Understanding Charge Radii with Machine Learning

B. Maheshwari, P. Van Isacker

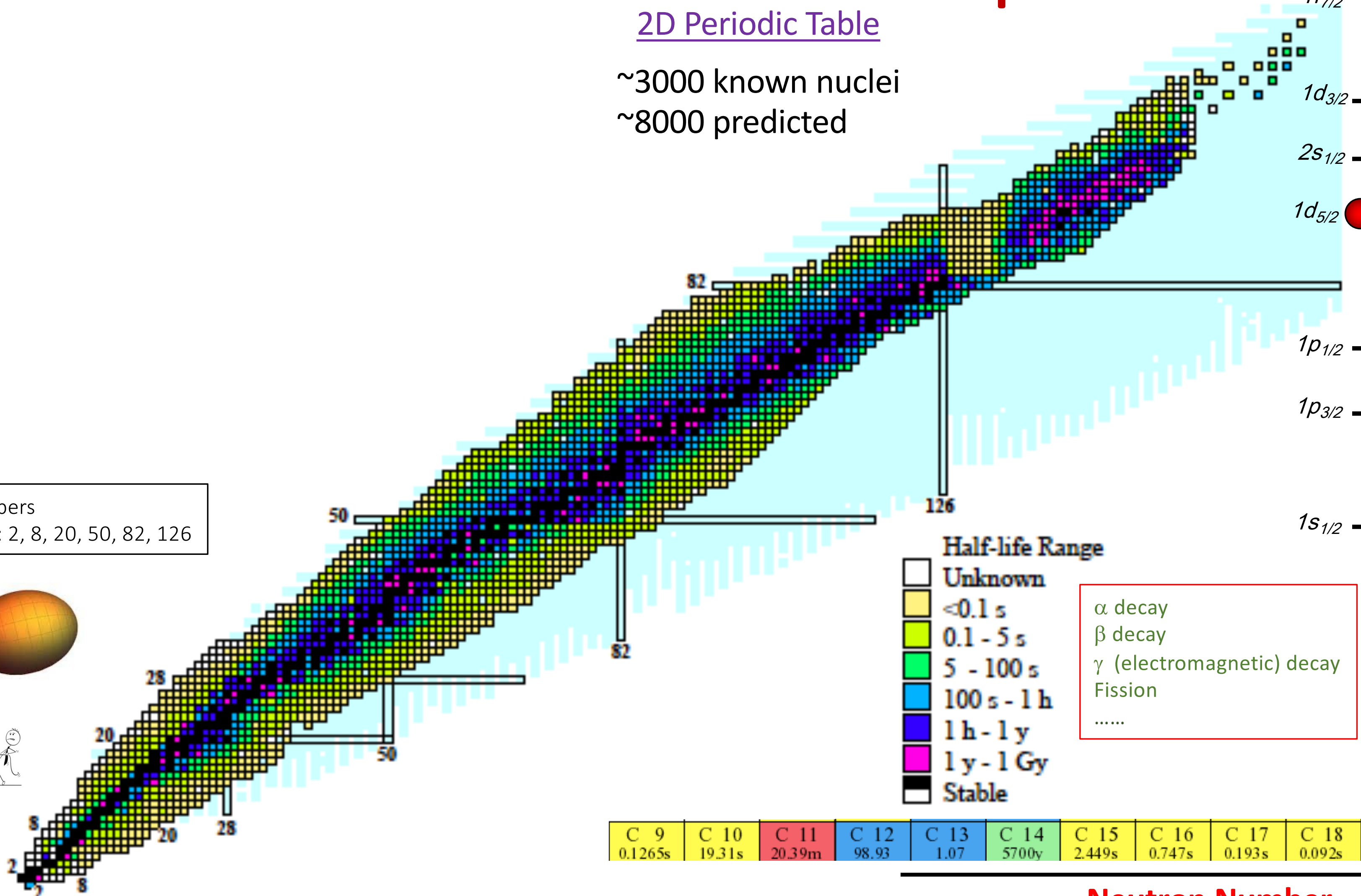
GANIL

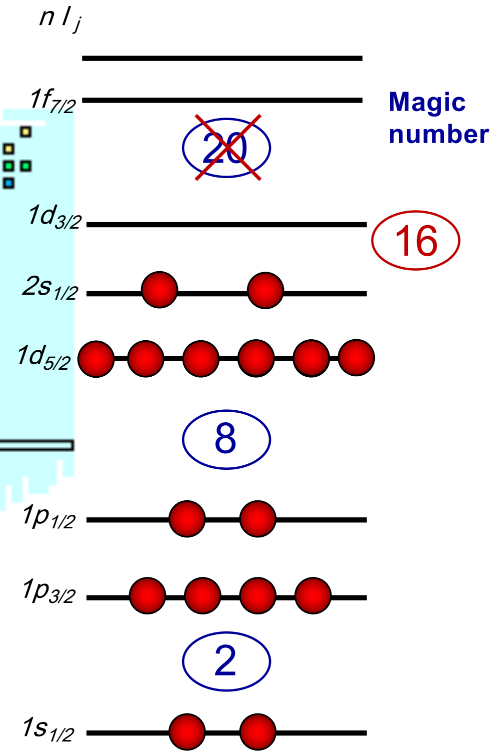
Nuclear Landscape

2D Periodic Table

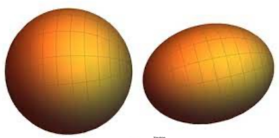
~3000 known nuclei
~8000 predicted

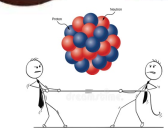
Proton Number



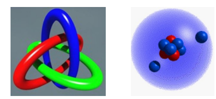


"Magic" numbers
Closed shells : 2, 8, 20, 50, 82, 126





α decay
β decay
γ (electromagnetic) decay
Fission
……



C 9	C 10	C 11	C 12	C 13	C 14	C 15	C 16	C 17	C 18	C 19	C 20		C 22
0.1265s	19.31s	20.39m	98.93	1.07	5700y	2.449s	0.747s	0.193s	0.092s	0.049s	0.014s		0.0061s

Neutron Number

Physics-driven Supervised Machine-Learning

Hidden Correlations to Expressions

(ML)

