Kevin Andrés Huillca Aparicio

Structural NanoFingerprint Generation

MASTER’S DEGREE FINAL PROJECT

Directed by Dr. Francesc Serratosa

Master’s degree in Computer Security Engineering and Artificial Intelligence

UNIVERSITAT ROVIRA I VIRGILI

Tarragona

2023

Abstract

The goal of this project is to focus on the development of a publicly accessible website application to aid in the generation of NanoFingerprints, which will later be used to train models for toxicity prediction of different nanocompounds. In this paper a description of the development of the web portal has been included as well as the different functionalities that a user can interact with. Additionally, the NanoFingerprint generated have been used to train models for toxicity prediction by means of Linear Regression. Models that have shown better performance have been added to the website as means of providing additional tools for aiding in the investigation of nanocompounds. This project has been developed under the SbD4Nano funded project together with Universitat Rovira I Virgili.

Resum

La finalitat d'aquest projecte és treballar en el desenvolupament d'un portal web de lliure accés al públic que permeti generar NanoFingerprints, que seran emprats per entrenar models per predir la toxicitat de diferents nanocomponents. En aquest document s'ha descrit el desenvolupament del portal web tenint en compte les funcionalitats disponibles amb les quals un usuari pot interactuar. A més, els NanoFingerprints generats s'han fet servir per entrenar models de predicció de toxicitats mitjançant Regressions Lineals. Els models que han mostrat millors resultats han estat afegit al portal web per tal de proveir d'eines addicional per ajudar amb la investigació de nanocomponents. Aquest projecte ha estat desenvolupat conjuntament amb la Universitat Rovira i Virgili sota el projecte finançat de SbD4Nano.

Resumen

La finalidad de este proyecto ha sido el desarrollo de un portal web de libre acceso al público que permitiese generar NanoFingerprints, que luego se utilizan para entrenar modelos de predicción de toxicidad de diferentes nanocomponentes. En este documento se ha descrito el desarrollo de la página web, teniendo en cuenta las funcionalidades disponibles que son accesibles a nivel usuario. Además, los NanoFingerprints generados se han utilizado para entrenar dichos modelos de predicción mediante el uso de Regresiones Lineales. Los modelos que han dado mejores resultados han sido añadidos en la página web para proveer de herramientas adicionales para ayudar en la investigación de nanocomponentes. Este Proyecto ha sido desarrollado juntamente con la Universitat Rovira I Virgili bajo el proyecto financiado de SbD4Nano.

Acknowledgements

I Would like to thank Dr. Francesc Serratosa for his support when working through the different milestones of this project. Additionally, I am grateful to have had the opportunity to take part in a European project such as Sb4DNano. Finally, I would like to thank my family and friends for their patience and understanding at all times.

Table of Contents

Table of Figures

1. Introduction

The following project has been developed under the SbD4Nano[1] funded project together with Universitat Rovira I Virgili. The presented work is divided in 2 main sections, as part of the study presented in [1], which aims to provide alternative toxicity prediction models by the analysis of nanocompounds as attributed graphs. These sections are:

- Development of a public website application
- Presentation of alternative toxicity prediction models using NanoFingerprints

A representation of the work presented in this thesis is shown below:

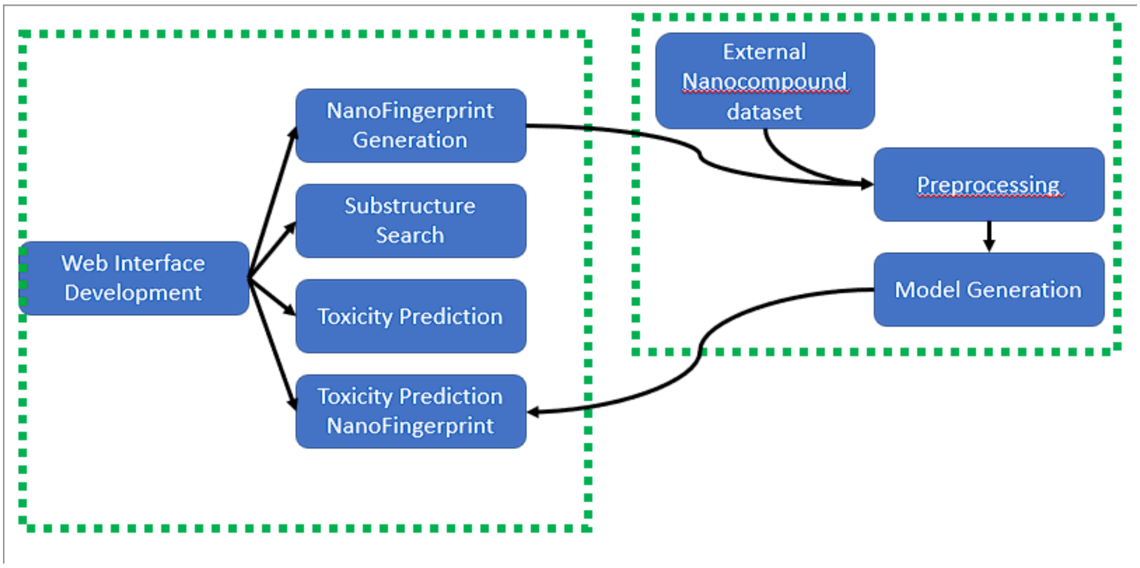


Figure 1. Project diagram

In terms of the web portal, the requirement is to develop the publicly accessible framework in charge of generating NanoFingerprints, ensuring all sections present in the definition of the NanoFingerprint are included considering the hardware constraints linked to the server hosting the site. NanoFingerprint is defined as a vector generated as a result of the analysis of a nanocompound considering it a graph.

Second part of the project consists of using the generated NanoFingerprints to define new models for toxicity prediction. Website will host known models and will include at the same time the proposed ones using NanoFingerprints, offering an alternative to the currently present in the literature.

[1] https://www.sbd4nano.eu/project-overview

1.1. Structure of thesis

This document is structured as follows: Section 2 corresponds to the presentation of the algorithms that are going to be analysed and compared to when proposing the toxicity models including the generated NanoFingerprints. The definition of the NanoFingerprint is also included, as well as an explanation of the different elements that form the array. A mention on the models that are going to be part of the case study are also described. Section 3 focuses on the presentation of the web portal as well as the different elements that form it, both at a backend and a frontend level. Additionally, a description of the datasets obtained joining both samples included in the literature as well as the NanoFingerprints generated is included. Furthermore, a definition of the metrics used to validate the models are shared. In Section 4, the preprocessing steps as well as the results associated with the models generated are included. Finally, Section 5 a discussion on the conclusions extracted from the project as well as possible future work development is presented.

2. Related work

2.1. Background

Multiple research involving the investigation of nanoparticles effect on living organisms has been presented, to have a better understanding not only of the implications they may have in nature but also considering the increase in use at both a commercial and industrial level [2]. As part of this project, 3 different approaches to analyse toxicity levels of a given nanocompound are explored and analysed.

The approach presented in [3] provides multiple MLR analyses in order to predict the impact TiO_2 and ZnO have on cells when inducing the release of lactate dehydrogenase (LDH) which is directly related to damage on tissue membrane cells when in direct contact. Three different MLR models are described, included below:

$$LDH(\mathrm{TiO2} + \mathrm{ZnO}) = 0.66 + 0.005 * X0 - 5 * X2 + 0.003 * X4 \quad (1)$$

$$LDH(\mathrm{TiO2}) = 0.599 + 0.004 * X0 + 0.003 * X4 \quad (2)$$

$$LDH(\mathrm{ZnO}) = 1.041 + 0.001 * X1 + 0.001 * X2 + 0.001 * X4 \quad (3)$$

X0 refers to the size in nanometres of the nanocompound, X1 is listed as the size of the same nanocompound in water, X2 corresponds to the size of the nanocompound in Phosphate buffered saline (PBS), and X4 is defined as the concentration of the sample, measured in mg/l.

The toxicity analysis presented in [4] is based on multiple nanocompounds, in this case other qualities of the samples are considered. No prediction of LDH is considered, instead it is stated that the if the output of the MLR is >0.5, the nanocomponent can be considered as TOXIC, otherwise it will be considered NOT TOXIC. Below is included the Linear regression model, which was included in NanoTox[2] nanotoxicology pipeline:

$$Toxicity = 29.964 - Size * 4.39 + HydroSize * 1.564 - SurfaceCharge * 1.7914 - SurfaceArea * 4.754 - Ec * 12.662 + Time * 1.0105 + Dose * 6.259 + Eneg * 7.983 - Noxygen * 10.557 \quad (4)$$

In both [5] [6], study is focused on the toxicity level for metal-oxide samples and their impact on human cells. Models proposed differ in both as a different enthalpy is used, below additional details regarding both models:

$$Puzyn - \log\left(\frac{1}{EC50}\right) = 2.59 + -50 * EnthalpyGaseousCation \quad (5)$$

$$Gajewicz - \log\left(\frac{1}{EC50}\right) = 2.47 + 0.24 * EnthalpyMetalOxide + 0.39 * X^c \quad (6)$$

X^c is defined as the Mulliken's electronegativity (in eV). All these parameters are included in the datasets extracted from the included articles.

2.2. Nanofingerprint structure

Presented in [1], a NanoFingerprint is a vector representation generated as a result of the analysis of a nanocompound. This nanocomponent can be understood as an attributed graph, by considering atoms as nodes and chemical bonds as edges between nodes. Given the information that can be attached to these nodes as well as the mathematical characteristics linked to graphs, a substructure graph is defined only using the shell of the nanoparticles, which is known to hold most of the information linked to the toxicity of the compound [7].

The NanoFingerprint array can be divided in subsections, based on the relationship between nodes and the group of subgraphs that are generated. These subgraphs are defined as follows:

- O(x): number of oxygen atoms that have x bonds.
- M(x): number of metal atom that have x bonds.

[2] https://github.com/NanoTox

- O(x, y): subgraph with a central oxygen atom connected to x oxygen atoms and y metal atoms.
- M(x, y): subgraph with a central metal atom connected to x oxygen atoms and y metal atoms.
- O(x, y)-O(x', y'): subgraph composed of O(x, y) and O(x', y') with central oxygen atoms connected.
- M(x, y)-M(x', y'): subgraph composed of M(x, y) and M(x', y') with central metal atoms connected.
- O(x, y)-M(x', y'): subgraph composed of O(x, y) and M(x', y') with central oxygen atom connected to central metal atom.

Considering these definitions, subsections are presented below. Note that index of the different elements have been added as well for better understanding of the size of the NanoFingerprint:

Section 1	1. Shell thickness in Angstrom
	2. Maximum number of bond per atom: **MAX**
	3. Size in Angstrom
	4. Atomic number of the metal
	5. Number of oxygen atoms
	6. Number of metal atoms
Section 2	7.Number of O(1)
	…
	MAX+6. Number of O(**MAX**)
	MAX+7. Number of M(**1**)
	…
	2MAX+6. Number of M(**MAX**)

Figure 2. NanoFingerprint Sections 1 & 2

Section 3	2MAX+7. Number of O(**0,1**)
	...
	$(MAX+1)^2$+2MAX+6. Number of O(**MAX, MAX**)
	$(MAX+1)^2$+2MAX+7. Number of M(**0, 1**)
	...
	2$(MAX+1)^2$+2MAX+6. Number of M(**MAX, MAX**)
Section 4	2$(MAX+1)^2$+2MAX+7. Number of O(**0,1**)-O(**0,1**)
	...
	$(MAX+1)^4$+2$(MAX+1)^2$+2MAX+6. Number of O(**MAX,MAX**)-O(**MAX,MAX**)
	$(MAX+1)^4$+2$(MAX+1)^2$+2MAX+7. Number of M(**0,1**)-M(**0,1**)
	...
	2$(MAX+1)^4$+2$(MAX+1)^2$+2MAX+6. Number of M(**MAX,MAX**)-M(**MAX,MAX**)
	2$(MAX+1)^4$+2$(MAX+1)^2$+2MAX+7. Number of O(**0,1**)-M(**0,1**)
	...
	3$(MAX+1)^4$+2$(MAX+1)^2$+2MAX+6. Number of O(**MAX,MAX**)-M(**MAX,MAX**)

Figure 3. NanoFingerprint Sections 3 & 4

Considering the sections above, it can be stated that the maximum length of a NanoFingerprint is linked to MAX value, which is the maximum number of bonds to be analysed. As such, the length of the NanoFingerprint is computed considering the following equation:

$$NanoFingerprint\ length = 3 * (MAX + 1)^4 + 2 * (MAX + 1)^2 + 2 * MAX + 6 \quad (1)$$

MAX Value	NanoFingerprint Length
4	1939
5	3976
6	7319
7	12436
8	19869
9	30224
10	44191

Table 1. Impact on NanoFingerprint length given a MAX value

For A visual representation of a nanocompound has been added below to aid in the understanding of the different substructures that are analysed and extracted to form the NanoFingerprint:

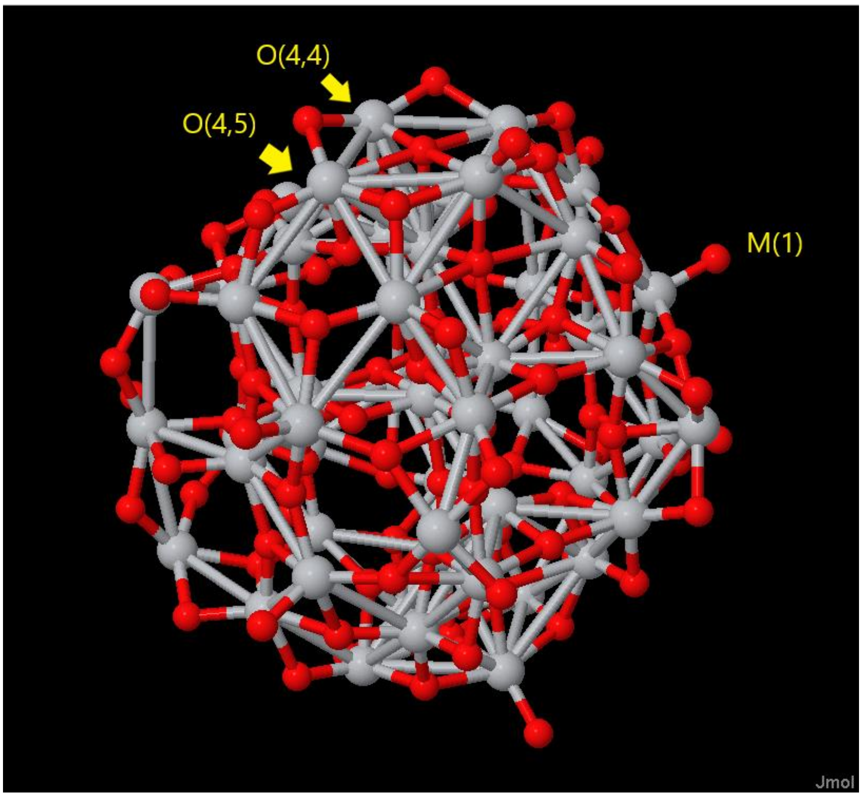


Figure 4. Nanocompound representation example

Multiple substructures have been highlighted using molviewer[3] in Matlab. It is seen two different configurations where an oxygen atom is a central node connecting with both metal atoms and other oxygen atoms. At the same time, we can observe multiple M (1), indicating a metal atom with a single bond.

3. Methodology

3.1. Atena Web Portal

A publicly accessible web portal was previously developed by a Software Engineer, allowing users to generate a shortened version of the NanoFingerprint, which contained only Section 1 and 2 (described above). Given the website was still in preliminary stages of development, a preliminary assessment was needed to understand the requirements and modifications to be included, more considering the feasibility of adding them given the constraints associated

[3] https://uk.mathworks.com/help/bioinfo/ref/molviewer.html

with the system hosting the website itself. Site is currently accessible through atena.urv.cat/model/.

Focus was put on developing the backend algorithm running the website. Since initial system was written in Go language, it was considered whether using a different programming language would have an impact on performance. As seen in [8] , Golang does make better use of CPU and memory available, which is linked to the fact of compiling it to a single binary file, compared to NodeJs. Additionally, including modifications on top of the initial system in Go reduced the time needed to include all requirements.

3.1.1. Characteristics

As stated above, web portal backend has been developed Golang App, together with Fiber framework[4]. Frontend consists of Bootstrap/JS as a Single Page Application. Compiled file for website is currently hosted in a Linux Virtual Machine in Atena host. Modifications applied to code have to be performed from a separate device labelled as Aura. This configuration was established to restrict outbound traffic from Atena, limiting only SSH as well as HTTPS inbound traffic. Guidelines on how to operate both Aura and Atena are included in Appendix A. Structure of the project goes as seen below:

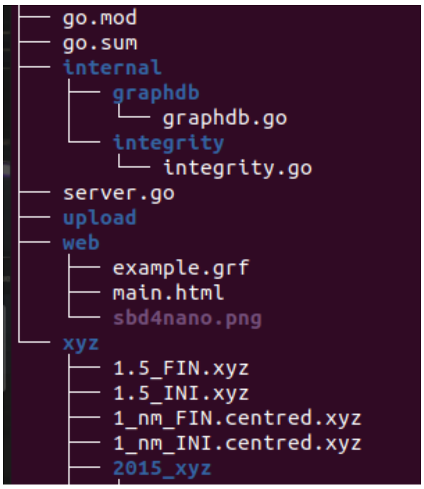


Figure 5. Project structure

Graphdb.go is in charge of performing all the required calculations to compute the toxicity level given a set of parameters, as well as generating the NanoFingerprint from a given nanocompound, using Neo4j[5] to perform a graph analysis of the nanocompound.

[4] https://gofiber.io/
[5] https://neo4j.com/

Server.go works as the main file of the application, in charge of starting the server and handling parameters submitted by user as well as providing the results through the web portal.

Main.html corresponds to the frontend part of the web portal. As stated earlier, it is an SPA. In terms of maintenance and update of the web portal, this is the preferable approach, more considering the different functionalities available.

3.1.2. Development environment

Given site was already accessible by external users, no modification was performed directly in the server hosting the site. Instead, a local testing environment was used to visualise the impact of the additional features as well as modifications performed.

A local Linux virtual machine was built from the scratch on a Windows 10 host using Oracle VM VirtualBox[6], replicating the structure already in place in production. Initially the idea was to include the hardware constraints present in production, however this led to higher processing time when handling some of the nanocompounds for further analysis. Below are hardware details from the host device in the local environment:

Hostname	DESKTOP-4K5CDJT
Processor	AMD Ryzen7-3800X 8-Core Processor 3.90 GHz
Graphic Card	AMD RX 6700 XT
RAM Available	24.0 GB

Out of the available resources, around 18GB RAM have been allocated for the Linux VM, as well as using five out of all the processors available. These values have been fined-tuned as new NanoFingerprints were generated, given some of these could not be completed due to not having enough memory available. Below a comparison between testing environment and production environment:

	PROD Environment	DEV Environment
# Processors	1	5
RAM	4 GB	18GB

[6] https://www.virtualbox.org/wiki/VirtualBox

3.2. Webpage sections

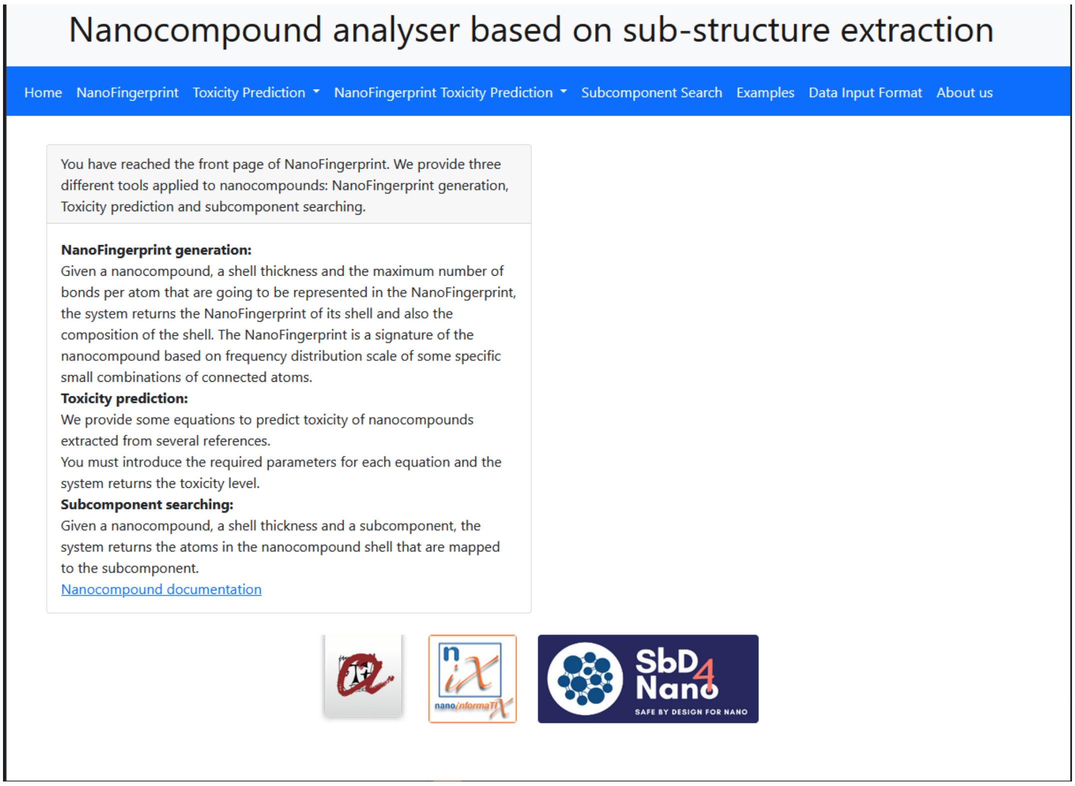


Figure 6. Atena web portal homepage

When accessing web portal, users are presented with a set of tabs that allows to gather different metrics as well as generation of the NanoFingerprint. Sections below include a description of the different tabs present as well as the functionalities included from the final user point of view.

