptors = Star Graph Trace Invariants of Raman Spectra

Model output: Mitochondrial Oxygen Mass Flux



RF material risk assessments

Cristian Robert Munteanu
c.munteanu@udc.es

Other applications

- ✓ Prediction of Drug-Decorated Nanoparticle Delivery Systems

Int. J. Mol. Sci. **2021**, *22*(21), 11519; https://doi.org/10.3390/ijms222111519

- ✓ Theoretical search/screening of molecule/materials/polymers/nanoparticles with a specific property (biological activity, optical, mechanical, reactive, etc.)

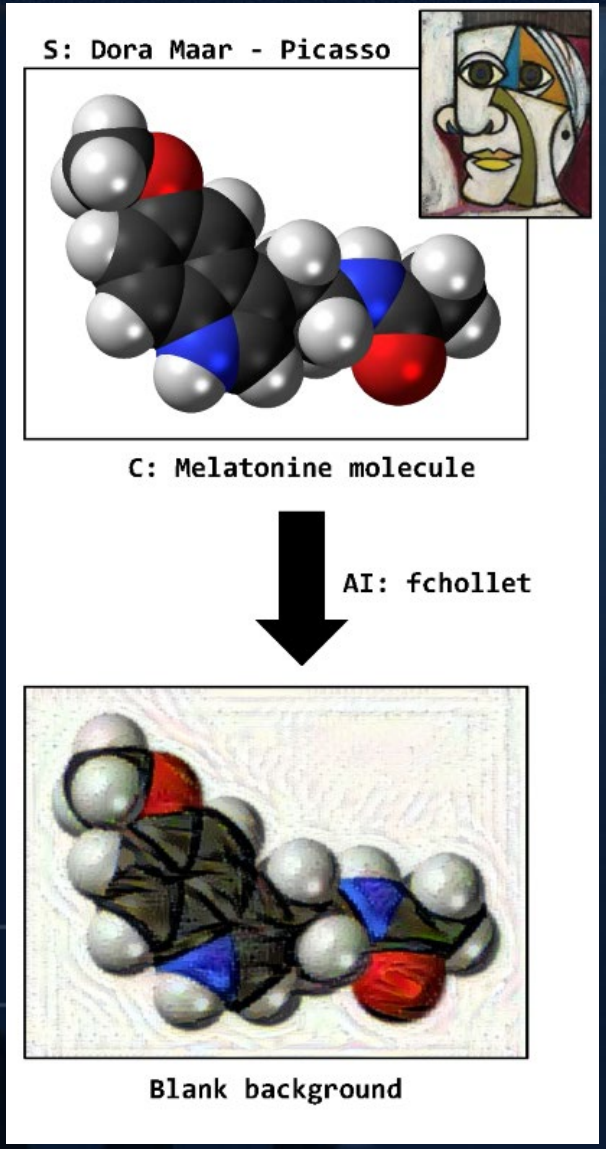
Front. Pharmacol., **2021**, 12:598925, https://doi.org/10.3389/fphar.2021.598925
Molecules, **2020**, 25, 5172, https://www.mdpi.com/1420-3049/25/21/5172

- ✓ Prediction of fuel mixtures properties (ex: searching for optimal fuel burning)
- ✓ Evaluation of material / molecular toxicity in order to replace or reduce the animal experimentation
- ✓ Generation of toxicity standards based on rule-based models
- ✓ Prediction of reaction efficacy or final product's properties

- ✓ Deep learning for images
  - Automatic detection of material defects or properties
  - Detection of specific nanoparticles / nanorobots in fluids
  - Sensor development based on image detection (ex: air composition / toxicity based on images)

  https://play.google.com/store/apps/details?id=com.xlandc.polypdetect

RNASA – IMEDIR
Computer Science Faculty
University of A Coruna
Spain



AI prediction of molecular properties

Institute of Physical Chemistry
"Ilie Murgulescu"
Bucuresti, 20 July 2022

Cristian Robert Munteanu
c.munteanu@udc.es