Hybrid Numerical-Symbolic Regression

Small and skewed nuclear physics data sets

Literature: A single random split for such small datasets

Unreliable or over-optimistic predictions and are sensitive to random seed

Physics-driven Supervised Machine-Learning

Hidden Correlations to Expressions

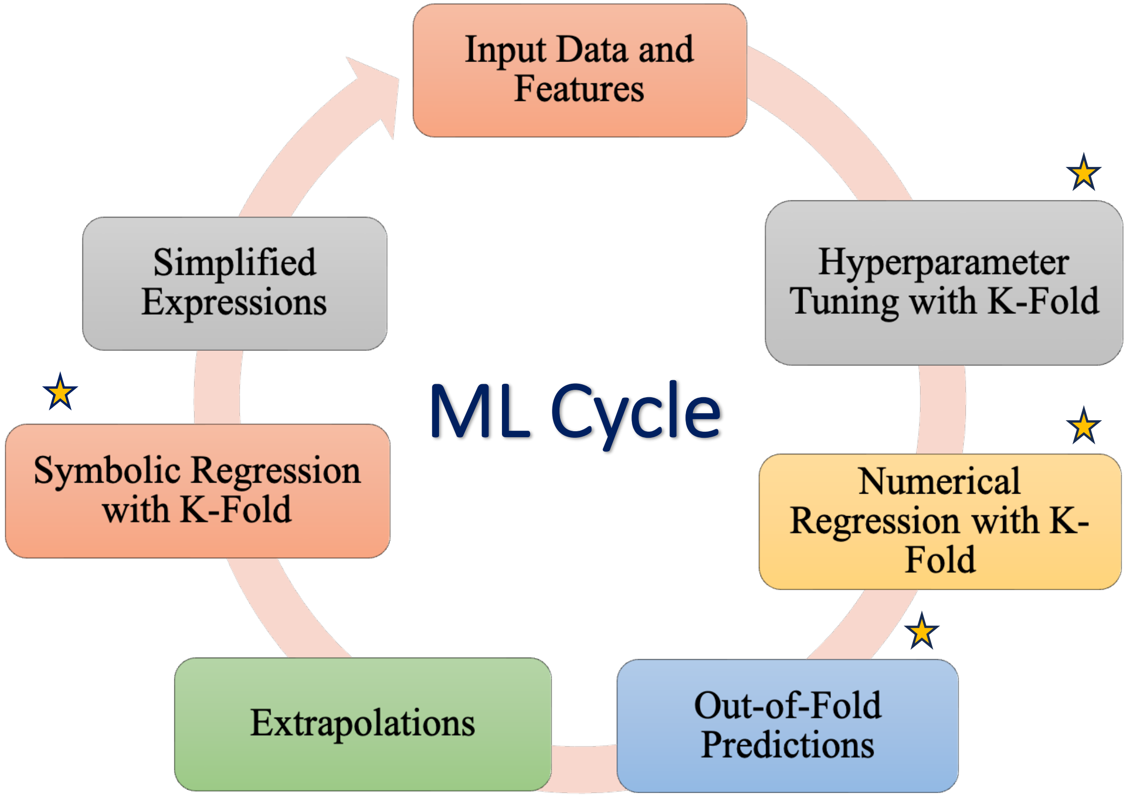
(ML)

Hybrid Numerical-Symbolic Regression

Small and skewed nuclear physics data sets

Literature: A single random split for such small datasets

Unreliable or over-optimistic predictions and are sensitive to random split



New developments in ML applications to Nuclear Physics

Input Data and Features

Physics-driven Features:

Proton number dependence $Z^{1/3}$,

Mass number dependence $A^{1/3}$,

Neutron number N,

Total binding energy BE,

Isospin asymmetry parameter $I = (N-Z)/A$,

Casten factor $CF = (N_p N_n)/(N_p + N_n)$

and Pairing gap P

Observable:

Charge-radii

Automating Hyperparameter Tuning with Optuna

Recipe for the model

Manually finding the best model settings (e.g., learning_rate, n_estimators) is slow and inefficient.

Optuna automates this search.

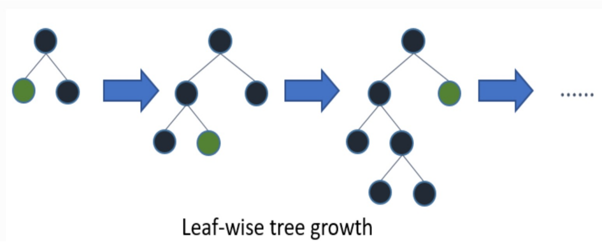
It intelligently "learns" from past results to find the best settings (hyperparameters) faster.

To *minimize RMSE* by finding the optimal LGBM and GPR parameters.

LGBM search space
n_estimators (100-1000),
learning_rate (log, 0.01-0.2),
num_leaves (20-256),
max_depth (5-20),
min_child_samples (5-100),
subsample (0.6-1.0),
colsample_bytree (0.6-1.0), and
regularization, reg_alpha (log, 0.001-10.0),
reg_lambda (log, 0.001-10.0)

GPR search space
constant_value (log, 1e-10 to 1e5),
length_scale (log, 1e-10 to 1e5),
\nu ([0.5, 1.5, 2.5, 3.5, 4.5]),
noise_level (log, 1e-5 to 1e5),
regularization term \alpha (1e-10 to 1e5)