3.2.1. NanoFingerprint

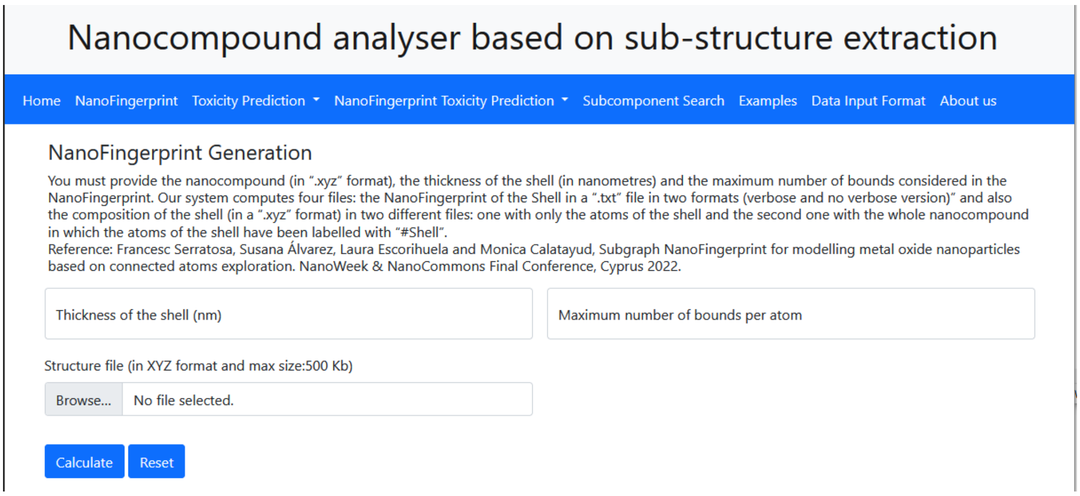


Figure 7. NanoFingerprint generation main page

This section allows the generation of the NanoFingerprint given a nanocompound. User is required to load a .xyz file and provide a thickness as well as a maximum number of bonds that are desired to be analysed. If user fails to input the data in the required format, portal will notify of the error and request the user to provide corrected information. There is a limitation in file size, directly tied to the capabilities of the system. These constraints are not present in the testing environment, as the available resources provided better manoeuvrability handling nanocompounds with higher file size. As such, production web portal is limited to a file size of 500Kb due to hardware constraints, whereas testing environment allows to increase this limit to 10Mb.

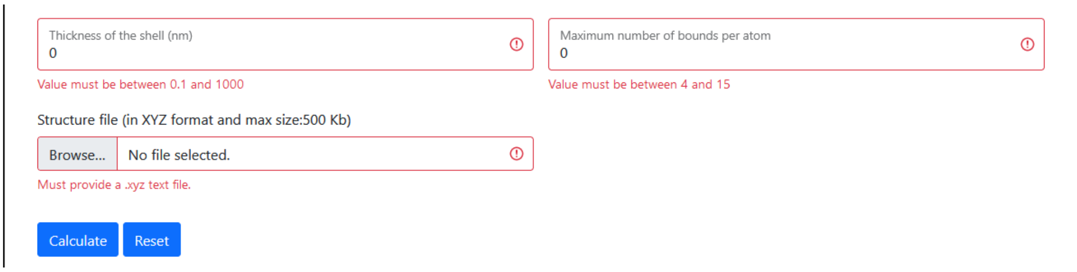


Figure 8. Error catch example

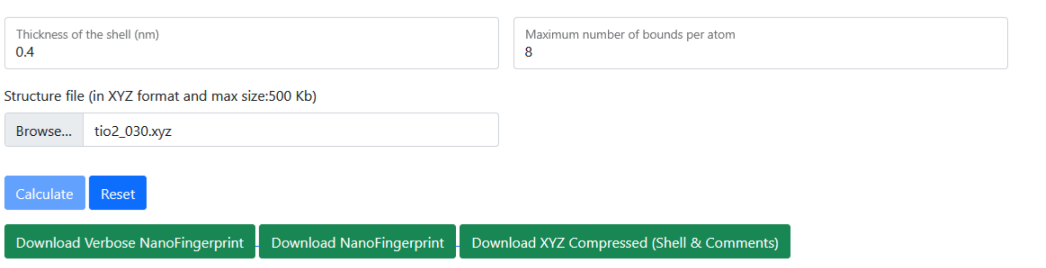
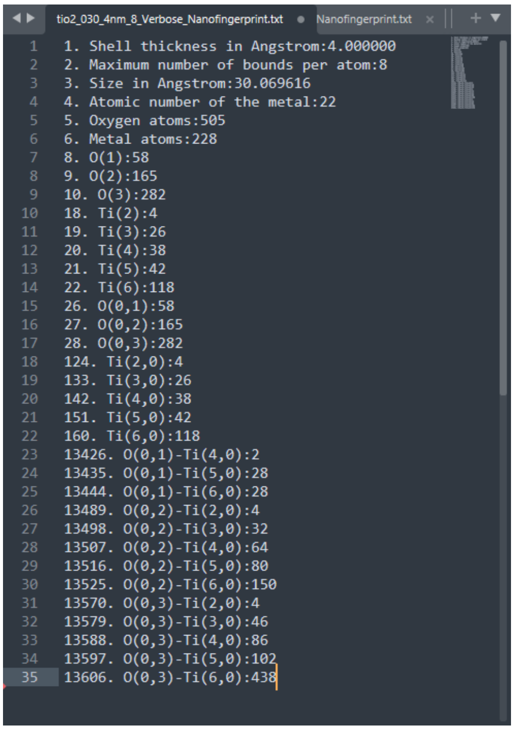


Figure 9. Results availability example

From a backend perspective, the file is locally stored in the server and multiple transactions are performed through Neo4j, not only by adding the atoms as nodes to the graph dB, but also when finding the relationship between nodes and finding the different subgraph based on a set of relationship between atoms, as seen in [1]. These are locally stored using UUID package, which avoids conflict when generating new files as uuid is used to name the directory where these files will be located.

Once all the required operations are performed, user is prompted with 3 downloadable files: Verbose NanoFingerprint, NanoFingerprint, and XYZ Compressed. First file includes a list of non-zero values associated with the NanoFingerprint, granting an overview of the subgraph relationship found given the input file. The second file provides the whole NanoFingerprint, including all zero values as well as the values listed in the Verbose file.

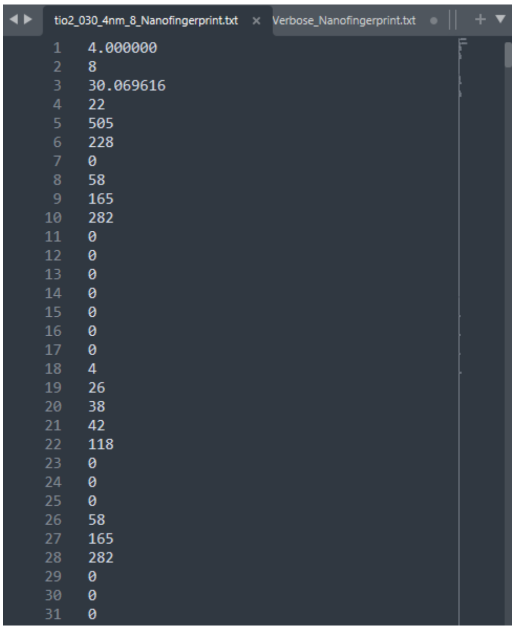
Both files are provided as txt files. Finally, the third available file is a .zip file containing both the input nanocompound .XYZ file (with only those atoms that are part of the shell), and the same input file labelling those atoms that were identified as part of the shell.



tio2_030_4nm_8_Verbose_Nanofingerprint.txt   Nanofingerprint.txt

```
1. Shell thickness in Angstrom:4.000000
2. Maximum number of bounds per atom:8
3. Size in Angstrom:30.069616
4. Atomic number of the metal:22
5. Oxygen atoms:505
6. Metal atoms:228
8. O(1):58
9. O(2):165
10. O(3):282
18. Ti(2):4
19. Ti(3):26
20. Ti(4):38
21. Ti(5):42
22. Ti(6):118
26. O(0,1):58
27. O(0,2):165
28. O(0,3):282
124. Ti(2,0):4
133. Ti(3,0):26
142. Ti(4,0):38
151. Ti(5,0):42
160. Ti(6,0):118
13426. O(0,1)-Ti(4,0):2
13435. O(0,1)-Ti(5,0):28
13444. O(0,1)-Ti(6,0):28
13489. O(0,2)-Ti(2,0):4
13498. O(0,2)-Ti(3,0):32
13507. O(0,2)-Ti(4,0):64
13516. O(0,2)-Ti(5,0):80
13525. O(0,2)-Ti(6,0):150
13570. O(0,3)-Ti(2,0):4
13579. O(0,3)-Ti(3,0):46
13588. O(0,3)-Ti(4,0):86
13597. O(0,3)-Ti(5,0):102
13606. O(0,3)-Ti(6,0):438
```

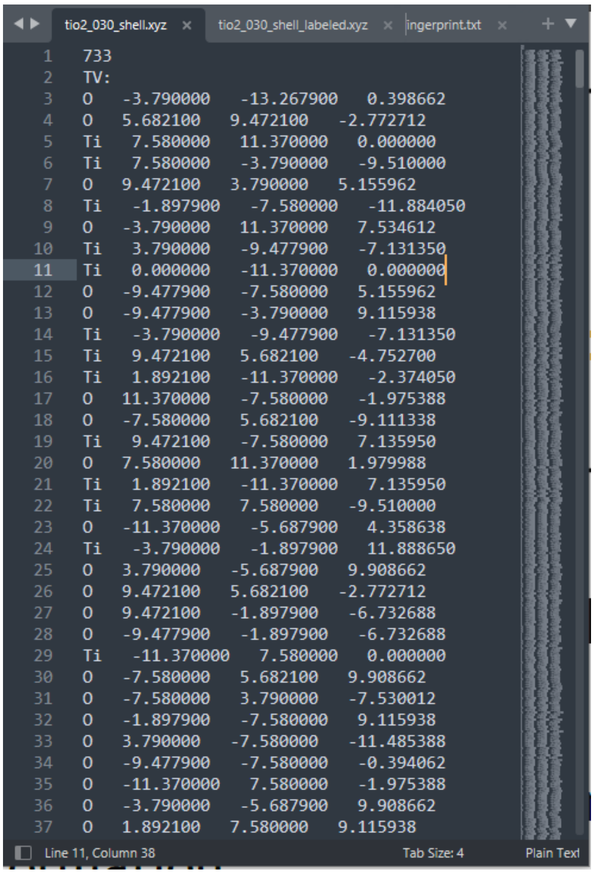
Figure 10. Verbose NanoFingerprint example



tio2_030_4nm_8_Nanofingerprint.txt   Verbose_Nanofingerprint.txt

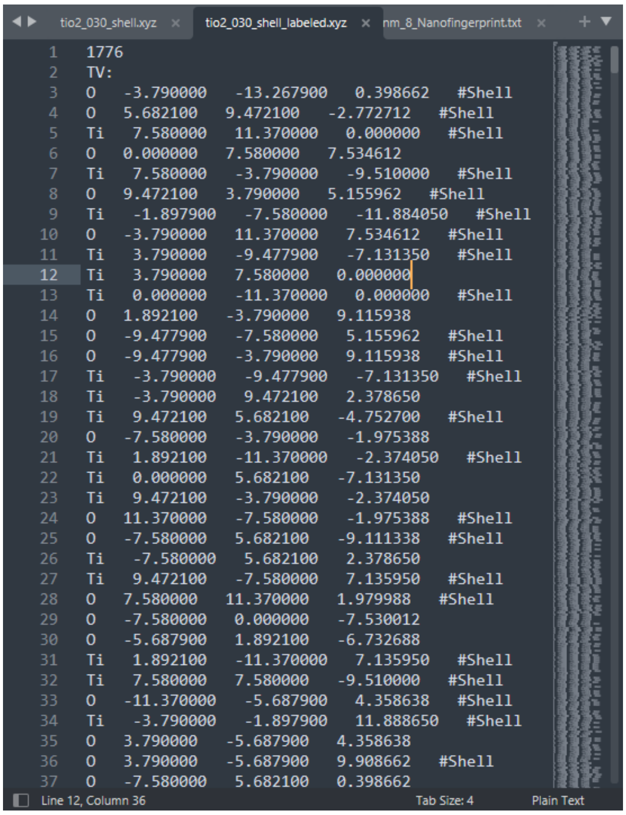
```
4.000000
8
30.069616
22
505
228
0
58
165
282
0
0
0
0
0
0
0
4
26
38
42
118
0
0
0
58
165
282
0
0
0
```

Figure 11. Nanofingerprint example




```
733
TV:
O   -3.790000   -13.267900   0.398662
O   5.682100   9.472100   -2.772712
Ti   7.580000   11.370000   0.000000
Ti   7.580000   -3.790000   -9.510000
O   9.472100   3.790000   5.155962
Ti   -1.897900   -7.580000   -11.884050
O   -3.790000   11.370000   7.534612
Ti   3.790000   -9.477900   -7.131350
Ti   0.000000   -11.370000   0.000000
O   -9.477900   -7.580000   5.155962
O   -9.477900   -3.790000   9.115938
Ti   -3.790000   -9.477900   -7.131350
Ti   9.472100   5.682100   -4.752700
Ti   1.892100   -11.370000   -2.374050
O   11.370000   -7.580000   -1.975388
O   -7.580000   5.682100   -9.111338
Ti   9.472100   -7.580000   7.135950
O   7.580000   11.370000   1.979988
Ti   1.892100   -11.370000   7.135950
Ti   7.580000   7.580000   -9.510000
O   -11.370000   -5.687900   4.358638
Ti   -3.790000   -1.897900   11.888650
O   3.790000   -5.687900   9.908662
O   9.472100   5.682100   -2.772712
O   9.472100   -1.897900   -6.732688
O   -9.477900   -1.897900   -6.732688
Ti   -11.370000   7.580000   0.000000
O   -7.580000   5.682100   9.908662
O   -7.580000   3.790000   -7.530012
O   -1.897900   -7.580000   9.115938
O   3.790000   -7.580000   -11.485388
O   -9.477900   -7.580000   -0.394062
O   -11.370000   7.580000   -1.975388
O   -3.790000   -5.687900   9.908662
O   1.892100   7.580000   9.115938
```

Figure 12. XYZ file with shell-only atoms




```
1776
TV:
O   -3.790000   -13.267900   0.398662   #Shell
O   5.682100   9.472100   -2.772712   #Shell
Ti   7.580000   11.370000   0.000000   #Shell
O   0.000000   7.580000   7.534612
Ti   7.580000   -3.790000   -9.510000   #Shell
O   9.472100   3.790000   5.155962   #Shell
Ti   -1.897900   -7.580000   -11.884050   #Shell
O   -3.790000   11.370000   7.534612   #Shell
Ti   3.790000   -9.477900   -7.131350   #Shell
Ti   3.790000   7.580000   0.000000
Ti   0.000000   -11.370000   0.000000   #Shell
O   1.892100   -3.790000   9.115938
O   -9.477900   -7.580000   5.155962   #Shell
O   -9.477900   -3.790000   9.115938   #Shell
Ti   -3.790000   -9.477900   -7.131350   #Shell
Ti   -3.790000   9.472100   2.378650
Ti   9.472100   5.682100   -4.752700   #Shell
O   -7.580000   -3.790000   -1.975388
Ti   1.892100   -11.370000   -2.374050   #Shell
Ti   0.000000   5.682100   -7.131350
Ti   9.472100   -3.790000   -2.374050
O   11.370000   -7.580000   -1.975388   #Shell
O   -7.580000   5.682100   -9.111338   #Shell
Ti   -7.580000   5.682100   2.378650
Ti   9.472100   -7.580000   7.135950   #Shell
O   7.580000   11.370000   1.979988   #Shell
O   -7.580000   0.000000   -7.530012
O   -5.687900   1.892100   -6.732688
Ti   1.892100   -11.370000   7.135950   #Shell
Ti   7.580000   7.580000   -9.510000   #Shell
O   -11.370000   -5.687900   4.358638   #Shell
Ti   -3.790000   -1.897900   11.888650   #Shell
O   3.790000   -5.687900   4.358638
O   3.790000   -5.687900   9.908662   #Shell
O   -7.580000   5.682100   0.398662
```

Figure 13. XYZ file with labels indicating shell atoms.

3.2.2. Toxicity Prediction

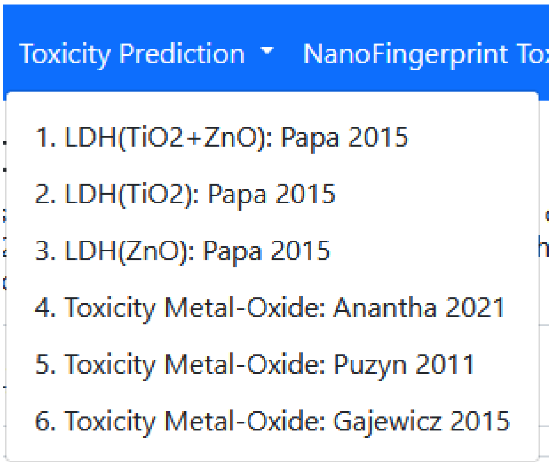


Figure 14. Toxicity Prediction

Within the tab there is a total of six separate options to be considered. The parameters required change depending on the model evaluated. As seen in [3], three different equations are presented, which have been added as individual fields since the input parameters differ between them.

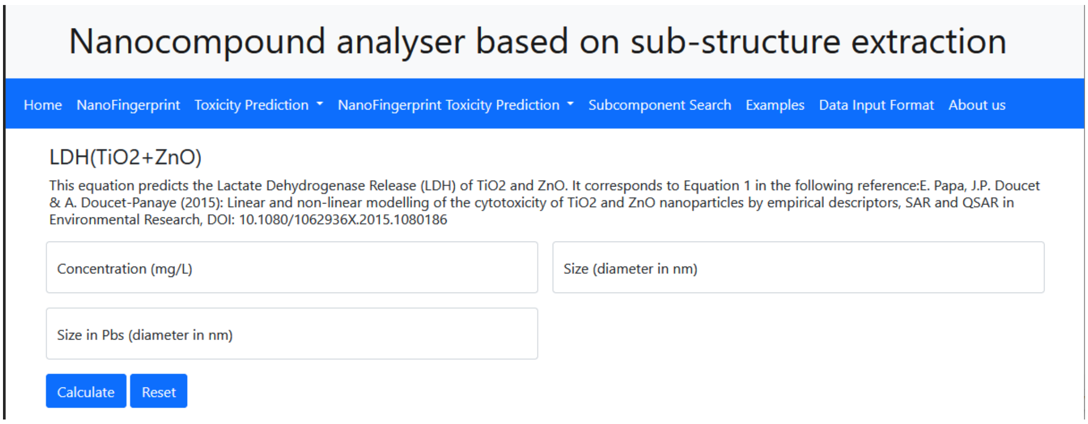


Figure 15. LDH (TiO2 + ZnO) toxicity calculation

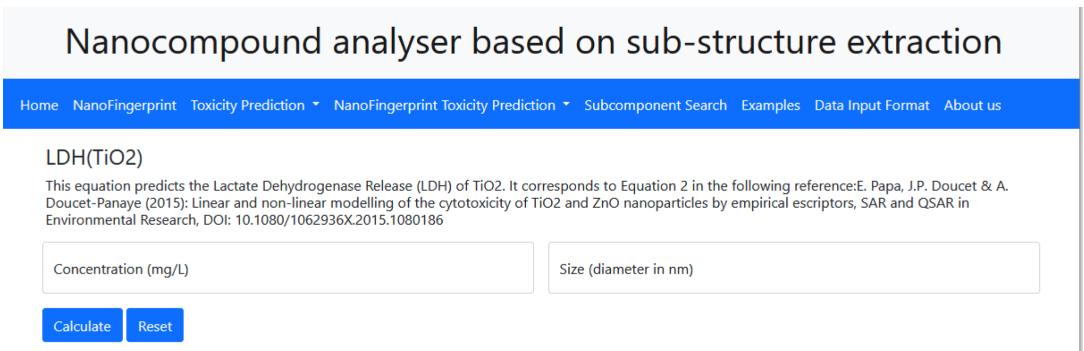


Figure 16. LDH (TiO2) toxicity calculation

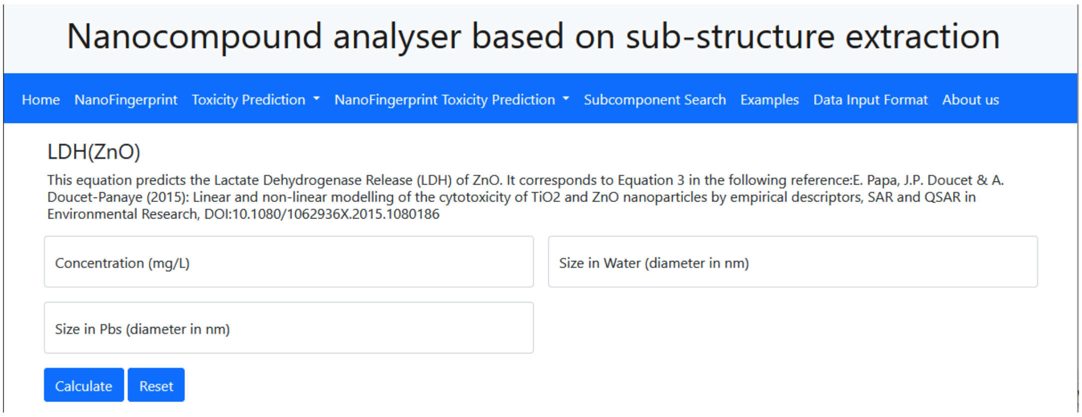


Figure 17. LDH (ZnO) toxicity calculation

These values are limited to fall within a given interval, if there is a failure to input the correct value, no toxicity will be calculated, and user will be prompted to correct the input values, as seen in the example below.

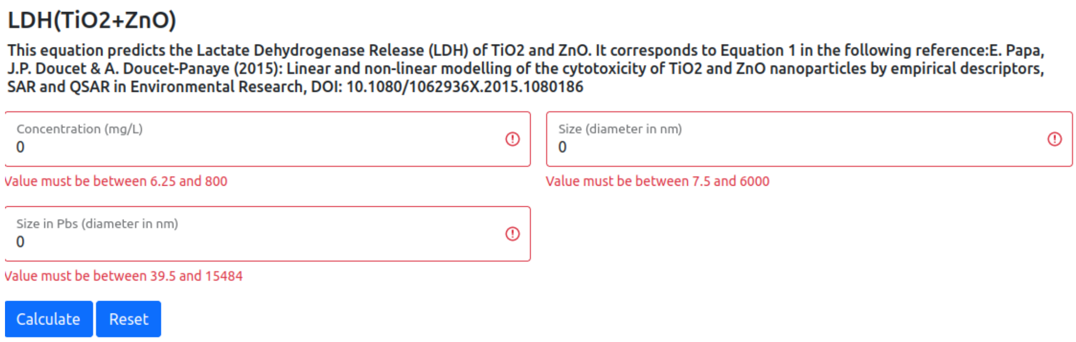


Figure 18. Example of incorrect input parameters submission

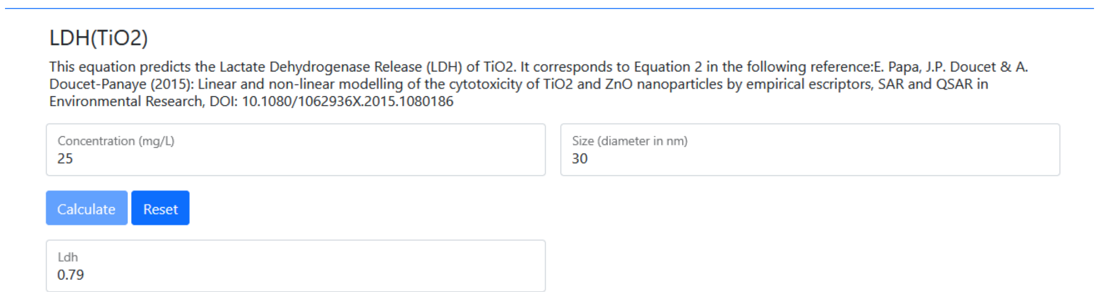


Figure 19. Example of output for LDH (TiO2)

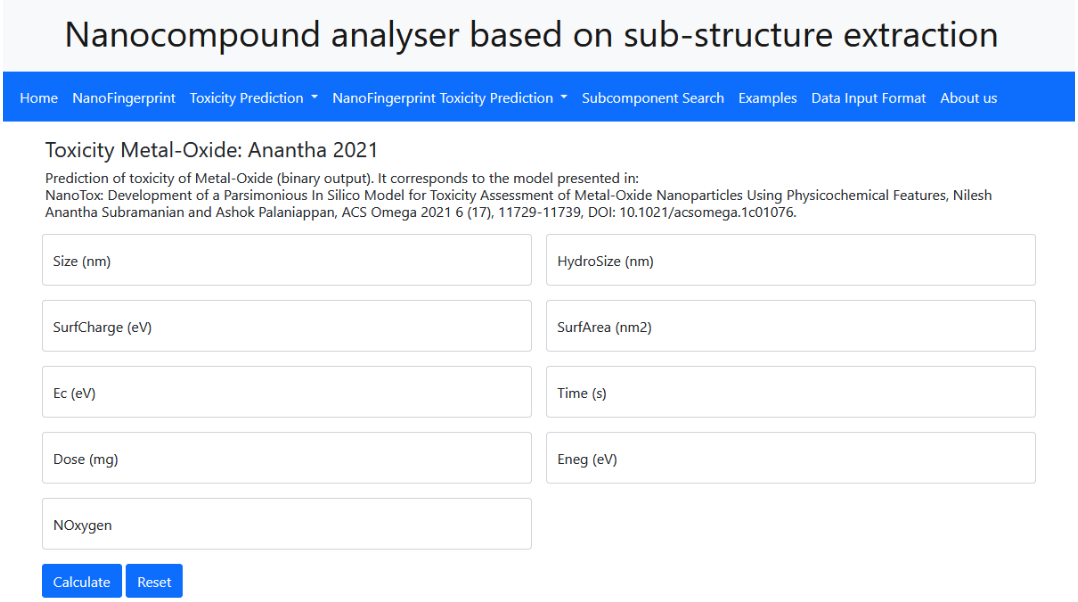


Figure 20. Anantha toxicity prediction

Anantha section does not provide a numeric value as it would occur with the other models, but rather a statement indicating whether the analysed nanocompound is either toxic or not, as seen in the example below:

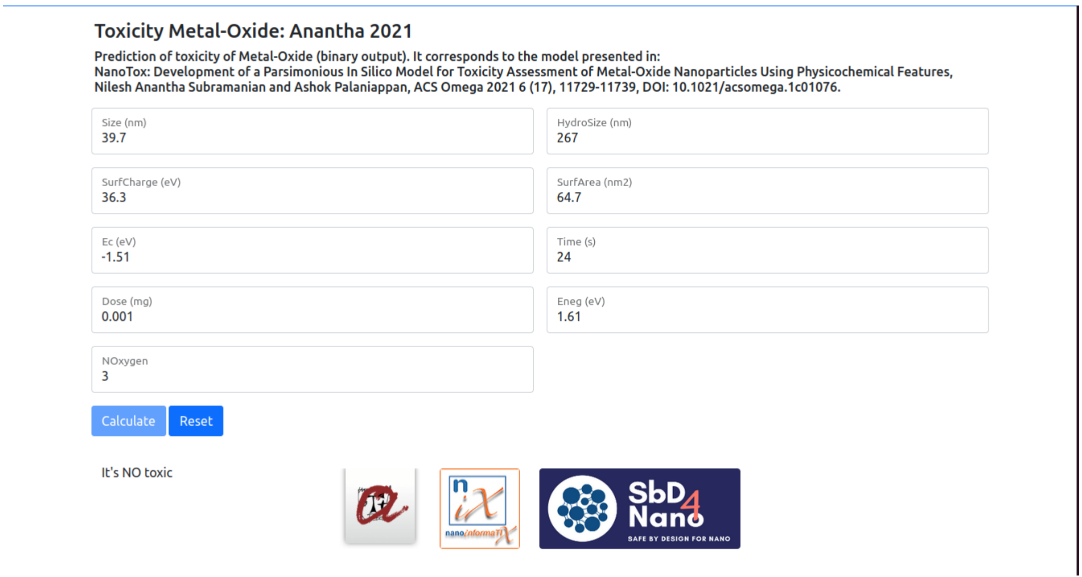


Figure 21. Anantha prediction example

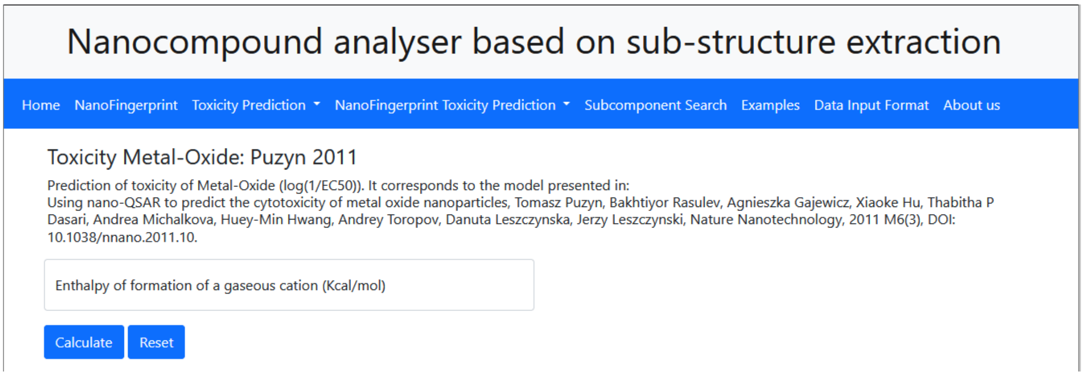


Figure 22. Toxicity prediction using equation from Puzyn

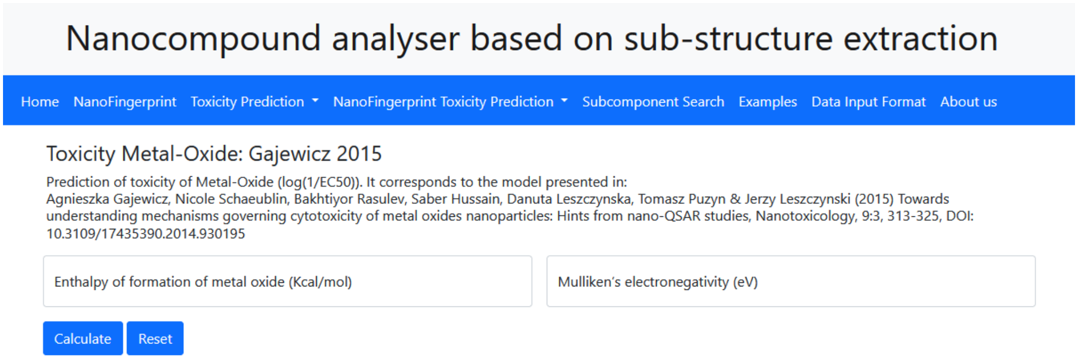


Figure 23. Toxicity prediction using equation from Gajewicz

In the backend, different REST API calls are performed depending on the selected options. Each call will perform a different operation considering input values. These operations are a code representation of the equations presented in [4] [3] [5] [6]. Example of the code used for computing the toxicity levels is appended below. Once computed these are returned to the webportal in JSON format, which then is:

```
func ldh4tz(c *fiber.Ctx) error {
    input := l4tz{}
    if e := c.BodyParser(&input); e != nil {
        return status400(c, e)
    }
    if e := integrity.CheckFields(input, "l4tz"); e != "" {
        return c.JSON(ldh{Ldh: 0, Error: e})
    }

    log.Println(c.Path(), input.Concentration, input.Size, input.SizeInPbs)

    return c.JSON(ldh{Ldh: roundFloat(0.66+0.003*input.Concentration+0.005*input.Size-5*input.SizeInPbs, 2), Error: ""})
}
```

Figure 24. Toxicity prediction for (TiO2+ZnO) model

```go
func ldh4tio2(c *fiber.Ctx) error {
    input := l4t{}
    if e := c.BodyParser(&input); e != nil {
        return status400(c, e)
    }
    if e := integrity.CheckFields(input, "l4t"); e != "" {
        return c.JSON(ldh{Ldh: 0, Error: e})
    }

    log.Println(c.Path(), input.Concentration, input.Size)

    return c.JSON(ldh{Ldh: roundFloat(0.599+0.003*input.Concentration+0.004*input.Size, 2), Error: ""})
}

func ldh4zno(c *fiber.Ctx) error {
    input := l4z{}
    if e := c.BodyParser(&input); e != nil {
        return status400(c, e)
    }
    if e := integrity.CheckFields(input, "l4z"); e != "" {
        return c.JSON(ldh{Ldh: 0, Error: e})
    }

    log.Println(c.Path(), input.Concentration, input.Size, input.Pbs)

    return c.JSON(ldh{Ldh: roundFloat(1.041+0.001*input.Size-0.001*input.Pbs+0.001*input.Concentration, 2), Error: ""})
}
```

Figure 25. Toxicity models for TiO2 and ZnO

3.2.3. NanoFingerprint Toxicity Prediction

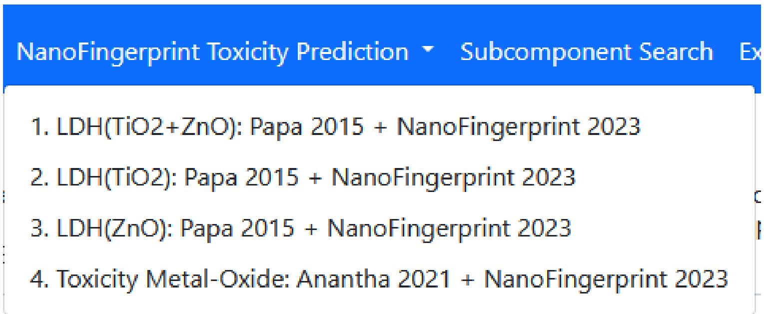


Figure 26. NanoFingerprint toxicity prediction options

In this scenario, toxicity levels are computed using not only the parameters previously presented but adding as well a previous computed NanoFingerprint. 4 different options are presented below, being 3 of these linked to [3] equations.

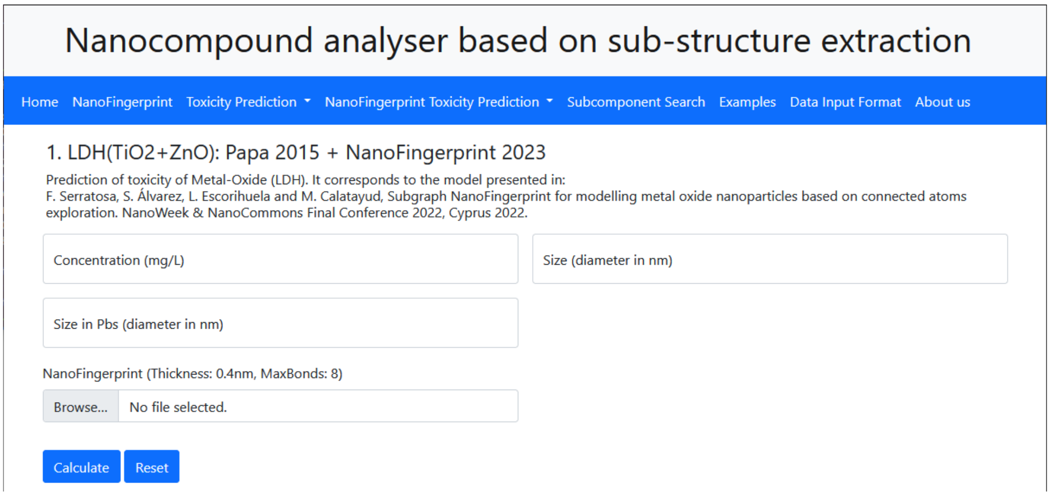


Figure 27. TiO2 + ZnO + NanoFingerprint

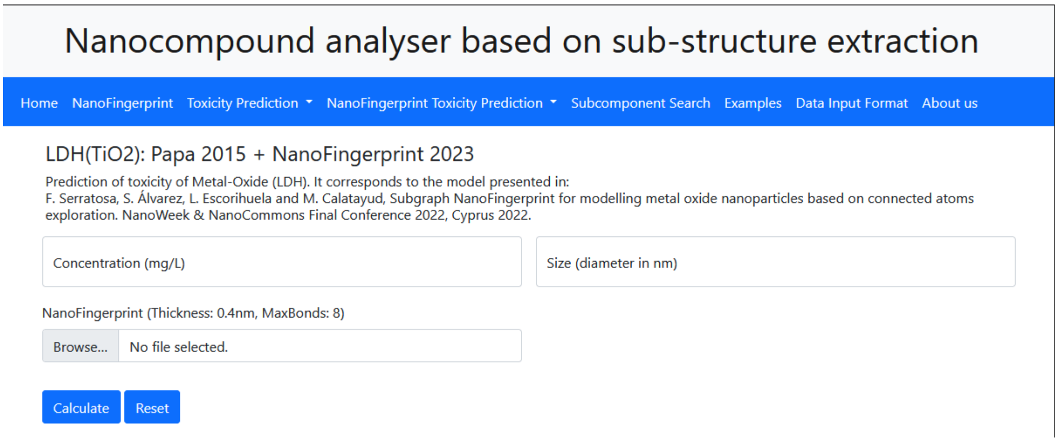


Figure 28. TiO2 + NanoFingerprint

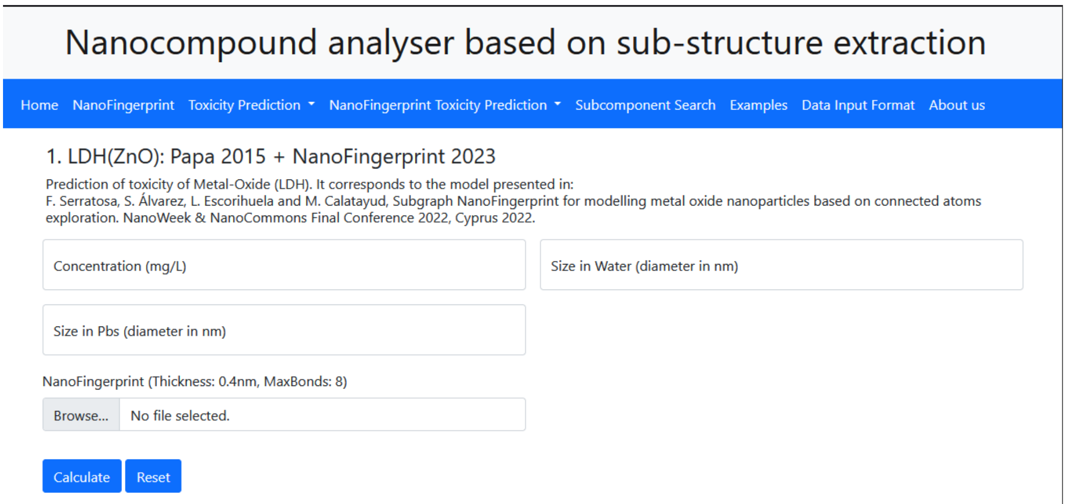


Figure 29. ZnO + NanoFingerprint

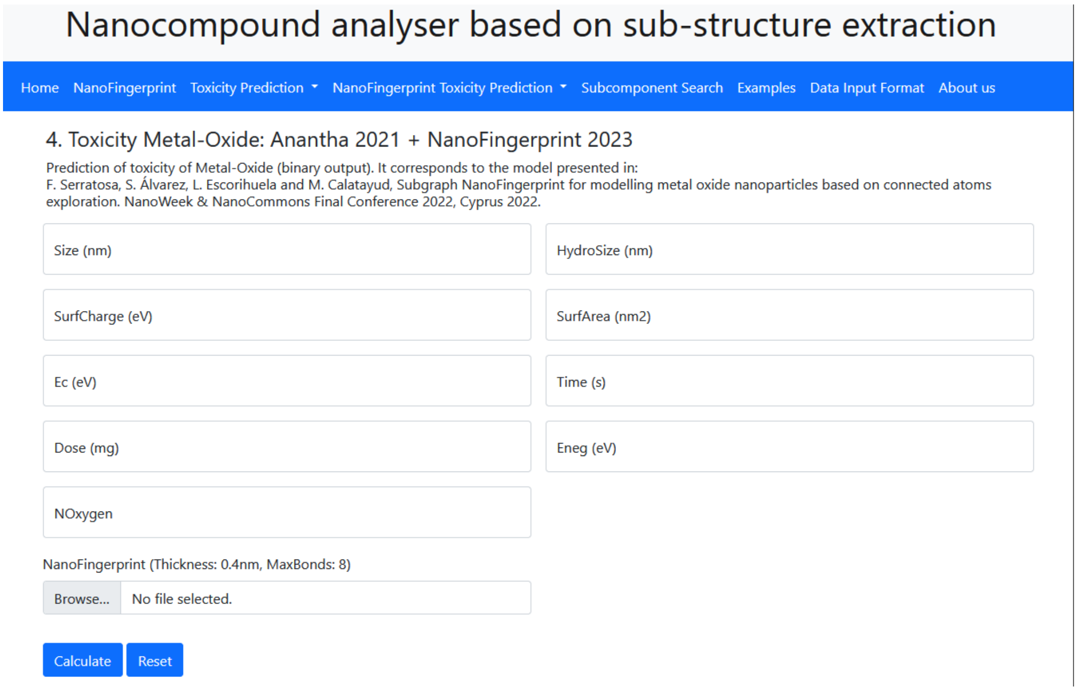


Figure 30. Anantha equation + NanoFingerprint

Pre-processing techniques are applied to the input NanoFingerprint before applying optimal models for toxicity prediction. Additional details regarding these models as well as the pre-processing techniques used can be seen in Results section. Like the previous tab description, all models except "Anantha" provide a toxicity estimation.