Light Gradient Boosting Machine



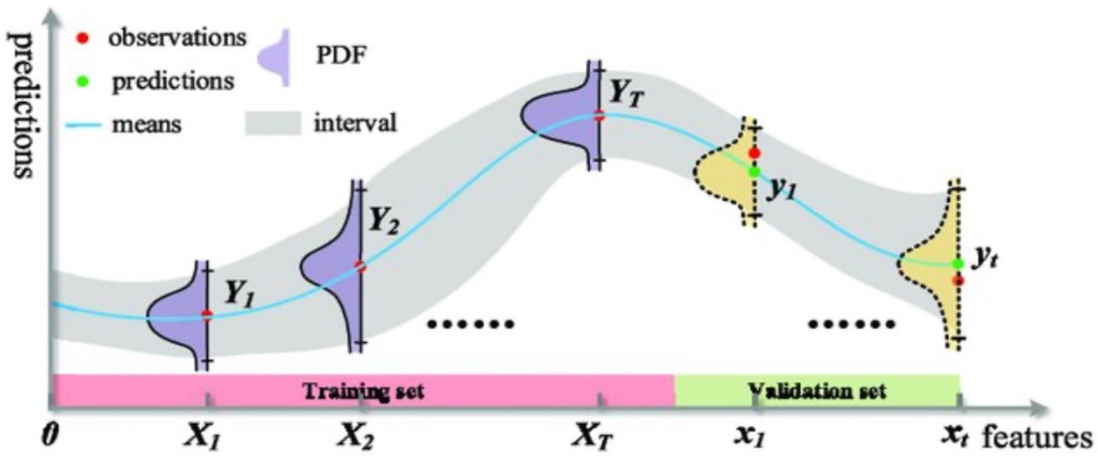
A high-performance gradient boosting model, builds decision trees. (a series of 'what if' questions to get to a prediction)

Known for its speed and high accuracy on tabular data.

Builds trees "leaf-wise" to capture complex patterns efficiently.

Well known for finding non-linear relationships in the features.

Gaussian Process Regression



R is determined by a smooth, general-purpose function (the Matern kernel), whose overall amplitude is scaled by a specific amount (the ConstantKernel), PLUS a small amount of random experimental noise(the WhiteKernel).

A series of continuous random variables subject to the Gaussian distribution constitute the Gaussian process.

In the case of discrete Gaussian process, deriving the Gaussian distribution parameters of unknown samples based on known sample information is Gaussian process regression (GPR).

Doesn't just predict a value; it provides a "confidence interval" (or uncertainty) for each prediction.

K-Fold Cross-Validation

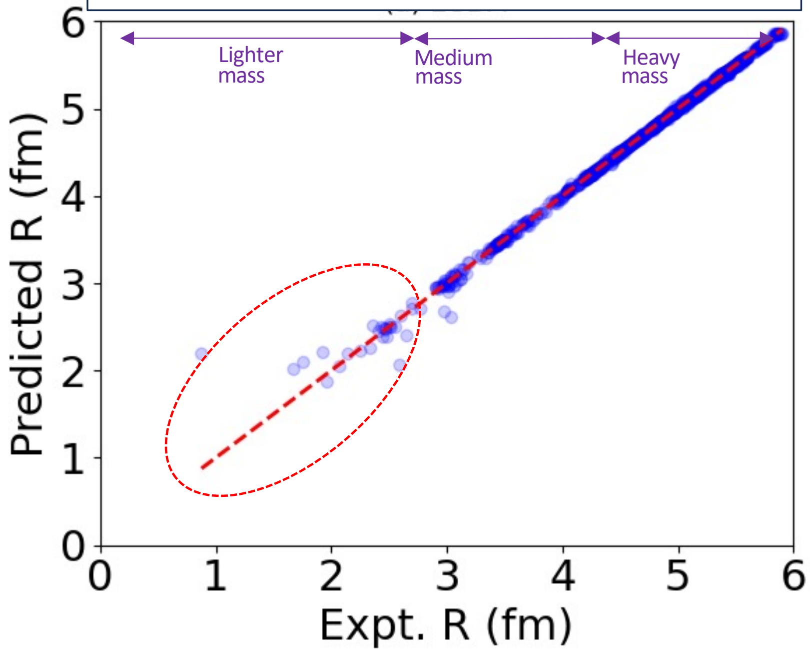
Cross-validation is a model assessment technique used to evaluate a machine learning algorithm's performance in making predictions on new datasets that it has not been trained on.

This is done by partitioning the known dataset, using a subset to train the algorithm and the remaining data for testing.

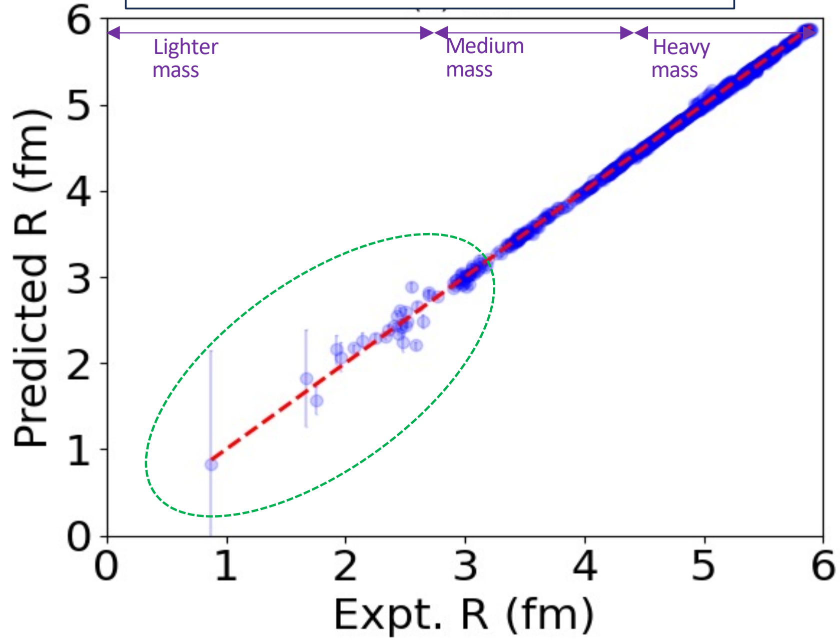
Numerical Regression Results

First global explanation of nuclear charge-radii including lighter mass nuclei

LGBM, Light Gradient Boosting Machine



GPR, Gaussian Process Regression



GPR is better for the lighter mass nuclei and even for the proton

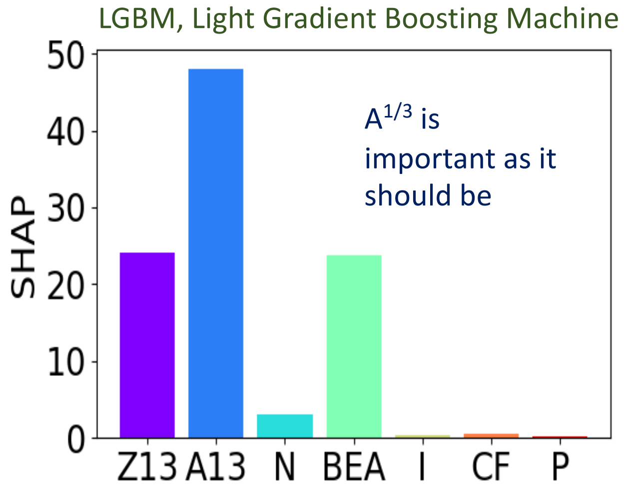
956 Input
3557 Output

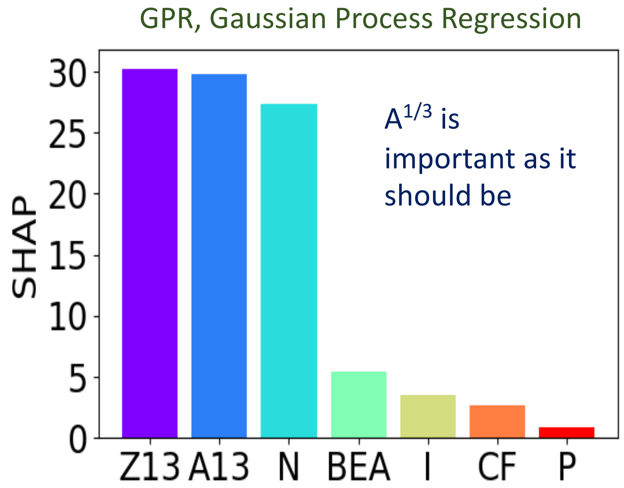
Extrapolations

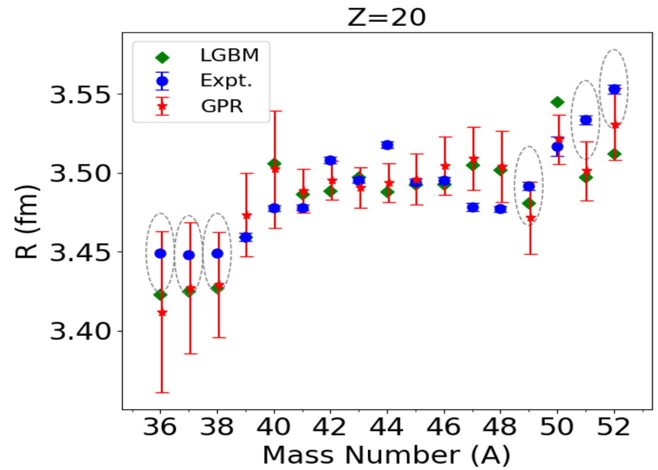
Important for Future Experiments

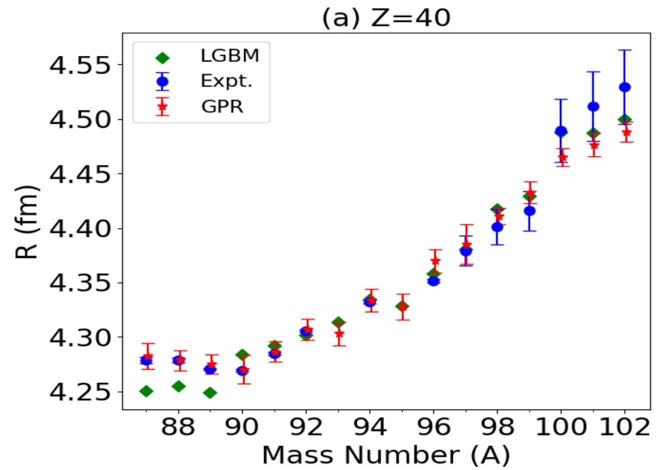
B. Maheshwari, P. Van Isacker, Submitted to Phys. Lett. B

Feature Importance Analysis









Ellipses present the data which was not a part of input (extrapolations)

GPR works better due to the uncertainities offered

Symbolic Regression

Starts by creating thousands of random, simple formulas
Tests them, 'kills' the bad ones, and 'breeds' the good ones together —swapping, adding, and mutating parts—over many generations

A "white-box" model that finds the underlying mathematical equation from data

Uses an evolutionary algorithm to build and test thousands of simple formulas

The final result is a simple, interpretable formula that is both accurate and easy to understand, (e.g., $R \sim A^{1/3} + \ldots$)

Discovering Physics Expressions

Predictions + Extrapolations

Symbolic Regression

Simplified Expression

$$R_{LGBM} = a_{1l}A^{1/3} + a_{2l}BE - a_{3l}N + c_l$$

a_{1l} =0.540, a_{2l} =1.655, a_{3l} = 0.013, c_l=1.265

$$R_{GPR} = a_{1g}A^{1/3} + a_{2g}BE + a_{3g}Z^{1/3} + c_g$$

a_{1g} =0.386, a_{2g} =0.186, a_{3g} = 0.536, c_g=0.617

B. Maheshwari, P. Van Isacker, Submitted to Phys. Lett. B

Traditionally known radius dependence on mass number is successfully reverse engineered

One can do such global analysis of other observables/properties

Symbolic vs. Experimental & Numerical Regression