3.2.4. Subcomponent Search

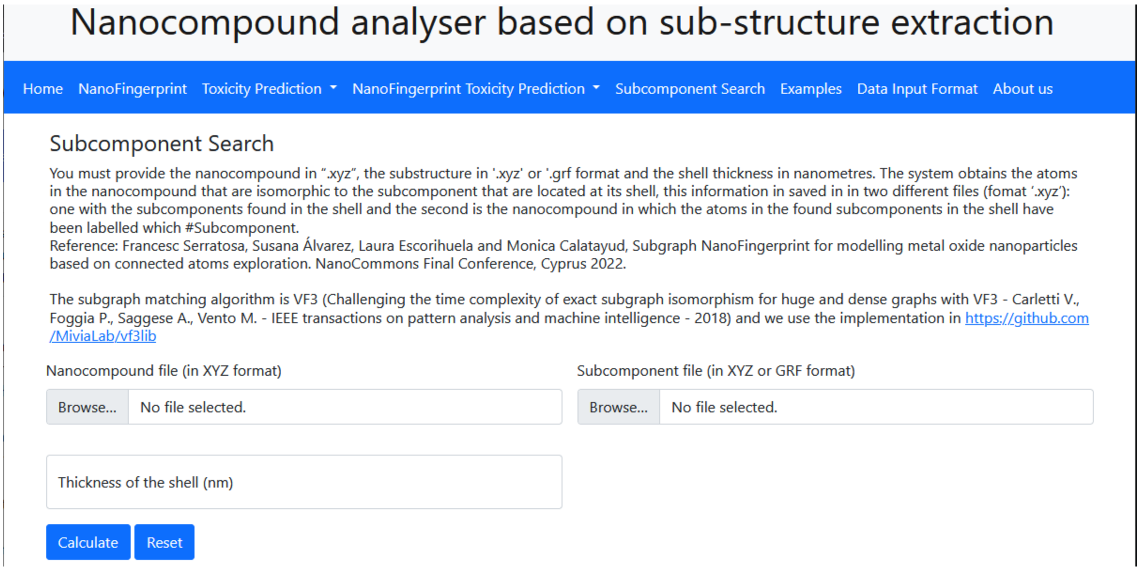


Figure 31. Subcomponent search section

Section allows the user to find patterns in the shell of the nanoncompound within a given nanocompound and a substructure. Function for this computation is not included in graphdb.go but rather in a set of multiple files written in C and based on VF3[7]. Server.go oversees the execution of the C code through the exec library.

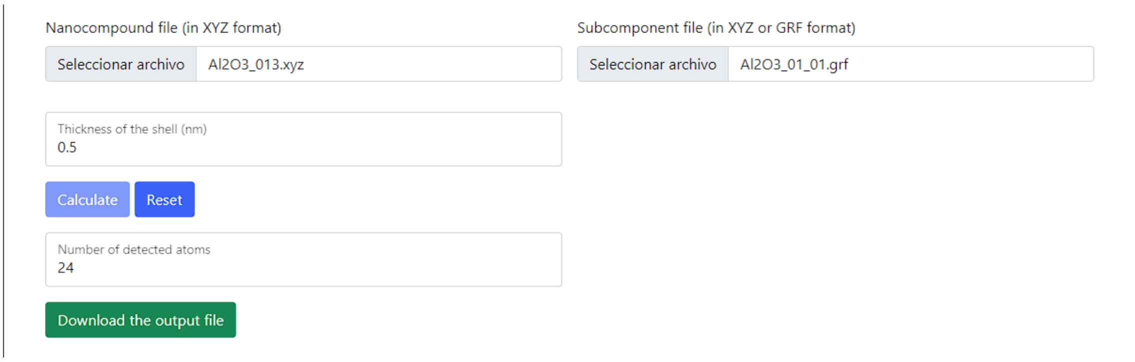
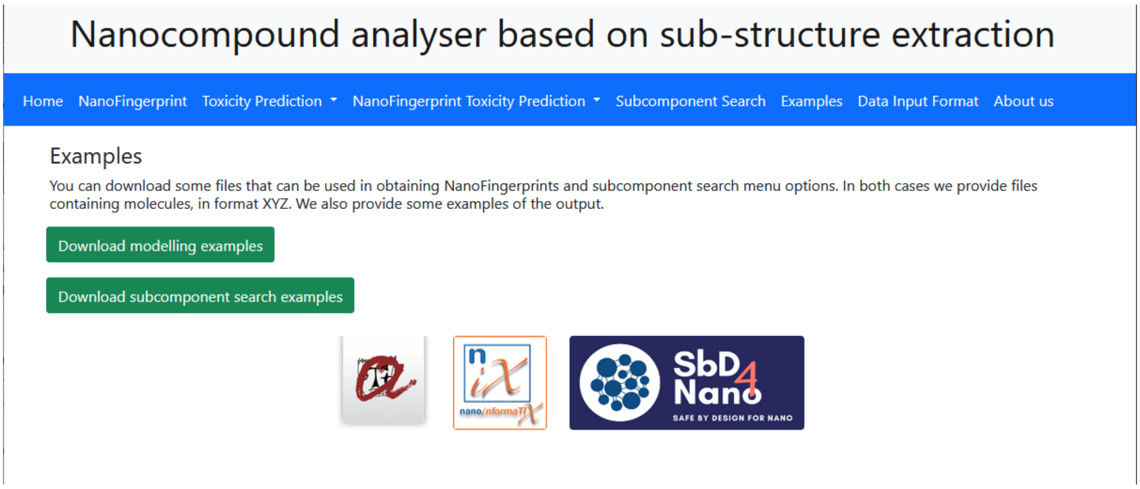


Figure 32. Substructure search result example

3.2.5. Examples



Users are provided with different samples to test both “NanoFingerprint” and “Subcomponent Search” tabs. Modelling examples files contain both images of some nanocompounds and XYZ files linked to the same nanocompounds to be used as input. At the same time, the corresponding NanoFingerprint has been added as part of the output examples. The second file contains both nanocompounds as well as substructures associated to these (in .GRF format). Some output examples are also added as a demonstration of the expected results.

[7] https://github.com/MiviaLab/vf3lib

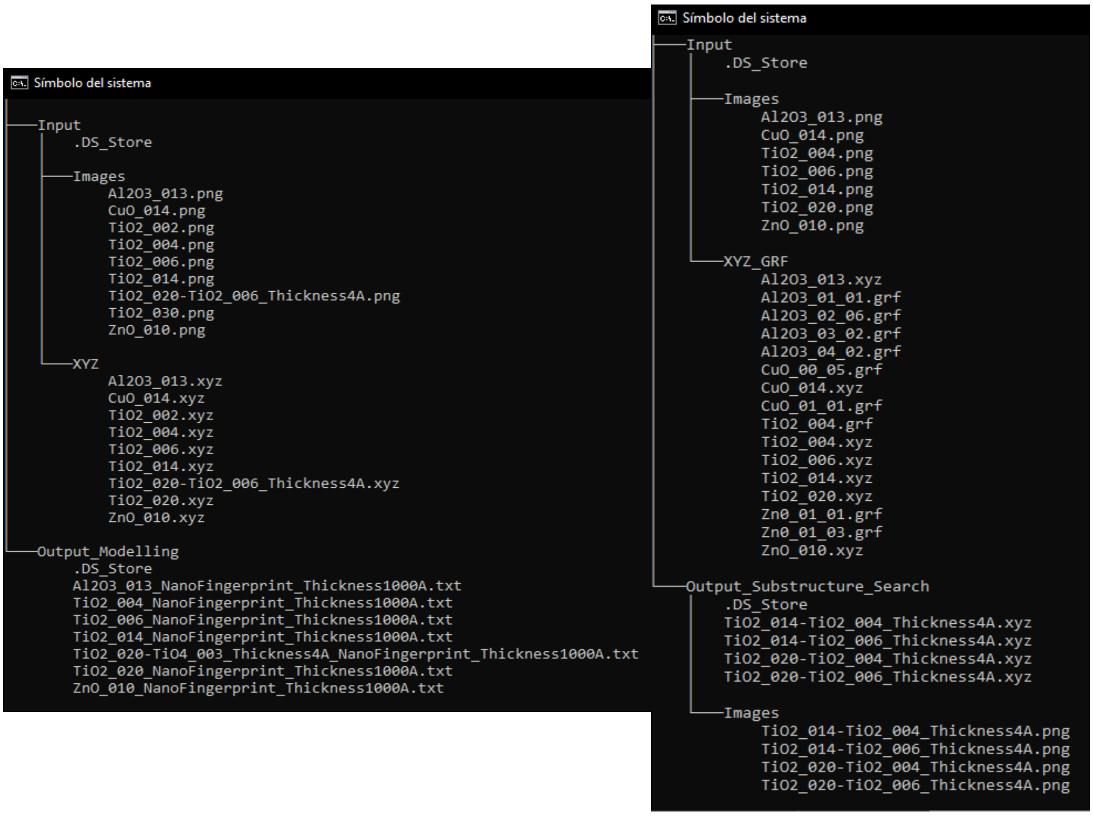


Figure 33. Directory tree for the Modelling and Subcomponent examples

3.2.6. Data Input Format

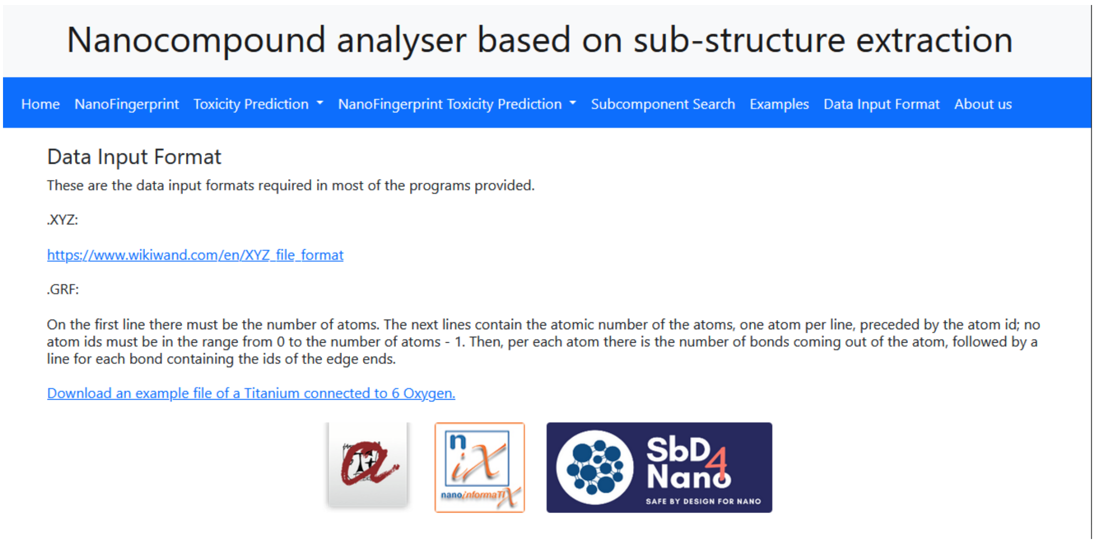


Figure 34. Expected input files details.

3.3. Nanofingeprint Generation

As stated earlier, graphdb.go file is in charge of performing the required operations to generate the NanoFingerprint given a set of input parameters. Once the request is sent through the webportal, server.go captures the values from the form and performs a go routine call to another function (coVectorCalc), which runs in parallel to main thread.

Within this function, a call to the main function in graphdb.go file, named GenVector, is performed.

```go
func vectorCalc(c *fiber.Ctx) error {

    thickness := c.FormValue("thickness")
    bounds := c.FormValue("bounds")
    file, _ := c.FormFile("file")
    strId := uuid.New().String()
    // fileName := fmt.Sprintf(uploadDir+"%s", strId+".xyz")
    filename_form := c.FormValue("filename")
    fileName := fmt.Sprintf(uploadDir+"%s", filename_form)
    c.SaveFile(file, fileName)
    go coVectorCalc(thickness, bounds, fileName, strId, filename_form)

    return c.JSON(modId{ID: strId, Error: ""})
}
```

Figure 35. Form input variables parsing

```go
func coVectorCalc(thickness string, bounds string, fileName string, strId string, base_name string) {
    var thick float64
    var bo int
    var err error
    if thick, err = strconv.ParseFloat(thickness, 64); err != nil {
        log.Panic(err)
    }
    if bo, err = strconv.Atoi(bounds); err != nil {
        log.Panic(err)
    }
    fileToDownload, fileToDownloadNoVer, err, fileShellZIP := graphdb.GenVector(fileName, thick, false, bo, downloadDir, strId, base_name)
    lock.Lock()
    defer lock.Unlock()
    if err == nil {
        fpReq[strId] = mod{File: fileToDownload, FileNoVer: fileToDownloadNoVer, Error: "", FileShell: fileShellZIP}
    } else {
        fpReq[strId] = mod{File: "", FileNoVer: "", Error: "Processing of the data failed"}
    }

}
```

Figure 36. Go routine call to CoVectorCalc

GenVector reads the input file submitted through the form and parses the contents of the file to compute the distance between atoms, as well as determining whether the atom corresponds to the shell part of the nanocompound or not. From this function, a call to computeShell is performed, which creates a Neo4j session. During the session, the atoms identified as shell are added to a graph database and queries are submitted to Neo4j to find those nodes that match the specifications listed in Sections 3 and 4 from [1]. Once the required operations are performed, Neo4j session is closed and database is deleted, to ensure no remaining data is left behind that could impact future NanoFingerprint generations. Results from Neo4j are written to 2 different files, one containing the whole NanoFingerprint array, and a second file specifying only those non-zero values found during the analysis. Furthermore, 2 more files are generated, which are stored as a ZIP file containing details regarding the shell itself, in .XYZ format. Details on GenVector function as well as secondary functions within graphdb.go are included in Appendix A.

3.3. Dataset

For this project, 2 different datasets have been used, one using the samples presented in [3] and adding the NanoFingerprints associated with the nanocompounds listed in the dataset, and a second dataset using the parameters included in [4], adding as well the NanoFingerprints for these nanocompounds. Regarding the first dataset, which will be labelled as "Papa_db"for simplicity when referring to it on this document, is formed by a total of forty-two samples. Out of these, nine different nanocompounds were identified.

Second dataset, which will be referred as "Anantha_db" going forward on this document, is formed by twenty-one different nanocompounds and a total of 483 samples. Both datasets have NanoFingerprints with thickness of 0.4nm and maximum number of bonds set to 8 as default values. Considering the bigger the value of MaxBonds the lower the computational efficiency, no information loss was observed when using a value of 8, as no substructure was found with more than 8 bonds. Most of the samples used for the toxicity prediction in both datasets showed a maximum of 6 bonds with other metal or oxygen atoms. Regarding thickness, it was set to 0.4nm as it was the most common value seen on the nanocompounds analysed. Thickness reference was initially extracted from the "Nanoparticle shell-depth calculator" portal[8]. Given these parameters, the NanoFingerprint array length was equivalent to 19867, as seen in [1]. This implies that the size of Papa_db is [42, 19874], whereas Anatha_db is [483, 19893]. These sizes correspond to the raw data without any preprocessing techniques applied. Since the number of samples available in Papa_db was low, cross-validation was performed to analyse the features as an attempt to avoid overfitting the model. Prior initial analysis consisted in dividing data in training, validation and test by assigning 70% of samples to training, and 15% for both validation and test and testing a set of models after performing a data-driven analysis. Cross-validation was also performed for Anantha_db. Additionally, recursive feature elimination (RFE) has been applied. As presented in [9], this technique allows to prune features considering the performance on the model and a ranking criterion. The iterative process of reducing features that do not contribute to a better model performance allows to find and rank those features that do so, especially for a limited number of samples. Results associated to this initial approach are included in section 4.

[8] https://nanogen.me/shell-depth

3.4. Metrics

3.3.1. CCC

Presented in [10], metric is presented as an alternative to validate QSAR models, which are described as classification algorithms used mainly in chemical engineering. Formula for the metric is presented below:

$$CCC = \frac{2 * \sum_{i=1}^{n}\left(y_i - \mu_y\right) * (y\_predicted_i - \mu_{y\_predicted})}{\sum_{i=1}^{n}\left(y_i - \mu_y\right)^2 + \sum_{i=1}^{n}(y_{predicted_i} - \mu_{y_{predicted}})^2 + n * (\mu_y - \mu_{y_{predicted}})^2}$$

In this case, CCC <= 0 would indicate model does not fit data properly and fails to predict values for unseen data. If CCC is closer to 1, it can be confirmed with high confidence that model is a good candidate. Information would need to be correlated with other model parameters.

3.3.2. RMSE

Allows to measure the accuracy of a model by comparing the predicted data obtained from a model when analysing unseen data with the ground truth associated with these testing samples. The benefit of using this metric is linked to the impact large errors has on the metric, as the distance between prediction and actual value is later squared.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_i - y_{predicted_i})^2}{n}}$$

3.3.3. MAE

Another metric that has been considered for model evaluation has been "Mean Absolute Error", which measures the average of the error in predicting the output, this gives a straightforward measure of the error size across the analysed samples. A larger MAE value indicates a poor performance of the model, since difference between prediction and expected value would be higher.

$$\frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_{predicted_i}\right|$$

3.3.4. R^2

The coefficient of determination, or r-squared, has been widely used in QSAR model analysis – it provides details regarding how well the model account for the variability of the expected values. r-squared of 0 would indicate model does not account for the samples analysed.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y\_predicted_i)^2}{\sum_{i=1}^{n}\left(y_i - \mu_y\right)^2}$$

4. Experimental Validation

4.1. Environment

In order to perform the analysis of the multiple datasets obtained, Python has been used. Multiple libraries were used to interact with the datasets, such as pandas and numpy. Models have been trained using scikit-learn library. Regarding the hardware features of the device used, the CPU was 8-core and a total of 24GB RAM available. In this case however, given the volume of data, no limitation was seen as per the resources used. Code used for preprocessing and modelling are included in Appendix A.

VSCode was the software using for working in the modelling script as well as the files required for the web portal, as stated in a previous section, web portal functionality was tested in a local Linux VM.

4.2. Web Portal Development

As stated in the definition of the project, the web portal was able to perform limited operations when generating the NanoFingerprint. Both sections 3 and 4 are seen to be the most exhaustive in terms of number of operations required, and these were not initially added to the production environment as it would disable the access to the portal itself. As part of the development of the website, these were later included with no direct impact on the performance of the server and are available to be downloaded by the end user. Additionally, user was given the option to download a new .XYZ with atoms linked to the nanocompound shell, as well as the same .XYZ but labelling what atoms were part of the shell. Both files are zipped to be downloaded as a single file. In order to avoid any possible disability of the website due to high processing time, files loaded to portal have been limited to 500Kb when calculating the NanoFingerprint.

Some additional functionalities were added linked to parsing the input file, which was seen to fail in some instances when these did not include a comment in the second line. This was corrected to ensure all files could be properly parsed as well as handling the error without crashing the web portal.

A better file management was added – initially, uuid generated from Google UUID was used to label the files that would later be downloaded. However, this added the difficulty of identifying what nanocompound it was referring to as well as what parameters were used for both thickness and maximum bonds. The proposed method consists of using uuid as a directory name instead, keeping name convention used in input file as well as adding what parameters were included, as seen in the example below:

```
vboxuser@ubuntu-desktop:~/atenaserver/downloads/39ff738d-3501-41b0-b84e-3f479e3691a9$ tree
.
├── SiO2_015_4nm_8_Nanofingerprint.txt
├── SiO2_015_4nm_8_Verbose_Nanofingerprint.txt
├── SiO2_015_compressed.zip
├── SiO2_015_shell_labeled.xyz
└── SiO2_015_shell.xyz

0 directories, 5 files
```

Proposed Linear Regression models have been included in webportal as alternative toxicity prediction algorithms using the information provided by the NanoFingerprints. Website has been entirely rebuilt to allocate all changes presented above.

4.3. Data Preprocessing

Given the vast number of features available after appending the NanoFingerprints to the different databases, feature selection was required, specially to remove those features that had redundant zeros across all samples. Some nanocompounds were initially removed due to not having the associated XYZ file available. For Papa_db, this was linked to the compound file size, which was around GBs of data. For Anantha_db, some nanocompounds could not generate a NanoFingerprint due to memory constraints in both testing and production environments.

Three different subsets have been created for comparing results with the ones provided in [3], whereas Anantha_db did not require any subset. First subset consists only of nanocompounds labelled as "TiO2". Second subset is formed by "ZnO" components, and third one is composed by both nanocompound types. Data was initially split in training and test. When applying data-driven analysis to "TiO2" dataset, no impact in the final number of rows were seen, however size of available features did change in both "ZnO" and "TiO2+ZnO" datasets. In all instances, cross-validation together with recursive feature elimination has been performed to see as well what features would work best given the training data provided. Following sections provide an example of possible training partitions after only applying data-driven analysis and removing features that are known to impact the model negatively. These are part of a data-driven analysis and not definitive but provides a guidance on how preprocessing affect to the final remaining features. In the following section, when presenting the models obtained, a mention on the features used as well as the results obtained for the model are included.

4.3.1. Papa – TiO2 dataset

Below is an example of some rows under TiO2 subset:

1	Name	Eng.Size (X0)	Size in Water (X1)	Size in PBS (X2)	Concentration (X4)	Zeta Potential (X5)	Y - LDH	Thickness (nm)	MaxBounds	Size (Angstrom)	Atomic Number Metal	Number Oxigen Atoms	Number Metal Atoms
2	TiO2_30	30	125	1250	25	-10	0.9	4	8	30.069616	22	505	228
3	TiO2_30	30	102	987	25	-12	1	4	8	30.069616	22	505	228
4	TiO2_30	30	281	1543	50	-15	0.75	4	8	30.069616	22	505	228
5	TiO2_30	30	101	1045	50	-9	0.7	4	8	30.069616	22	505	228
6	TiO2_30	30	299	1754	100	-11	1.04	4	8	30.069616	22	505	228
7	TiO2_30	30	134	961	100	-11	1.09	4	8	30.069616	22	505	228
8	TiO2_30	30	600	1876	200	-12	1.15	4	8	30.069616	22	505	228
9	TiO2_30	30	298	1165	200	-12	1.2	4	8	30.069616	22	505	228
10	TiO2_45	45	129	2567	25	-9	0.9	4	8	45.902255	22	1320	674
11	TiO2_45	45	129	2309	25	-10	0.85	4	8	45.902255	22	1320	674
12	TiO2_45	45	201	2431	50	-9	0.75	4	8	45.902255	22	1320	674
13	TiO2_45	45	201	2987	50	-11	0.78	4	8	45.902255	22	1320	674
14	TiO2_45	45	451	2941	100	-11	1.4	4	8	45.902255	22	1320	674

Figure 37. TiO2 dataset example (Part 1)

1	Number O(1)	Number O(2)	Number O(3)	Number O(4)	Number O(5)	Number O(6)	Number O(7)	Number O(8)	Number M(1)	Number M(2)	Number M(3)	Number M(4)	Number M(5)	Number M(6)	Number M(7)	Number M(8)
2	58	165	282	0	0	0	0	0	0	4	26	38	42	118	0	0
3	58	165	282	0	0	0	0	0	0	4	26	38	42	118	0	0
4	58	165	282	0	0	0	0	0	0	4	26	38	42	118	0	0
5	58	165	282	0	0	0	0	0	0	4	26	38	42	118	0	0
6	58	165	282	0	0	0	0	0	0	4	26	38	42	118	0	0
7	58	165	282	0	0	0	0	0	0	4	26	38	42	118	0	0
8	58	165	282	0	0	0	0	0	0	4	26	38	42	118	0	0
9	58	165	282	0	0	0	0	0	0	4	26	38	42	118	0	0
10	60	368	892	0	0	0	0	0	0	0	96	108	128	342	0	0
11	60	368	892	0	0	0	0	0	0	0	96	108	128	342	0	0
12	60	368	892	0	0	0	0	0	0	0	96	108	128	342	0	0
13	60	368	892	0	0	0	0	0	0	0	96	108	128	342	0	0
14	60	368	892	0	0	0	0	0	0	0	96	108	128	342	0	0

Figure 38. TiO2 dataset example (Part 2)

When working with "TiO2", different features were initially discarded as part of feature analysis. Initial step removed the first 4 values from Section 1 (*Shell Thickness, Maximum number of bonds, Size in Angstrom, and Atomic number of the metal*). Regarding "Size in Angstrom", this information is already provided by X0 feature, as seen above. The other listed features had a standard deviation of 0, which would not help in the classification of training samples.

2 main approaches were taken into consideration: first approach consisted in joining all values under the same section and type, whereas second would consist of just removing those features with 0 values in all samples. Even though additional

As part of the first approach we take for instance the following example: in Section 3 one of the substructures is defined as O (1,1), which would refer to the subgraph where there is an oxygen atom as a central atom connected to another oxygen atom as well as a metal atom. The idea was to group all found subgraphs that had any similar central atom, either metal or oxygen, by joining the results obtained for any combination of metal and oxygen atoms. This reduced the number of features linked to the NanoFingerprint to only 12 different features. Below included a snippet of the remaining DB:

	0	1	2	3	4	5	6	Section2_O	Section2_M	Section3_O	Section3_M	Section4_O	Section4_M	Section4_OM
0	30.0	125.0	1250.0	25.0	-10.0	505.0	228.0	505.0	228.0	505.0	228.0	0.0	0.0	1064.0
1	30.0	102.0	987.0	25.0	-12.0	505.0	228.0	505.0	228.0	505.0	228.0	0.0	0.0	1064.0
2	30.0	281.0	1543.0	50.0	-15.0	505.0	228.0	505.0	228.0	505.0	228.0	0.0	0.0	1064.0
3	30.0	101.0	1045.0	50.0	-9.0	505.0	228.0	505.0	228.0	505.0	228.0	0.0	0.0	1064.0
4	30.0	299.0	1754.0	100.0	-11.0	505.0	228.0	505.0	228.0	505.0	228.0	0.0	0.0	1064.0
5	30.0	134.0	961.0	100.0	-11.0	505.0	228.0	505.0	228.0	505.0	228.0	0.0	0.0	1064.0
6	30.0	600.0	1876.0	200.0	-12.0	505.0	228.0	505.0	228.0	505.0	228.0	0.0	0.0	1064.0
7	30.0	298.0	1165.0	200.0	-12.0	505.0	228.0	505.0	228.0	505.0	228.0	0.0	0.0	1064.0
8	45.0	129.0	2567.0	25.0	-9.0	1320.0	674.0	1320.0	674.0	1320.0	674.0	0.0	0.0	3048.0
9	45.0	129.0	2309.0	25.0	-10.0	1320.0	674.0	1320.0	674.0	1320.0	674.0	0.0	0.0	3048.0
10	45.0	201.0	2431.0	50.0	-9.0	1320.0	674.0	1320.0	674.0	1320.0	674.0	0.0	0.0	3048.0
11	45.0	201.0	2987.0	50.0	-11.0	1320.0	674.0	1320.0	674.0	1320.0	674.0	0.0	0.0	3048.0
12	45.0	451.0	2941.0	100.0	-11.0	1320.0	674.0	1320.0	674.0	1320.0	674.0	0.0	0.0	3048.0
13	45.0	451.0	1934.0	100.0	-9.0	1320.0	674.0	1320.0	674.0	1320.0	674.0	0.0	0.0	3048.0
14	45.0	876.0	1965.0	200.0	-11.0	1320.0	674.0	1320.0	674.0	1320.0	674.0	0.0	0.0	3048.0
15	45.0	876.0	2109.0	200.0	-10.0	1320.0	674.0	1320.0	674.0	1320.0	674.0	0.0	0.0	3048.0
16	125.0	136.0	3215.0	25.0	-11.0	10766.0	5400.0	10766.0	5400.0	10766.0	5400.0	0.0	0.0	24552.0
17	125.0	136.0	2667.0	25.0	-10.0	10766.0	5400.0	10766.0	5400.0	10766.0	5400.0	0.0	0.0	24552.0
18	125.0	149.0	3782.0	50.0	-10.0	10766.0	5400.0	10766.0	5400.0	10766.0	5400.0	0.0	0.0	24552.0
19	125.0	149.0	2144.0	50.0	-15.0	10766.0	5400.0	10766.0	5400.0	10766.0	5400.0	0.0	0.0	24552.0
20	125.0	343.0	3871.0	100.0	-12.0	10766.0	5400.0	10766.0	5400.0	10766.0	5400.0	0.0	0.0	24552.0
21	125.0	343.0	2890.0	100.0	-9.0	10766.0	5400.0	10766.0	5400.0	10766.0	5400.0	0.0	0.0	24552.0
22	125.0	967.0	3813.0	200.0	-9.0	10766.0	5400.0	10766.0	5400.0	10766.0	5400.0	0.0	0.0	24552.0
23	125.0	967.0	2671.0	200.0	-8.0	10766.0	5400.0	10766.0	5400.0	10766.0	5400.0	0.0	0.0	24552.0

Figure 39. TiO2 joined sections dataset

Notice that Section 2 and Section 3 results are identical to the total number of oxygen atoms and metal atoms. Checking the generated verbose file for TiO2_30 for example, it is seen that the substructures found on these sections are mostly referred to either oxygen atoms bonded to only other metal atoms, or metal atoms only bonded to only oxygen atoms.

```
1. Shell thickness in Angstrom:4.000000
2. Maximum number of bounds per atom:8
3. Size in Angstrom:30.069616
4. Atomic number of the metal:22
5. Oxygen atoms:505
6. Metal atoms:228
7. O(1):58
8. O(2):165
9. O(3):282
16. Ti(2):4
17. Ti(3):26
18. Ti(4):38
19. Ti(5):42
20. Ti(6):118
24. O(0,1):58
25. O(0,2):165
26. O(0,3):282
122. Ti(2,0):4
131. Ti(3,0):26
140. Ti(4,0):38
149. Ti(5,0):42
158. Ti(6,0):118
13424. O(0,1)-Ti(4,0):2
13433. O(0,1)-Ti(5,0):28
13442. O(0,1)-Ti(6,0):28
13487. O(0,2)-Ti(2,0):4
13496. O(0,2)-Ti(3,0):32
13505. O(0,2)-Ti(4,0):64
13514. O(0,2)-Ti(5,0):80
13523. O(0,2)-Ti(6,0):150
13568. O(0,3)-Ti(2,0):4
13577. O(0,3)-Ti(3,0):46
13586. O(0,3)-Ti(4,0):86
13595. O(0,3)-Ti(5,0):102
13604. O(0,3)-Ti(6,0):438
```

Figure 40. TiO2_30 Verbose NanoFingerprint

From the figure above, Section 2 is represented by rows 7 – 14, whereas Section 3 is represented by rows 15 – 22. When comparing the results in both sections, it is seen that same values are obtained. Going back to the partially preprocessed dataset, both "section4_O" and "section4_M" are filled with zeros. Correlating this information with the example above, it is seen that substructures found for TiO2 nanocompounds are mostly between oxygen and metal atoms, no substructure was found with labels O(x,y) – O(x',y') or Ti(x,y) – Ti(x',y') to be part of the shell of the nanocompound.

Removing column features for both Section 2 and Section 3 as well as Section4_O and Section4_M leaves the dataset as follows:

1		0	1	2	3	4	5	6	Section4_OM
2	0	30	125	1250	25	-10	505	228	1064
3	1	30	102	987	25	-12	505	228	1064
4	2	30	281	1543	50	-15	505	228	1064
5	3	30	101	1045	50	-9	505	228	1064
6	4	30	299	1754	100	-11	505	228	1064
7	5	30	134	961	100	-11	505	228	1064
8	10	45	201	2431	50	-9	1320	674	3048
9	11	45	201	2987	50	-11	1320	674	3048
10	12	45	451	2941	100	-11	1320	674	3048
11	13	45	451	1934	100	-9	1320	674	3048
12	14	45	876	1965	200	-11	1320	674	3048
13	19	125	149	2144	50	-15	10766	5400	24552
14	20	125	343	3871	100	-12	10766	5400	24552
15	21	125	343	2890	100	-9	10766	5400	24552
16	22	125	967	3813	200	-9	10766	5400	24552
17	23	125	967	2671	200	-8	10766	5400	24552

Figure 41. TiO2 training dataset example by joining sections

About the second approach, consider Figure 37 and Figure 38. As an example, *Number O(6)* has been removed as it would interfere with the output expected. First four values from the NanoFingerprint have been also removed as explained in the previous approach. The remaining features after selecting most significant features in the example are as follows:

1	Name	Eng.Size (X0)	Size in Water (X1)	Size in PBS (X2)	Concentration (X4)	Zeta Potential (X5)	Y - LDH	Number Oxigen Atoms	Number Metal Atoms	Number O(1)	Number O(2)	Number O(3)	Number M(2)	Number M(3)	Number M(4)	Number M(5)	Number M(6)
2	TiO2_30	30	125	1250	25	-10	0.9	505	228	58	165	282	4	26	38	42	118
3	TiO2_30	30	102	987	25	-12	1	505	228	58	165	282	4	26	38	42	118
4	TiO2_30	30	281	1543	50	-15	0.75	505	228	58	165	282	4	26	38	42	118
5	TiO2_30	30	101	1045	50	-9	0.7	505	228	58	165	282	4	26	38	42	118
6	TiO2_30	30	299	1754	100	-11	1.04	505	228	58	165	282	4	26	38	42	118
7	TiO2_30	30	134	961	100	-11	1.09	505	228	58	165	282	4	26	38	42	118
8	TiO2_30	30	600	1876	200	-12	1.15	505	228	58	165	282	4	26	38	42	118
9	TiO2_30	30	298	1165	200	-12	1.2	505	228	58	165	282	4	26	38	42	118
10	TiO2_45	45	129	2567	25	-9	0.9	1320	674	60	368	892	0	96	108	128	342
11	TiO2_45	45	129	2309	25	-10	0.85	1320	674	60	368	892	0	96	108	128	342
12	TiO2_45	45	201	2431	50	-9	0.75	1320	674	60	368	892	0	96	108	128	342
13	TiO2_45	45	201	2987	50	-11	0.78	1320	674	60	368	892	0	96	108	128	342
14	TiO2_45	45	451	2941	100	-11	1.4	1320	674	60	368	892	0	96	108	128	342

Figure 42. TiO2 dataset after preprocessing example

Applying the technique mentioned above across the whole subset, size is reduced to [24, 36]. Example of a training dataset considering split of 70:15:15, assigning 15% of samples to test, 15% to validation, and remaining to training subset has been added below:

1		0	1	2	3	4	5	6	7	8	9	16	17	18	19	20	24	25	26	122
2	12	45	451	2941	100	-11	1320	674	60	368	892	0	96	108	128	342	60	368	892	0
3	21	125	343	2890	100	-9	10766	5400	716	2820	7230	32	464	794	916	3194	716	2820	7230	32
4	0	30	125	1250	25	-10	505	228	58	165	282	4	26	38	42	118	58	165	282	4
5	6	30	600	1876	200	-12	505	228	58	165	282	4	26	38	42	118	58	165	282	4
6	13	45	451	1934	100	-9	1320	674	60	368	892	0	96	108	128	342	60	368	892	0
7	22	125	967	3813	200	-9	10766	5400	716	2820	7230	32	464	794	916	3194	716	2820	7230	32
8	20	125	343	3871	100	-12	10766	5400	716	2820	7230	32	464	794	916	3194	716	2820	7230	32
9	18	125	149	3782	50	-10	10766	5400	716	2820	7230	32	464	794	916	3194	716	2820	7230	32
10	19	125	149	2144	50	-15	10766	5400	716	2820	7230	32	464	794	916	3194	716	2820	7230	32
11	23	125	967	2671	200	-8	10766	5400	716	2820	7230	32	464	794	916	3194	716	2820	7230	32
12	9	45	129	2309	25	-10	1320	674	60	368	892	0	96	108	128	342	60	368	892	0
13	7	30	298	1165	200	-12	505	228	58	165	282	4	26	38	42	118	58	165	282	4
14	16	125	136	3215	25	-11	10766	5400	716	2820	7230	32	464	794	916	3194	716	2820	7230	32
15	2	30	281	1543	50	-15	505	228	58	165	282	4	26	38	42	118	58	165	282	4
16	11	45	201	2987	50	-11	1320	674	60	368	892	0	96	108	128	342	60	368	892	0
17	3	30	101	1045	50	-9	505	228	58	165	282	4	26	38	42	118	58	165	282	4

Figure 43. TiO2 training dataset example (Part 1)

1	122	131	140	149	158	13424	13433	13442	13487	13496	13505	13514	13523	13568	13577	13586	13595	13604
2	0	96	108	128	342	0	24	36	0	96	148	212	280	0	192	284	404	1372
3	32	464	794	916	3194	32	142	542	0	376	1074	1506	2684	64	1016	2070	2932	12114
4	4	26	38	42	118	2	28	28	4	32	64	80	150	4	46	86	102	438
5	4	26	38	42	118	2	28	28	4	32	64	80	150	4	46	86	102	438
6	0	96	108	128	342	0	24	36	0	96	148	212	280	0	192	284	404	1372
7	32	464	794	916	3194	32	142	542	0	376	1074	1506	2684	64	1016	2070	2932	12114
8	32	464	794	916	3194	32	142	542	0	376	1074	1506	2684	64	1016	2070	2932	12114
9	32	464	794	916	3194	32	142	542	0	376	1074	1506	2684	64	1016	2070	2932	12114
10	32	464	794	916	3194	32	142	542	0	376	1074	1506	2684	64	1016	2070	2932	12114
11	32	464	794	916	3194	32	142	542	0	376	1074	1506	2684	64	1016	2070	2932	12114
12	0	96	108	128	342	0	24	36	0	96	148	212	280	0	192	284	404	1372
13	4	26	38	42	118	2	28	28	4	32	64	80	150	4	46	86	102	438
14	32	464	794	916	3194	32	142	542	0	376	1074	1506	2684	64	1016	2070	2932	12114
15	4	26	38	42	118	2	28	28	4	32	64	80	150	4	46	86	102	438
16	0	96	108	128	342	0	24	36	0	96	148	212	280	0	192	284	404	1372
17	4	26	38	42	118	2	28	28	4	32	64	80	150	4	46	86	102	438

Figure 44. TiO2 training dataset example (Part 2)

4.3.2. Papa – ZnO dataset

Same approaches are applied to the subset of only "ZnO" nanocompounds, in this case however, given the different composition of the nanocompound in comparison with "TiO2" dataset, the non-zero features differ. In this example, remaining subset size would be of [7, 147], leaving only to samples for validating the model, whereas when removing 0s the training subset would of size [7, 11] when joining sections.

4.3.3. Papa – TiO2+ZnO dataset

When considering all rows in the dataset, a possible configuration of available training samples when joining sections would be [27, 11]. Size when removing 0s is in this case of [27, 168].

4.3.4. Anantha dataset

From Anatha_db, no subset was required, in this case there were more nanocompounds to be considered, which implied that each model generated would remove different zeros depending on the available training samples used. To grasp the size of a possible available dataset, a possible training dataset observed was [306, 473], leaving a total of 54 samples for testing.

4.4. Model results

Main focus was put on multilinear regression as a quick and direct implementation was possible on the webportal since only the model coefficients had to be. Additional models have been investigated considering the definitions provided in Section 3.4. In this section the results obtained using only data-driven analysis as well as the combination of this analysis and recursive feature elimination are presented.

4.4.1. Papa – TiO2 model

When using the joined dataset without performing any additional feature analysis, the following results were obtained:

Index	train_CCC	train_RMSE	train_MAE	train_r2_score	validation_CCC	validation_RMSE	validation_MAE	validation_r2_score	test_CCC	test_RMSE	test_MAE	test_r2
0	0.9031	0.1161	0.0884	0.8234	0.5232	0.2256	0.1919	0.4578	0.5170	0.1997	0.1727	0.3159
1	0.8376	0.1475	0.1159	0.7205	0.9706	0.0796	0.0736	0.9348	0.3414	0.2195	0.2050	0.1729
2	0.8440	0.1344	0.1096	0.7301	0.8635	0.1768	0.1478	0.7399	0.6312	0.1686	0.1594	0.5122
3	0.8555	0.1236	0.0862	0.7474	0.7248	0.2335	0.1835	0.0646	0.3866	0.2497	0.1903	-0.0702
4	0.9246	0.0963	0.0595	0.8598	0.1987	0.3780	0.3350	0.0271	0.1979	0.2853	0.2176	-0.3966
5	0.9133	0.1162	0.0874	0.8405	0.2970	0.2442	0.1995	0.1722	0.4848	0.2277	0.1737	0.1104
6	0.8867	0.1227	0.0849	0.7964	0.7470	0.2509	0.2192	0.2413	0.2798	0.2522	0.2244	-0.0911
7	0.8073	0.1451	0.1176	0.6769	0.8505	0.1741	0.1549	0.4733	0.4763	0.2064	0.1843	0.2686
8	0.9005	0.1226	0.0888	0.8190	0.4364	0.2753	0.2473	-3.1821	0.3561	0.2096	0.1688	0.2461
9	0.9354	0.0895	0.0723	0.8786	0.4536	0.2655	0.2132	-1.2871	0.2280	0.2498	0.2193	-0.0709
10	0.8867	0.1227	0.0849	0.7964	0.7470	0.2509	0.2192	0.2413	0.2798	0.2522	0.2244	-0.0911
11	0.8441	0.1451	0.1123	0.7302	0.7272	0.1229	0.1162	0.4115	0.2543	0.2228	0.2097	0.1481
12	0.8826	0.1383	0.1029	0.7899	0.3316	0.1643	0.1587	-0.2011	0.5763	0.1986	0.1797	0.3231
13	0.8567	0.1259	0.0894	0.7493	0.8311	0.2003	0.1545	0.7468	0.4897	0.1972	0.1604	0.3324
14	0.8553	0.1420	0.1110	0.7471	0.9159	0.1247	0.1050	0.7444	0.5233	0.2070	0.1791	0.2645

Figure 45. Model metrics using TiO2 dataset joining sections

Analysing the r-squared metric for training, it is observed that nearly in all instances model represents with high fidelity the data provided in validation and test, however it fails to provide consistent results for the unseen data. When checking the r-squared for validation set, it is seen that in most cases model overfitted, since for those instances with negative or close to 0 r-squared, model fails to predict these values properly. Checking CCC, a similar pattern is observed. Only model 3 is seen to show consistency with a model not overfitted, as values for both CCC and r-squared do not extremely differ as seen with the remaining models. Considering MAE and RMSE, the difference across all predicted and expected samples does not show high discrepancy between expected and predicted value.

For the same model but including only the parameters that were flagged as the optimal for LDH prediction for TiO_2 (X0, X4), better results are obtained in comparison with the previous one. Models listed below show in most cases no overfit. Instances with negative r-squared for validation indicate that samples used in validation had no reference with the ones using in training, and model fails to fit the subset. Considering that samples were chosen at random, it was observed that validation samples belonged just to one of the nanocompounds, which lead to higher errors, as seen in both MAE and RMSE. Based on the features used, which were proven in [3] to provide the best results when using Linear Regression, models using only these parameters together with a compacted version of the fingerprint will mostly provide better predictions.

Index	train_CCC	train_RMSE	train_MAE	train_r2_score	validation_CCC	validation_RMSE	validation_MAE	validation_r2_score	test_CCC	test_RMSE	test_MAE	test_r2
0	0.7925	0.1680	0.1300	0.6564	0.6847	0.1249	0.0911	0.5560	0.8308	0.1270	0.1199	0.7238
1	0.8191	0.1372	0.0997	0.6936	0.6901	0.2516	0.2008	0.5241	0.8170	0.1303	0.1229	0.7093
2	0.8401	0.1489	0.1145	0.7242	0.7548	0.2183	0.2087	0.4713	0.8144	0.1445	0.1264	0.6427
3	0.8729	0.1482	0.1068	0.7744	-0.3776	0.2198	0.1876	-1.1927	0.8617	0.1224	0.0975	0.7436
4	0.8103	0.1747	0.1465	0.6811	0.9272	0.0786	0.0702	0.7754	0.9081	0.0959	0.0942	0.8427
5	0.7637	0.1643	0.1296	0.6177	0.8519	0.1703	0.1313	0.6847	0.8455	0.1343	0.1177	0.6913
6	0.8624	0.1463	0.1182	0.7581	0.5742	0.2146	0.1393	0.0776	0.8979	0.1002	0.0976	0.8281
7	0.8498	0.1478	0.1043	0.7388	0.4986	0.2291	0.2014	-0.0159	0.8378	0.1363	0.1167	0.6821
8	0.8541	0.1601	0.1212	0.7454	-0.0067	0.1628	0.1390	-2.2051	0.8354	0.1314	0.1207	0.7045
9	0.8760	0.1409	0.1081	0.7793	0.0115	0.2344	0.1708	-0.8990	0.8381	0.1272	0.1208	0.7230
10	0.8875	0.1312	0.1022	0.7977	0.5466	0.2645	0.2394	-0.3257	0.9396	0.0768	0.0752	0.8991
11	0.8182	0.1575	0.1174	0.6924	0.7294	0.1761	0.1394	0.5264	0.8757	0.1069	0.1045	0.8044
12	0.7850	0.1641	0.1280	0.6461	0.8511	0.1399	0.0993	0.7554	0.8731	0.1128	0.1025	0.7824
13	0.7804	0.1745	0.1441	0.6399	0.9464	0.0755	0.0691	0.8871	0.8951	0.1014	0.0973	0.8241
14	0.8512	0.1460	0.1102	0.7409	0.4316	0.2069	0.1964	0.2575	0.9137	0.0922	0.0872	0.8545

Figure 46.TiO2 modelling results using proposed features by authors

Index	train_CCC	train_RMSE	train_MAE	train_r2_score	validation_CCC	validation_RMSE	validation_MAE	validation_r2_score	test_CCC	test_RMSE	test_MAE	test_r2
0	0.7834	0.1519	0.1269	0.6439	0.9466	0.0722	0.0565	0.9102	0.7358	0.2048	0.1797	0.6202
1	0.8618	0.1292	0.1053	0.7572	0.5858	0.1941	0.1642	0.2402	0.7367	0.2011	0.1678	0.6340
2	0.7589	0.1366	0.1125	0.6115	0.8485	0.1729	0.1504	0.7775	0.6657	0.2184	0.2002	0.5680
3	0.8332	0.1378	0.1062	0.7141	0.8212	0.1486	0.1378	0.5927	0.7704	0.1959	0.1720	0.6526
4	0.7833	0.1481	0.1216	0.6438	0.9090	0.1024	0.1011	0.8019	0.7375	0.2024	0.1734	0.6292
5	0.8472	0.1355	0.1008	0.7348	0.5817	0.2083	0.2063	0.0988	0.7695	0.1990	0.1763	0.6416
6	0.7977	0.1525	0.1314	0.6634	0.9227	0.0801	0.0666	0.8690	0.7124	0.2079	0.1835	0.6087
7	0.8444	0.1271	0.1063	0.7307	-0.1031	0.1889	0.1727	-0.7316	0.7229	0.2050	0.1769	0.6196
8	0.8066	0.1498	0.1207	0.6759	0.8606	0.0865	0.0805	0.7497	0.7452	0.2025	0.1748	0.6288
9	0.8444	0.1271	0.1063	0.7307	-0.1031	0.1889	0.1727	-0.7316	0.7229	0.2050	0.1769	0.6196
10	0.8472	0.1355	0.1008	0.7348	0.5817	0.2083	0.2063	0.0988	0.7695	0.1990	0.1763	0.6416
11	0.7834	0.1519	0.1269	0.6439	0.9466	0.0722	0.0565	0.9102	0.7358	0.2048	0.1797	0.6202
12	0.8644	0.1331	0.0976	0.7612	0.1022	0.2130	0.1807	-6.2376	0.7797	0.1954	0.1751	0.6542
13	0.8016	0.1504	0.1240	0.6689	0.8995	0.0958	0.0881	0.8093	0.7468	0.1994	0.1701	0.6399
14	0.8644	0.1331	0.0976	0.7612	0.1022	0.2130	0.1807	-6.2376	0.7797	0.1954	0.1751	0.6542

Figure 47.TiO2 modelling results without features with std=0

Index	train_CCC	train_RMSE	train_MAE	train_r2_score	validation_CCC	validation_RMSE	validation_MAE	validation_r2_score	test_CCC	test_RMSE	test_MAE	test_r2
0	0.8073	0.1386	0.1053	0.6768	0.7130	0.1669	0.1558	0.6371	0.6970	0.1800	0.1621	0.5040
1	0.8819	0.1327	0.1116	0.7887	0.1559	0.1929	0.1604	-0.2916	0.4595	0.2470	0.2227	0.0663
2	0.7092	0.1493	0.1236	0.5495	0.8928	0.1499	0.1412	0.8311	0.4980	0.2182	0.2065	0.2713
3	0.8047	0.1459	0.1210	0.6733	0.9047	0.1434	0.1374	0.8237	0.4981	0.2295	0.2148	0.1937
4	0.8779	0.1362	0.1080	0.7823	0.7327	0.1850	0.1805	0.1024	0.6279	0.2070	0.1832	0.3442
5	0.8540	0.1393	0.1119	0.7452	0.7889	0.1623	0.1539	0.6629	0.6580	0.1896	0.1732	0.4499
6	0.8227	0.1476	0.1159	0.6988	0.7341	0.1242	0.1220	0.3484	0.6105	0.2116	0.1882	0.3145
7	0.8434	0.1277	0.1145	0.7292	0.6239	0.2267	0.2006	0.5312	0.3353	0.2651	0.2453	-0.0759
8	0.8166	0.1405	0.1074	0.6901	0.8809	0.1591	0.1425	0.7992	0.6706	0.1872	0.1703	0.4634
9	0.8434	0.1277	0.1145	0.7292	0.6239	0.2267	0.2006	0.5312	0.3353	0.2651	0.2453	-0.0759
10	0.8318	0.1394	0.1085	0.7120	0.6553	0.1655	0.1486	0.3228	0.6548	0.1898	0.1746	0.4484
11	0.8600	0.1364	0.1038	0.7544	0.5338	0.1944	0.1758	-0.5882	0.6399	0.2059	0.1731	0.3513
12	0.8549	0.1452	0.1178	0.7465	0.8204	0.1444	0.1366	0.5771	0.5788	0.2168	0.1976	0.2807
13	0.8591	0.1503	0.1255	0.7529	0.7396	0.1262	0.1218	-0.2134	0.5494	0.2221	0.2046	0.2448
14	0.8587	0.1263	0.1132	0.7525	0.6614	0.2186	0.1879	0.5452	0.4214	0.2442	0.2223	0.0872

Figure 48.TiO2 modelling results with proposed features in article

Performance of models when removing features with null standard deviation is more consistent across models. As per values of CCC and r-squared, models represent most of the samples in all instances. Even though at first glance it may appear that leaving only the proposed parameters for TiO2 calculation + NanoFingerprint does perform better, results associated to test samples show that most models do not manage to represent this subset. R-squared is lower than 0.50 in all instances, but that does not necessarily imply model is overfitting, it just implies that it does not fully represent the test subset. The error linked to misclassification is higher as per RMSE and MAE. Overall, in comparison with using the previous approach, models do provide better model generalisation.

Main problems attributed to TiO2 subset is linked to the small dataset available and high number of features. In most instances, especially when using joining sections approach, models tended to overfit and provide low results for both testing and validation.

As an alternative to the previous analysis, recursive feature elimination applied to the dataset was considered. recursively removing features that would not be useful for the model.

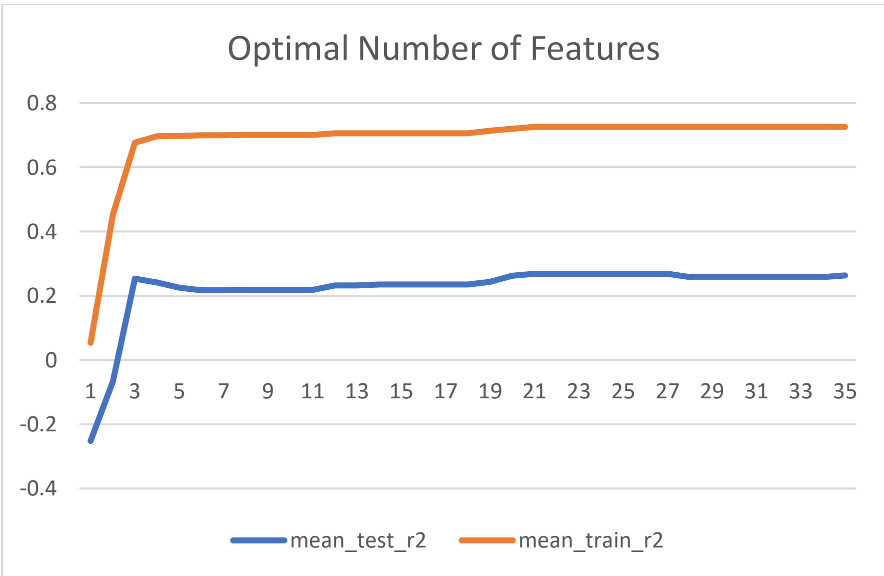


Figure 49. RFE applied to TiO2 dataset

Considering the graph above, it is seen that the model performs better when using more than 20 features. Each number of features has a rank assigned, which helps to determine what model performed better and what number of features was used. In this case the best model was obtained when using 21 features. The metric used to validate these results was r-squared. It is seen that the second best in the ranking was when using 26 features. Below are the metrics associated to these 2 models:

#Features	CCC	RMSE	MAE	R^2
21	0.8366444	0.175046	0.127087	0.593989
26	0.8366444	0.175046	0.127087	0.593989

Table 2. TiO2 results using RFE

Even though models appear to provide the same results for the metrics used, there's a slight difference. Full value is included in the appendix B.

4.4.2. Papa – ZnO model

Index	train_CCC	train_RMSE	train_MAE	train_r2_score	validation_CCC	validation_RMSE	validation_MAE	validation_r2_score	test_CCC	test_RMSE	test_MAE	test_r2
0	1	9.74E-16	8.66E-16	1	0.2136	0.1750	0.1685	-14.1301	0.8197	0.0512	0.0491	0.6367
1	1	3.91E-16	3.11E-16	1	0.1862	0.4569	0.3793	-68.0164	-0.7912	0.2411	0.2360	-7.0431
2	1	0	0	1	-0.2237	0.1236	0.1097	-0.6926	-0.0844	0.1842	0.1543	-3.6941
3	1	3.02E-16	2.89E-16	1	0.2136	0.1750	0.1685	-14.1301	0.8197	0.0512	0.0491	0.6367
4	1	1.79E-16	1.55E-16	1	-0.6667	0.1588	0.1426	-4.9669	-0.7698	0.2263	0.2160	-6.0867
5	1	4.97E-17	2.22E-17	1	0.1200	0.0979	0.0931	-3.7365	0.0583	0.1571	0.1374	-2.4147
6	1	1.72E-16	1.33E-16	1	-0.5303	0.4002	0.3983	-18.7740	0.5442	0.2004	0.1970	-4.5592
7	1	1.49E-16	1.11E-16	1	0.3251	0.1833	0.1673	-8.3335	0.6275	0.0892	0.0890	-0.1016
8	1	4.97E-17	2.22E-17	1	0.3251	0.1833	0.1673	-8.3335	0.6275	0.0892	0.0890	-0.1016
9	1	1.11E-16	6.66E-17	1	-0.0377	0.1210	0.1060	-584.2731	-0.7541	0.1866	0.1731	-3.8194
10	1	3.22E-16	2.22E-16	1	0.1862	0.4569	0.3793	-68.0164	-0.7912	0.2411	0.2360	-7.0431
11	1	0	0	1	0.6472	0.1204	0.1147	0.2049	-0.1990	0.2010	0.1579	-4.5925
12	1	1.11E-16	6.66E-17	1	0.2136	0.1750	0.1685	-14.1301	0.8197	0.0512	0.0491	0.6367
13	1	1.11E-16	6.66E-17	1	0.1200	0.0979	0.0931	-3.7365	0.0583	0.1571	0.1374	-2.4147
14	1	3.02E-16	2.89E-16	1	0.2136	0.1750	0.1685	-14.1301	0.8197	0.0512	0.0491	0.6367

Figure 50. ZnO modelling results using joined sections

Index	train_CCC	train_RMSE	train_MAE	train_r2_score	validation_CCC	validation_RMSE	validation_MAE	validation_r2_score	test_CCC	test_RMSE	test_MAE	test_r2
0	1	5.73E-16	4.66E-16	1	-0.1110	0.3305	0.2950	-24.8494	0.0454	0.7532	0.7203	-225.9108
1	1	0	0	1	0.1266	0.1260	0.1211	0.0606	0.2596	0.0887	0.0858	-2.1449
2	1	1.11E-16	6.66E-17	1	-0.9204	0.1371	0.1363	-1.9375	-0.0110	0.2174	0.2103	-17.9055
3	1	1.72E-16	1.33E-16	1	-0.9204	0.1371	0.1363	-1.9375	-0.0110	0.2174	0.2103	-17.9055
4	1	1.11E-16	6.66E-17	1	0.5043	0.1584	0.1560	-1.5093	0.0287	0.2816	0.2803	-30.7199
5	1	9.93E-17	4.44E-17	1	-0.9204	0.1371	0.1363	-1.9375	-0.0110	0.2174	0.2103	-17.9055
6	1	7.02E-17	4.44E-17	1	0.6245	0.0981	0.0904	-2.1801	0.0277	0.2749	0.2735	-29.2370
7	1	1.99E-16	1.78E-16	1	0.4134	0.1841	0.1621	-10.2056	0.1929	0.2265	0.2145	-19.5179
8	1	4.28E-15	3.40E-15	1	0.0068	3.8248	3.5645	-11941.3584	0.0030	10.5325	10.0120	-44372.2126
9	1	2.33E-16	1.78E-16	1	0.2043	0.1704	0.1664	-71.6232	-0.2309	0.0857	0.0638	-1.9384
10	1	1.49E-16	1.11E-16	1	0.4134	0.1841	0.1621	-10.2056	0.1929	0.2265	0.2145	-19.5179
11	1	7.02E-17	4.44E-17	1	0.2043	0.1704	0.1664	-71.6232	-0.2309	0.0857	0.0638	-1.9384
12	1	4.68E-16	3.77E-16	1	0.2581	0.6572	0.6512	-42.1885	-0.0100	0.8172	0.8088	-266.0954
13	1	9.93E-17	4.44E-17	1	-0.1168	0.1288	0.1116	-7.1861	0.0255	0.2628	0.2609	-26.6339
14	1	4.28E-15	3.40E-15	1	0.0068	3.8248	3.5645	-11941.3584	0.0030	10.5325	10.0120	-44372.2126

Figure 51. ZnO modelling results using proposed features in article

As per the results seen above, model overfits in all instances, this is clearly seen analysing the results associated to the training dataset, where it is seen RMSE and MAE is close to 0. This implies that the models perfectly fit the data used for training. In contrast, results for both validation and test, especially when analysing the r-squared parameter show that these models do not represent these subsets. Negative r-squared is normally linked to missing bias, however all coefficients were considered when creating these models. The main reason behind these results is the low sample availability. Dataset is only formed by 9 samples, which makes it difficult to fit a model.

Index	train_CCC	train_RMSE	train_MAE	train_r2_score	validation_CCC	validation_RMSE	validation_MAE	validation_r2_score	test_CCC	test_RMSE	test_MAE	test_r2
0	1	1.79E-16	1.11E-16	1	0.0017	0.2786	0.2777	-123.2258	-0.6332	0.3360	0.3331	-12.9364
1	1	1.07E-15	8.88E-16	1	0.2222	0.7083	0.6594	-54.5843	0.0532	0.6424	0.6412	-49.9444
2	1	3.18E-16	2.00E-16	1	0.1494	0.1205	0.1202	0.1404	-0.1216	0.0956	0.0955	-0.1273
3	1	6.30E-16	6.00E-16	1	-0.5938	0.2616	0.2611	-15.1993	0.4759	0.2538	0.2350	-6.9530
4	1	1.12E-15	9.77E-16	1	0.0062	0.2719	0.2688	-2956.7265	-0.3437	0.3724	0.2871	-16.1220
5	1	1.40E-16	8.88E-17	1	0.3716	0.1443	0.1391	-1.0817	-0.8807	0.2298	0.2273	-5.5176
6	1	1.99E-16	1.78E-16	1	0.2315	0.1504	0.1504	-55.5834	-0.4788	0.1130	0.1130	-0.5755
7	1	6.28E-16	5.33E-16	1	-0.6026	0.2835	0.2689	-11.5616	0.1991	0.1423	0.1280	-1.4988
8	1	9.93E-17	4.44E-17	1	0.1286	0.1218	0.1216	0.1221	-0.3198	0.1049	0.1048	-0.3581
9	1	5.62E-15	5.06E-15	1	-0.0255	0.7656	0.7438	-161.8103	0.2203	0.7166	0.7162	-62.3939
10	1	1.58E-15	1.38E-15	1	-0.5584	0.2719	0.2687	-16.4948	-0.4573	0.4490	0.4415	-23.8877
11	1	1.79E-15	1.55E-15	1	0.2222	0.7083	0.6594	-54.5843	0.0532	0.6424	0.6412	-49.9444
12	1	1.72E-16	1.33E-16	1	-0.9647	0.0895	0.0887	-2.9559	-0.8819	0.2294	0.2270	-5.4981
13	1	5.62E-15	5.06E-15	1	-0.0255	0.7656	0.7438	-161.8103	0.2203	0.7166	0.7162	-62.3939
14	1	1.07E-15	8.88E-16	1	0.2222	0.7083	0.6594	-54.5843	0.0532	0.6424	0.6412	-49.9444

Figure 52. ZnO modelling results without features with std = 0

Index	train_CCC	train_RMSE	train_MAE	train_r2_score	validation_CCC	validation_RMSE	validation_MAE	validation_r2_score	test_CCC	test_RMSE	test_MAE	test_r2
0	1	2.31E-15	2.22E-15	1	-0.2261	0.3911	0.3114	-49.5685	0.0074	0.3983	0.3817	-6343.6067
1	1	1.07E-15	1.04E-15	1	-0.4669	0.1668	0.1615	-21.7050	-0.0671	0.1263	0.1054	-637.3584
2	1	2.67E-16	2.00E-16	1	-0.7460	0.1851	0.1711	-3.7421	-0.0420	0.1192	0.1016	-567.4191
3	1	8.31E-16	6.22E-16	1	-0.2506	0.3596	0.2768	-11.9291	0.0080	0.2987	0.2912	-3568.6766
4	1	3.48E-16	2.44E-16	1	-0.0762	0.1869	0.1728	-37.8077	-0.0307	0.0515	0.0499	-105.0978
5	1	1.70E-15	1.47E-15	1	-0.2285	0.4036	0.3006	-24.4573	0.0062	0.3448	0.3377	-4754.9200
6	1	3.88E-16	2.89E-16	1	-0.0307	0.2634	0.2554	-55.6288	-0.0428	0.0956	0.0856	-364.6176
7	1	5.41E-15	4.15E-15	1	-0.0011	0.2048	0.2044	-418.3495	0.0135	0.3822	0.3285	-5841.8663
8	1	6.90E-16	5.55E-16	1	-0.0167	0.2466	0.2392	-607.2517	-0.0307	0.1736	0.1455	-1203.9182
9	1	7.51E-16	6.00E-16	1	-0.0744	0.1954	0.1808	-41.4316	-0.0601	0.0459	0.0427	-83.3262
10	1	2.81E-16	1.78E-16	1	-0.0727	0.2432	0.2193	-93.6542	-0.0996	0.1047	0.1042	-437.1006
11	1	5.25E-16	4.44E-16	1	-0.2506	0.3596	0.2768	-11.9291	0.0080	0.2987	0.2912	-3568.6766
12	1	2.03E-15	1.47E-15	1	-0.2285	0.4036	0.3006	-24.4573	0.0062	0.3448	0.3377	-4754.9200
13	1	8.31E-16	6.22E-16	1	-0.2506	0.3596	0.2768	-11.9291	0.0080	0.2987	0.2912	-3568.6766
14	1	3.88E-16	3.77E-16	1	-0.0762	0.1869	0.1728	-37.8077	-0.0307	0.0515	0.0499	-105.0978

Figure 53. ZnO modelling results using proposed features in article

Similar to the pervious approach, models overfit – only 2 features have been removed from previous instance. Distance between predicted samples and actual values are seen to be considerable as per RMSE and MAE.

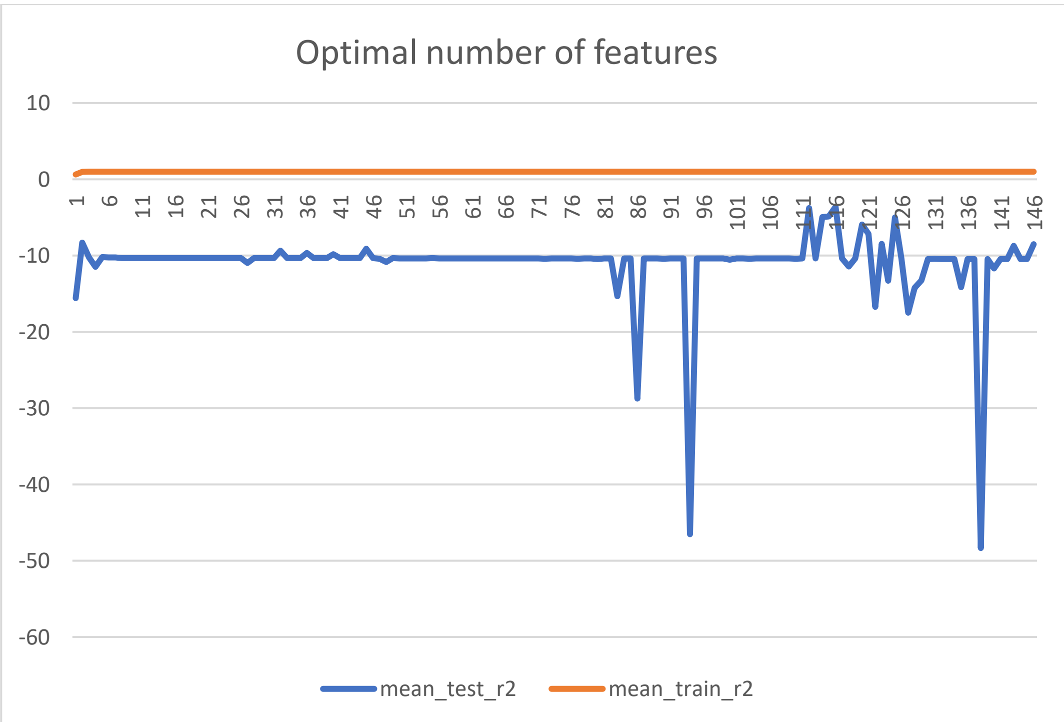


Figure 54. RFE applied to ZnO dataset

Considering the graph above, we observe similar results as the ones seen using data-driven analysis. Low sample availability for subset does not provide reliable models. Notice the spikes when using either 86, 94 or 138 features. Based on the metrics gathered from RFE, it is seen that 2 models had a better performance than the rest, however there are still not fitting the data analysed.

#Features	CCC	RMSE	MAE	R^2
112	0.43203	0.40218	0.34431	-4.94128
116	0.43203	0.40218	0.34431	-4.94128

5. Table 3. ZnO results using RFE

4.4.3. Papa – TiO2+ZnO model

Index	train_CCC	train_RMSE	train_MAE	train_r2_score	validation_CCC	validation_RMSE	validation_MAE	validation_r2_score	test_CCC	test_RMSE	test_MAE	test_r2
0	0.8468	0.1388	0.1104	0.7343	0.7598	0.1259	0.1228	0.4435	0.5001	0.1518	0.1511	-0.2623
1	0.8817	0.1129	0.0833	0.7885	0.4220	0.2471	0.2010	-0.4748	0.4686	0.1788	0.1775	-0.7506
2	0.8921	0.1157	0.0870	0.8052	0.3669	0.2447	0.1837	-0.3141	0.4949	0.1675	0.1650	-0.5366
3	0.8439	0.1264	0.1016	0.7300	0.8097	0.1804	0.1540	0.6291	0.5944	0.1327	0.1223	0.0354
4	0.8357	0.1391	0.1040	0.7178	0.7350	0.1486	0.1181	0.4524	0.5754	0.1340	0.1317	0.0165
5	0.8704	0.1281	0.1021	0.7706	0.4978	0.1750	0.1349	0.0573	0.6007	0.1391	0.1341	-0.0600
6	0.8601	0.1267	0.0844	0.7546	0.5200	0.2072	0.1819	-0.4715	0.4529	0.1641	0.1618	-0.4743
7	0.8535	0.1302	0.0990	0.7444	0.3878	0.1842	0.1457	-1.1609	0.5336	0.1510	0.1479	-0.2481
8	0.8061	0.1377	0.1159	0.6751	0.9149	0.1225	0.1003	0.8328	0.5490	0.1476	0.1411	-0.1931
9	0.8538	0.1384	0.1094	0.7448	0.6385	0.1641	0.1568	-0.6076	0.6468	0.1377	0.1273	-0.0389
10	0.9383	0.0881	0.0785	0.8838	-0.2190	0.3004	0.2407	-1.2416	0.5675	0.1406	0.1349	-0.0828
11	0.8850	0.1130	0.0873	0.7938	0.3201	0.2361	0.2236	-0.0663	0.7051	0.1020	0.0955	0.4305
12	0.8203	0.1385	0.1114	0.6954	0.7890	0.1426	0.1413	0.4441	0.6032	0.1410	0.1361	-0.0891
13	0.8468	0.1388	0.1104	0.7343	0.7598	0.1259	0.1228	0.4435	0.5001	0.1518	0.1511	-0.2623
14	0.8606	0.1372	0.1027	0.7553	0.2468	0.1708	0.1453	-3.7201	0.5642	0.1467	0.1401	-0.1788

Figure 55. TiO2+ZnO modelling results when joining sections

Index	train_CCC	train_RMSE	train_MAE	train_r2_score	validation_CCC	validation_RMSE	validation_MAE	validation_r2_score	test_CCC	test_RMSE	test_MAE	test_r2
0	0.8132	0.1442	0.1173	0.6853	0.8072	0.1287	0.0975	0.6737	0.7779	0.1150	0.1016	0.6487
1	0.7944	0.1450	0.1197	0.6589	0.8993	0.1115	0.0759	0.8317	0.7511	0.1227	0.1092	0.6003
2	0.7791	0.1402	0.1099	0.6381	0.8849	0.1414	0.1055	0.8029	0.6616	0.1509	0.1389	0.3956
3	0.8356	0.1237	0.0975	0.7176	0.6969	0.2104	0.1537	0.5499	0.7206	0.1271	0.1117	0.5710
4	0.8520	0.1266	0.0928	0.7421	0.5507	0.2048	0.1483	0.4087	0.6810	0.1439	0.1284	0.4502
5	0.8225	0.1499	0.1239	0.6985	0.8942	0.0663	0.0540	0.6587	0.7719	0.1226	0.1137	0.6009
6	0.7874	0.1498	0.1210	0.6493	0.9633	0.0654	0.0569	0.9326	0.7500	0.1249	0.1110	0.5856
7	0.8454	0.1338	0.0964	0.7322	0.5175	0.1863	0.1751	-0.2266	0.6161	0.1755	0.1732	0.1822
8	0.8320	0.1441	0.1071	0.7124	0.1391	0.1315	0.1048	-3.5341	0.7739	0.1224	0.1137	0.6021
9	0.8666	0.1136	0.0804	0.7646	0.4863	0.2293	0.1975	0.3562	0.7125	0.1321	0.1131	0.5364
10	0.8218	0.1324	0.0996	0.6976	0.6259	0.1786	0.1322	0.0469	0.6788	0.1522	0.1417	0.3847
11	0.8516	0.1271	0.0944	0.7416	0.5667	0.2018	0.1370	0.4140	0.6883	0.1421	0.1269	0.4634
12	0.8165	0.1309	0.0998	0.6899	0.8456	0.1720	0.1372	0.7105	0.7741	0.1202	0.1083	0.6160
13	0.8265	0.1438	0.1082	0.7043	0.7873	0.1278	0.1131	0.4984	0.7918	0.1205	0.1165	0.6142
14	0.8265	0.1438	0.1082	0.7043	0.7873	0.1278	0.1131	0.4984	0.7918	0.1205	0.1165	0.6142

Figure 56. TiO2+ZnO modelling results using proposed features in article

When joining both dataset under a single set, it is observed that no models fit the analysed data, considering CCC and r-squared metrics, it is seen that the results associated to training are way higher than the other sets. In both scenarios, RMSE and MAE should low impact on error. Regarding the second model, better results are obtained, in this case we are using the features proposed by the paper as well as a compacted NanoFingerprint. There are a couple of scenarios where r-squared is seen to be negative, which again indicates overfitting.

Index	train_CCC	train_RMSE	train_MAE	train_r2_score	validation_CCC	validation_RMSE	validation_MAE	validation_r2_score	test_CCC	test_RMSE	test_MAE	test_r2
0	0.8669	0.1256	0.0943	0.7651	0.4588	0.1926	0.1548	-0.2908	0.6704	0.1713	0.1447	0.5038
1	0.8234	0.1408	0.1055	0.6998	0.3075	0.1198	0.1075	-1.2067	0.8039	0.1316	0.1169	0.7071
2	0.8156	0.1233	0.0946	0.6887	0.7559	0.1953	0.1837	0.6276	0.6369	0.1754	0.1512	0.4797
3	0.8202	0.1414	0.1077	0.6951	0.8523	0.1156	0.0948	0.6243	0.8237	0.1263	0.1141	0.7304
4	0.7609	0.1386	0.1071	0.6140	0.9078	0.1265	0.1070	0.8514	0.6871	0.1601	0.1368	0.5667
5	0.8392	0.1382	0.1072	0.7229	0.6346	0.1488	0.1418	0.0351	0.7514	0.1495	0.1132	0.6220
6	0.8534	0.1280	0.1075	0.7442	0.6274	0.1901	0.1704	0.0282	0.8195	0.1280	0.1131	0.7229
7	0.8278	0.1380	0.1088	0.7062	0.5748	0.1626	0.1568	0.1206	0.7046	0.1625	0.1195	0.5533
8	0.8392	0.1382	0.1072	0.7229	0.6346	0.1488	0.1418	0.0351	0.7514	0.1495	0.1132	0.6220
9	0.7609	0.1386	0.1071	0.6140	0.9078	0.1265	0.1070	0.8514	0.6871	0.1601	0.1368	0.5667
10	0.7884	0.1398	0.1089	0.6507	0.8604	0.1377	0.1306	0.7294	0.6084	0.1842	0.1539	0.4262
11	0.7609	0.1386	0.1071	0.6140	0.9078	0.1265	0.1070	0.8514	0.6871	0.1601	0.1368	0.5667
12	0.8499	0.1324	0.1010	0.7390	0.4084	0.1715	0.1637	-0.5640	0.7632	0.1483	0.1326	0.6279
13	0.8336	0.1308	0.1090	0.7144	-1.85E-12	40598895679	18156378110	-2.61E+22	-7.49E-13	57415508886	36312756220	-5.57E+22
14	0.8597	0.1252	0.1042	0.7549	0.4323	0.1985	0.1869	-0.2548	0.7854	0.1419	0.1185	0.6595

Figure 57. TiO2+ZnO modelling results without features with std = 0

Checking the model below, where those features with std equal to 0 have been removed, it is observed that multiple models have good CCC and r-squared metrics. RMSE and Mae is higher than previous models, but models still manage to fit most of the unseen data.

Index	train_CCC	train_RMSE	train_MAE	train_r2_score	validation_CCC	validation_RMSE	validation_MAE	validation_r2_score	test_CCC	test_RMSE	test_MAE	test_r2
0	0.8390	0.1210	0.0966	0.7227	0.8824	0.1355	0.1285	0.6974	0.6012	0.2719	0.2638	-0.2532
1	0.9097	0.0949	0.0843	0.8344	0.3812	0.2203	0.1677	-0.5622	0.6071	0.2760	0.2667	-0.2910
2	0.8545	0.1110	0.0922	0.7460	0.8091	0.1628	0.1486	0.5503	0.6154	0.2752	0.2695	-0.2840
3	0.8892	0.1088	0.0890	0.7998	0.2086	0.1722	0.1362	-0.4319	0.5597	0.2842	0.2765	-0.3694
4	0.8766	0.1020	0.0798	0.7803	0.6677	0.1896	0.1721	0.5771	0.6635	0.2218	0.2117	0.1661
5	0.9019	0.0964	0.0840	0.8214	0.4732	0.2152	0.1569	0.0798	0.6156	0.2628	0.2512	-0.1707
6	0.9130	0.0943	0.0779	0.8399	0.5162	0.2358	0.2049	-0.3281	0.4870	0.3281	0.3175	-0.8250
7	0.8754	0.1088	0.0909	0.7784	0.4380	0.1743	0.1303	0.0687	0.5613	0.2802	0.2702	-0.3304
8	0.8912	0.1043	0.0877	0.8037	0.6283	0.2048	0.1699	-0.0145	0.5184	0.3219	0.3123	-0.7563
9	0.8545	0.1110	0.0922	0.7460	0.8091	0.1628	0.1486	0.5503	0.6154	0.2752	0.2695	-0.2840
10	0.8826	0.0996	0.0784	0.7899	0.4869	0.1818	0.1619	0.3913	0.5733	0.2677	0.2584	-0.2142
11	0.8690	0.1163	0.0943	0.7683	0.8079	0.1346	0.1091	0.4885	0.5876	0.2779	0.2707	-0.3086
12	0.8373	0.1237	0.1020	0.7202	0.9400	0.0922	0.0747	0.8344	0.6153	0.2679	0.2601	-0.2160
13	0.8657	0.1221	0.1004	0.7632	0.6234	0.0967	0.0810	0.3193	0.5504	0.2935	0.2859	-0.4596
14	0.9235	0.0849	0.0695	0.8579	0.3409	0.2616	0.2192	0.0378	0.7042	0.2003	0.1653	0.3203

Figure 58. TiO2+ZnO modelling results when using proposed features in article

When using the parameters advised within the paper as optimal for regression models, it is seen that r-squared for testing responds to negative values, which indicates that models create do not fit the data analysed. Checking CCC and r-squared parameters for training, we can observe a huge difference between values when comparing it to testing samples. This is an indicator of possible model being overfitted.

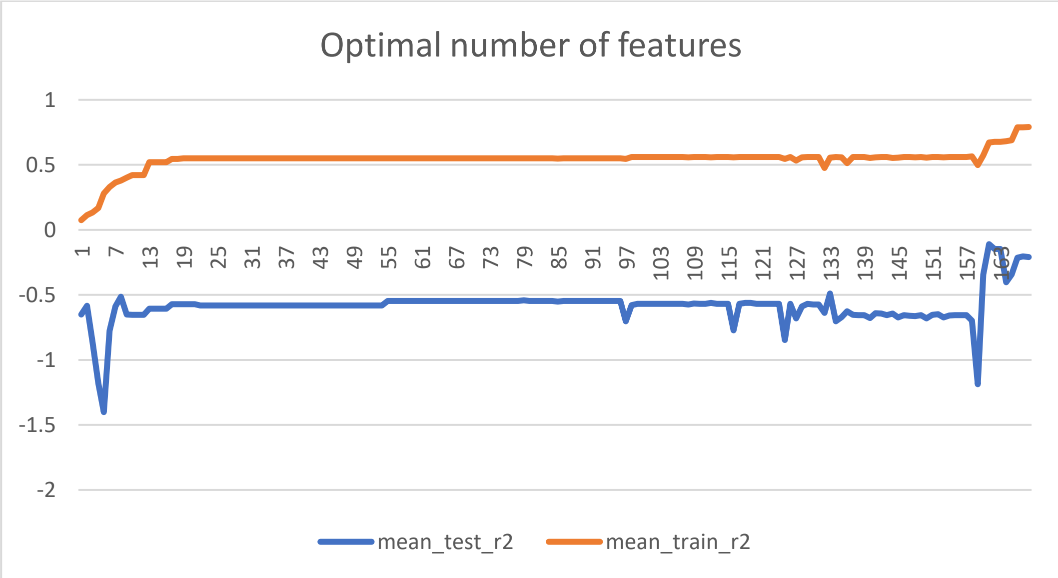


Figure 59. RFE applied to TiO2+ZnO dataset

When applying RFE to the whole dataset, we observe that all r-squared values for the validation dataset is below 0. No changes are perceived in model performance between using 13 features and using 150 approximately. As part rank provided by the technique, it was observed that best ranked features were found when using 160 or more feature. In this case we had around 168 features to analyse and find the optimal ones for the model. When checking the results obtained for the case of use 161 and 163, we observed the following:

#Features	CCC	RMSE	MAE	R^2
161	0.3986080	0.2526117	0.20380952	0.23287
163	0.2544713	0.3305967	0.2968154	-0.313

Table 4. TiO2+ZnO results using RFE

4.4.4. Anantha model

Index	train_RMSE	train_MAE	train_r2_score	validation_RMSE	validation_MAE	validation_r2_score	test_RMSE	test_MAE	test_r2
0	0.40336	0.16270	-0.80543	0.30429	0.09259	-0.22921	0.50452	0.25455	-1.52393
1	0.48795	0.23810	-1.16708	0.47140	0.22222	-1.31423	0.57208	0.32727	-1.41279
2	0.33923	0.11508	0.31535	0.36004	0.12963	-0.14136	0.38139	0.14545	0.18161
3	0.35074	0.12302	0.10261	0.38490	0.14815	-0.06764	0.40452	0.16364	-0.06348
4	0.50787	0.25794	-3.55007	0.49065	0.24074	-2.36793	0.50452	0.25455	-3.19601
5	0.30861	0.09524	0.34585	0.33333	0.11111	-0.01112	0.40452	0.16364	0.16322
6	0.38832	0.15079	-0.32849	0.40825	0.16667	-0.03909	0.42640	0.18182	-0.21833
7	0.32121	0.10317	0.34251	0.36004	0.12963	0.21424	0.40452	0.16364	0.20704
8	0.35635	0.12698	-0.05856	0.40825	0.16667	-1.15287	0.44721	0.20000	-0.09383
9	0.30211	0.09127	0.32460	0.40825	0.16667	0.25139	0.40452	0.16364	0.18478
10	0.29547	0.08730	0.33676	0.36004	0.12963	0.11367	0.38139	0.14545	0.15736
11	0.33923	0.11508	0.31535	0.36004	0.12963	-0.14136	0.38139	0.14545	0.18161
12	0.38832	0.15079	-0.32849	0.40825	0.16667	-0.03909	0.42640	0.18182	-0.21833
13	0.32733	0.10714	0.31940	0.36004	0.12963	-0.49289	0.42640	0.18182	-0.07571
14	0.38318	0.14683	-0.42213	0.43033	0.18519	-0.65147	0.42640	0.18182	-0.41614

Figure 60. Anantha modelling results

Given that Anantha is formed by multiple nancompounds, using LinearRegression does not manage to fully fit all sets. Considering r-squared, it is seen that most of the model provide negative value for this metric. In this case, metric CCC has not been used and analysis of models has been performed considering the other 3 remaining metrics. The decision behind this exclusion is linked to the fact that the toxicity prediction associated with this article is categorical – if the output has a value higher than 0.5 then nanocompound is considered toxic.

After presenting all models below is a table with metrics comparing our results with the results present in the different articles:

Study	**CCC**	**RMSE**	**r-squared**
Papa – TiO_2 LinearRegression	0.94	0.10	0.90
Our Study: Nanofingerprint + Papa (Joined Sections)	0.5763	0.1986	0.3231
Our Study: Nanofingerprint + Papa (Proposed parameters + Joined Sections)	0.9396	0.07	0.8991
Our Study: Nanofingerprint + Papa (Feature removal with std=0)	0.7797	0.1954	0.6542
Our Study: Nanofingerprint + Papa (Proposed parameters + feature removal)	0.697	0.18	0.50
Our Study: NanoFingerprint + Papa (RFE)	0.8364	0.175	0.5939

Figure 61. Compiled results for TiO_2

Study	**CCC**	**RMSE**	**r-squared**
Papa – ZnO LinearRegression	0.99	0.04	0.97
Our Study: Nanofingerprint + Papa (Joined Sections)	0.8197	0.0512	0.6367
Our Study: Nanofingerprint + Papa	0.25	0.08	-2.1

(Proposed parameters + Joined Sections)			
Our Study: Nanofingerprint + Papa (Feature removal with std=0)	-0.3198	0.10	-0.3585
Our Study: Nanofingerprint + Papa (Proposed parameters + feature removal)	-0.06	0.04	-83
Our Study: NanoFingerprint + Papa (RFE)	0.43	0.40	-4.94

Figure 62. Compiled results for ZnO

Study	**CCC**	**RMSE**	**r-squared**
Papa – TiO2+ZnO LinearRegression	0.92	0.08	0.96
Our Study: Nanofingerprint + Papa (Joined Sections)	0.7051	0.1020	0.43
Our Study: Nanofingerprint + Papa (Proposed parameters + Joined Sections)	0.7918	0.12	0.61
Our Study: Nanofingerprint + Papa (Feature removal with std=0)	0.8237	0.1263	0.7304
Our Study: Nanofingerprint + Papa (Proposed parameters + feature removal)	0.7042	0.2003	0.3203
Our Study: NanoFingerprint + Papa (RFE)	0.39	0.25	0.23

Figure 63. Compiled results for TiO2+ZnO

6. Conclusions & Future Work

In this project multiple functionalities have been developed for the production of the web portal framework. An initial assessment was performed to understand the required for the webportal development. Website is currently up and running and publicly accessible enabling users to compute NanoFingerprints or determine the Toxicity level for a given nanocompound. Back-end has been developed considering any future modifications in terms of adding either new functionalities or providing additional files to the end user. By creating subdirectories for each interaction with the portal itself we have ensured that the same data is not accessed by different users at the same time. The development of a local environment allowed to test any functionality without compromising access to the web portal.

The analysis of different toxicity prediction algorithms has been performed and further research on the usage of the NanoFingerprint for regression models have been presented. As per results obtained it is remarkable to state that the limitation on samples available for training and testing the model has led to poor results in comparison to the ones provided in the multiple articles studied. The introduction of Recursive Feature Elimination as means of selecting features that could benefit the model has provided a different perspective. It was observed for example that on TiO_2+ZnO dataset the number of features with the highest rank was around 160 features. When checking the features that were selected, it was observed that most of these were the located in the Section 4 of the NanoFingerprint. Further analysis with models mostly considering Section 4 could be beneficial in testing further models for toxicity prediction. Additionally, methods based on feature selection by means of ranking analysis can also provide a different insight on what part of the NanoFingerprint do provide further relevance to models. Due to time constraints, no further models could be analysed, however an extension of the models and focus on the utilization of algorithms based on Support Vector Machines would be an alternative to Linear Regression, which could at the same time be added to the website as a complement to the current model in place.

In regard to the website, future work should be focused on the visualisation of the shell as a mean to verify the results obtained from the NanoFingerprint. Multiple works are available using JavaScript and the integration of such functionality would not impact current resources available in the server.

7. References

[1] F. Serratosa, S. Álvarez, L. Escorihuela and M. Calatayud, “Subgraph NanoFingerprint for modelling metal oxide nanoparticles based on connected atoms exploration,” 2022.

[2] A. Nel, T. Xia, L. Mädler and N. Li, “Toxic potential of materials at the nanolevel,” *Science,* pp. 622-627, 2006.

[3] E. Papa, J. P. Doucet and A. Doucet-Panaye, “Linear and non-linear modelling of the cytotoxicity of TiO2 and ZnO nanoparticles by empirical descriptors,” *SAR and QSAR in Environmental Research,* vol. 26, pp. 647-665, 2015.

[4] N. Anantha and A. Palaniappan, “NanoTox: Developemnt of a Parsimonious In Silico Model for Toxicity Assessment of Metal-Oxide Nanoparticles Using PhysicoChemical Features,” *ACS Omega,* pp. 11729-11739, 2021.

[5] A. Gajewicz, N. Schaeublin, B. Rasulev, S. Hussain, D. Leszczynska, T. Puzyn and J. Leszczysnki, “Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies,” *Nanotoxicology,* pp. 313-325, 2015.

[6] T. Puzyn, B. Rasulev, A. Gajewicz, X. Hu, T. Dasari and A. Michalkova, “Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles.,” *Nat Nanotechnol,* vol. 6, 2011.

[7] K. Tämm, L. Sikk, J. Burk, R. Rallo, S. Pokhrel, L. Mädler, J. Scott-Fordsmand and P. Burk, “Parametrization of nanoparticles: development of full-particle nanodescriptors,” *Nanoscale,* 2016.

[8] F. Effendy and B. Adhilaksono, “Performance Comparison of Web Backend and Database: A Case Study of Node.JS, Golang and MySQL, Mongo DB,” *Recent Advances in Computer Science and Communications,* vol. 14, no. 6, pp. 1955-1961, 2021.

[9] I. Guyon and A. Elisseeff, “An Introduction of Variable and Feature Selection,” *Journal of Machine Learning Research,* vol. 3, pp. 1157-1182, 2003.

[10] N. Chirico and P. Gramatica, “Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient,” *Journal of Chemical Information and Modelling,* pp. 2320-2335, 2011.

[11] C. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[12] R. Duda, P. Hart and D. Stork, Pattern Classification, Second ed., John Wiley & Sons, Inc, 2001.

[13] G. Varoquaux, “Cross-validation failure: small sample sizes lead to large error bars,” 2017.

[14] “Golang Documentation,” [Online]. Available: https://go.dev/doc/. [Accessed 3 March 2023].

[15] “Shell Depth Calculator,” [Online]. Available: https://nanogen.me/shell-depth. [Accessed 7 April 2023].

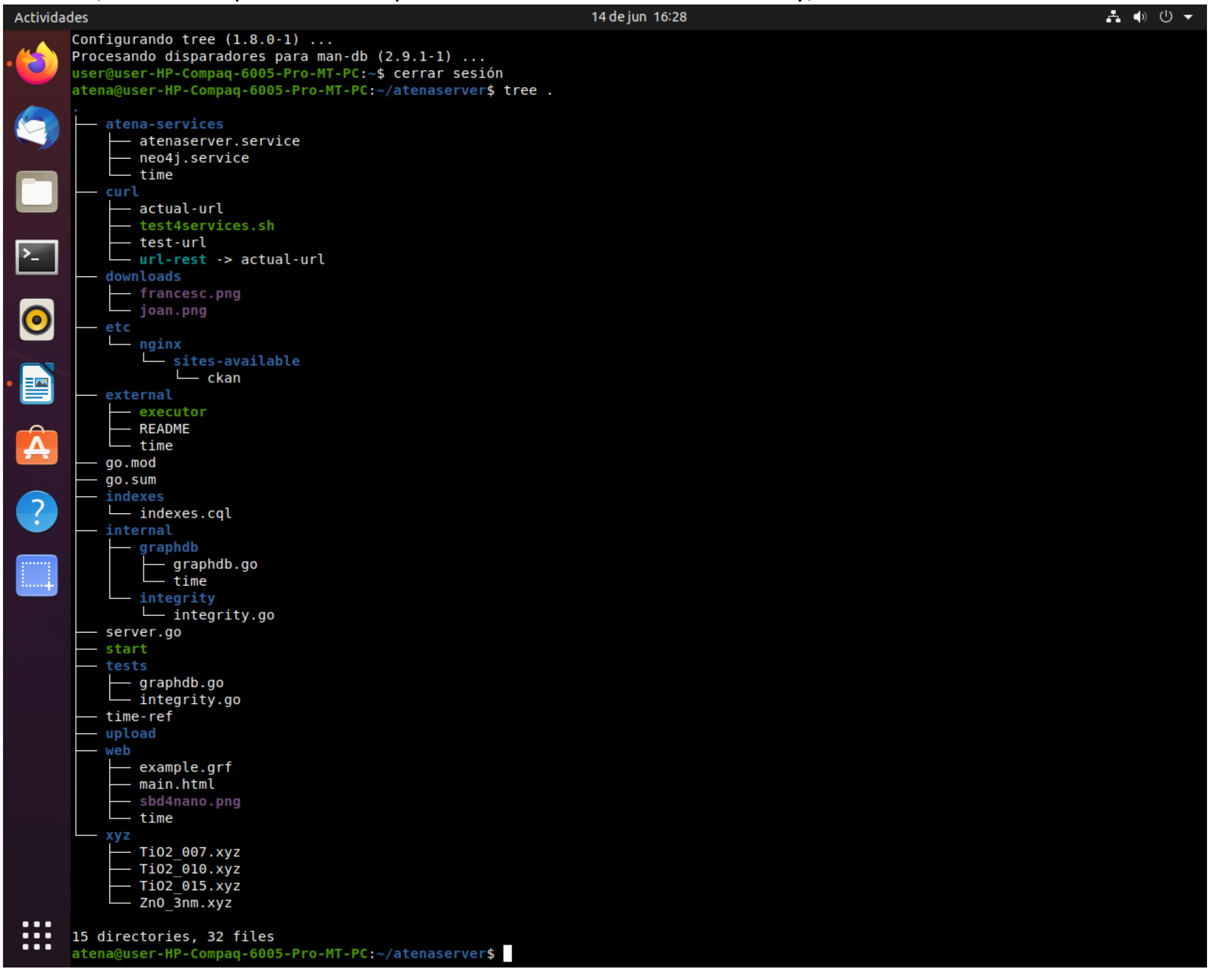
[16] I. Guyon, J. Weston, S. Barnhill and V. Vapnik, “Gene Selection for Cancer Classification using Support Vector Machines,” *Machine Learning,* vol. 46, pp. 389-422, 2002.

Appendix A

Access to Atena and Aura

Both Atena and Aura are Linux hosts located in the university labs. As stated in previous sections, Atena is the device in charge of hosting the website. As a security measure, any changes performed in Atena in terms of the website have to be performed through Aura.

In the scenario of requiring modifications in the workflow or presentation of data for the website, file are required to be uploaded in “atenaserver” directory, located in Aura.



Changes require of username and password, which can be provided by the local site admin providing a justification, as site is publicly accessible, and any unauthorized changes can lead to security violations.

Web Portal: GenVector function

```go
func GenVector(file string, shellThinkness float64, genXyz bool, maxBounds int, path string, strId string, base_name string) (fileRet string, fileRetNV string, err error, fileShell string) {

    defer func() {···
    }()
    shellThinkness = shellThinkness * 10
    var xyz []Atom
    content, err := os.ReadFile(file)
    if err != nil {···
    }
    lines := strings.Split(string(content), "\n")
    // check first line of .xyz file
    var head string
    if len(strings.Fields(lines[1])) > 3 {···
    } else {···
    }
    xyz = make([]Atom, len(lines))
    var median Atom

    for i, l := range lines {···
    }
    numXYZ := float64(len(xyz))
    median.X /= numXYZ
    median.Y /= numXYZ
    median.Z /= numXYZ
    maxRadius := maxRadiusDistInit(median, xyz)
    shell := make(map[int]*Atom)
    shellLine := maxRadius - shellThinkness
    if shellLine <= 0.0 {···
    }
    nearShellLine := shellLine - min_distance
    if nearShellLine <= 0.0 {···
    }
    length_shell := 0
    for i, _ := range xyz {···
    }

    visited := make(map[string]bool)
    bounds := make([]Bound, 0, 100) // bound format IDATOM-IDATOM. Ex: 22-66
    for id, atom := range shell {···
    }
    fmt.Println("Atoms ", len(shell))
    fmt.Println("bounds: ", len(bounds))

    NF_path := filepath.Join(path, strId)
    if err_dir := os.MkdirAll(NF_path, os.ModePerm); err_dir != nil {
        log.Fatal(err)
    }
    trimmed_basename := strings.TrimSuffix(base_name, filepath.Ext(base_name))
    fileZIP := genXYZ(shell, head, length_shell, NF_path, strId, trimmed_basename)
    fileToDownload, fileToDownloadNoVer, err := computeShell(shell, bounds, driver, shellThinkness, maxBounds, maxRadius*2, NF_path, strId, trimmed_basename)
    return fileToDownload, fileToDownloadNoVer, err, fileZIP
```

```go
func computeShell(shell map[int]*Atom, bounds []Bound, driver neo4j.Driver, shellThinkness float64, maxBounds int, size float64, vectorPath string, strId string, fileName string) (string, string, error) {

    fileOut := fileName + "_" + fmt.Sprintf("%v", shellThinkness) + "nm_" + fmt.Sprintf("%v", maxBounds) + "_Verbose_Nanofingerprint.txt"  // verbose fingerprint file - change from filename_base[2] from regex to filename
    fileOutNoVerbose := fileName + "_" + fmt.Sprintf("%v", shellThinkness) + "nm_" + fmt.Sprintf("%v", maxBounds) + "_Nanofingerprint.txt" // no verbose fingerprint file
    path_NF_Verbose := filepath.Join(vectorPath, fileOut)
    f, err := os.Create(path_NF_Verbose)
    path_NF := filepath.Join(vectorPath, fileOutNoVerbose)
    f2, err := os.Create(path_NF)

    check(err)
    defer f.Close()
    defer f2.Close()
    w := bufio.NewWriter(f)
    w2 := bufio.NewWriter(f2)

    sessionW := driver.NewSession(neo4j.SessionConfig{AccessMode: neo4j.AccessModeWrite})
    defer unsafeClose(sessionW)
    // loading the shell
    sessionW.WriteTransaction(func(tx neo4j.Transaction) (interface{}, error) {
        loadShell(tx, strId, fileName, shell, bounds) // updated to incldue fileName reference
        return nil, nil
    })
    // end loading the shell
    fmt.Fprintf(w, "1. Shell thickness in Angstrom:%f\n", shellThinkness)
    fmt.Fprintf(w, "2. Maximum number of bounds per atom:%d\n", maxBounds)
    fmt.Fprintf(w, "3. Size in Angstrom:%f\n", size)
    // no verbose fingerprint generation
    fmt.Fprintf(w2, "%f\n", shellThinkness)
    fmt.Fprintf(w2, "%d\n", maxBounds)
    fmt.Fprintf(w2, "%f\n", size)
    //metalDetected := ""
    // read only transaction to make the queries:
    session := driver.NewSession(neo4j.SessionConfig{AccessMode: neo4j.AccessModeRead})
    defer unsafeClose(session)
    // doing the readonly queries
    session.ReadTransaction(func(tx neo4j.Transaction) (interface{}, error) {···
    w.Flush()
    w2.Flush()
    sessionW.WriteTransaction(func(tx neo4j.Transaction) (interface{}, error) {
        deleteShell(tx, strId)
        return nil, nil
    })

    return filepath.Join(strId, fileOut), filepath.Join(strId, fileOutNoVerbose), nil

}
```

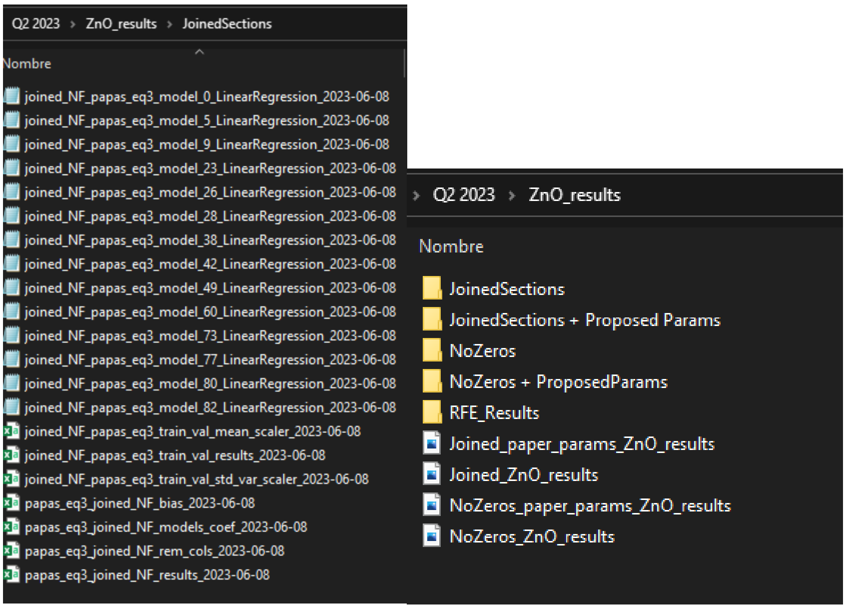
Some steps have been hidden to avoid oversharing contents of site backend – however, as seen in the snippets, GenVector function is in charge of performing the required operations, such as initializing files where the NanoFingerprint will be written to as well as computing what atoms are part of the shell, to later submit it and perform additional operations using Neo4j.

Code model analysis

Code used for performing model analysis has been added to GitHub – it can be found at: https://github.com/khuillca/NanoFingerprint-model-analysis.

Models Generated

A set of multiple models have been created for the experiments conducted in section 4. These models can be loaded on a python script to be analysed on further samples. These have been stored in 4 different directories, each of these have a set of results obtained from the models as well as the coefficients linked to each LinearRegression model generated.



All models and results have been shared with the director of the TFM, which will make them available to the public.

Appendix B

RFE Results

TiO2

	mean_fit_time	std_fit_time	mean_score_time	std_score_time	param_n_features_to_select	params	split0_test_r2	split1_test_r2	split2_test_r2	split3_test_r2	split4_test_r2	mean_test_r2	std_test_r2	rank_test_r2	split0_train_r2	split1_train_r2	split2_train_r2	split3_train_r2	split4_train_r2	mean_train_r2	std_train_r2
0	0.05125518	0.00050022	0.00300288	3.37E-07	1	{'n_features_to_select': 1}	-0.05548076	-0.16953452	-0.14336964	-0.85070966	-0.04107557	-0.25203403	0.30336885	35	0.04430053	0.08134835	0.00103552	0.13080274	0.01237219	0.05397187	0.0474948
1	0.05006728	4.16E-05	0.00280252	0.00040038	2	{'n_features_to_select': 2}	-1.00955567	0.31630052	0.0181279	-0.00414716	0.34696378	-0.06646213	0.49353668	34	0.47181351	0.41805398	0.54025596	0.4441168	0.3949922	0.45384649	0.05025736
2	0.04846296	0.00051433	0.00260234	0.00049037	3	{'n_features_to_select': 3}	0.30633734	0.67994614	-0.24243613	0.32548503	0.19532156	0.25293079	0.29636656	17	0.65161896	0.59856408	0.77617703	0.68476495	0.67378899	0.6769828	0.05779806
3	0.0470439	2.34E-06	0.00280204	0.00040181	4	{'n_features_to_select': 4}	0.37504934	0.67849598	-0.31550648	0.23638256	0.2343456	0.2417534	0.32222172	19	0.66662856	0.62632091	0.78329756	0.71223298	0.69560415	0.69681683	0.0521518
4	0.04527907	0.00035452	0.00280404	0.00040115	5	{'n_features_to_select': 5}	0.37504934	0.67849598	-0.31550648	0.23638256	0.15017233	0.22491874	0.32436072	27	0.66662856	0.62632091	0.78329756	0.71223298	0.7009167	0.69787934	0.05217038
5	0.04488792	0.00042985	0.00299463	1.78E-05	6	{'n_features_to_select': 6}	0.37504934	0.67849598	-0.31550648	0.19734089	0.15017233	0.21711041	0.32446067	33	0.66662856	0.62632091	0.78329756	0.72357721	0.7009167	0.70014819	0.05298558
6	0.04309664	0.00011475	0.00240235	0.00049014	7	{'n_features_to_select': 7}	0.37504934	0.67849598	-0.31550648	0.19734089	0.15017233	0.21711041	0.32446067	32	0.66662856	0.62632091	0.78329756	0.72357721	0.7009167	0.70014819	0.05298558
7	0.04140525	0.00051623	0.00251398	0.00044833	8	{'n_features_to_select': 8}	0.37504934	0.68551919	-0.31550648	0.19734089	0.15017233	0.21851505	0.32646406	29	0.66662856	0.63036488	0.78329756	0.72357721	0.7009167	0.70095698	0.05187164
8	0.04203839	0.00405338	0.00260234	0.00049017	9	{'n_features_to_select': 9}	0.37504934	0.68551919	-0.31550648	0.19734089	0.15017233	0.21851505	0.32646406	30	0.66662856	0.63036488	0.78329756	0.72357721	0.7009167	0.70095698	0.05187164
9	0.04444041	0.00742563	0.00280299	0.00040026	10	{'n_features_to_select': 10}	0.37504934	0.68551919	-0.31550648	0.19734089	0.15017233	0.21851505	0.32646406	28	0.66662856	0.63036488	0.78329756	0.72357721	0.7009167	0.70095698	0.05187164
10	0.03763442	0.00049051	0.0026022	0.00049045	11	{'n_features_to_select': 11}	0.37504934	0.68551919	-0.31550648	0.19734089	0.15017233	0.21851505	0.32646406	31	0.66662856	0.63036488	0.78329756	0.72357721	0.7009167	0.70095698	0.05187164
11	0.03643308	0.00049049	0.00260248	0.00049049	12	{'n_features_to_select': 12}	0.37504934	0.68551919	-0.24731295	0.19734089	0.15017233	0.23215376	0.30455889	26	0.66662856	0.63036488	0.81065227	0.72357721	0.7009167	0.70642792	0.06092087

1 2	0.0408 3691	0.003 8712 4	0.00320 287	0.0014 7105	13	{'n_features_ to_select': 13}	0.375 04934	0.685 51919	- 0.247 31295	0.197 34089	0.150 17233	0.2321 5376	0.304 5588 9	25	0.6666 2856	0.6303 6488	0.8106 5227	0.7235 7721	0.7009 167	0.7064 2792	0.060 9208 7
1 3	0.0336 3061	0.000 4904 7	0.00300 269	2.34E- 07	14	{'n_features_ to_select': 14}	0.375 04934	0.685 51919	- 0.231 84009	0.197 34089	0.150 17233	0.2352 4833	0.299 7114 1	21	0.6666 2856	0.6303 6488	0.8106 9877	0.7235 7721	0.7009 167	0.7064 3722	0.060 9367 8
1 4	0.0319 9759	0.000 6361 4	0.00260 248	0.0004 9	15	{'n_features_ to_select': 15}	0.375 04934	0.685 51919	- 0.231 84009	0.197 34089	0.150 17233	0.2352 4833	0.299 7114 1	20	0.6666 2856	0.6303 6488	0.8106 9877	0.7235 7721	0.7009 167	0.7064 3722	0.060 9367 8
1 5	0.0310 2822	0.000 6329 4	0.00200 191	3.57E- 07	16	{'n_features_ to_select': 16}	0.375 04934	0.685 51919	- 0.231 84009	0.197 34089	0.150 17233	0.2352 4833	0.299 7114 1	23	0.6666 2856	0.6303 6488	0.8106 9877	0.7235 7721	0.7009 167	0.7064 3722	0.060 9367 8
1 6	0.0298 2698	0.000 4002 3	0.00280 285	0.0004 0042	17	{'n_features_ to_select': 17}	0.375 04934	0.685 51919	- 0.231 84009	0.197 34089	0.150 17233	0.2352 4833	0.299 7114 1	24	0.6666 2856	0.6303 6488	0.8106 9877	0.7235 7721	0.7009 167	0.7064 3722	0.060 9367 8
1 7	0.0281 3549	0.000 6703 2	0.00260 229	0.0004 9023	18	{'n_features_ to_select': 18}	0.375 04934	0.685 51919	- 0.231 84009	0.197 34089	0.150 17233	0.2352 4833	0.299 7114 1	22	0.6666 2856	0.6303 6488	0.8106 9877	0.7235 7721	0.7009 167	0.7064 3722	0.060 9367 8
1 8	0.0268 2452	0.000 4005 7	0.00220 189	0.0004 005	19	{'n_features_ to_select': 19}	0.375 04934	0.685 51919	- 0.231 84009	0.238 14418	0.150 17233	0.2434 0899	0.299 1230 8	18	0.6666 2856	0.6303 6488	0.8106 9877	0.7615 6494	0.7009 167	0.7140 3477	0.064 8430 9
1 9	0.0255 3134	0.000 4479 5	0.00252 361	0.0004 4963	20	{'n_features_ to_select': 20}	0.472 62627	0.685 51919	- 0.231 84009	0.238 14418	0.150 17233	0.2629 2438	0.310 0581 3	9	0.6983 6737	0.6303 6488	0.8106 9877	0.7615 6494	0.7009 167	0.7203 8253	0.061 3510 8
2 0	0.0242 2209	0.000 4003 3	0.00280 247	0.0004 0047	21	{'n_features_ to_select': 21}	0.472 62627	0.714 04319	- 0.231 84009	0.238 14418	0.150 17233	0.2686 2918	0.317 9431 7	2	0.6983 6737	0.6585 3651	0.8106 9877	0.7615 6494	0.7009 167	0.7260 1686	0.053 6335 7
2 1	0.0231 5946	0.000 4471 1	0.00280 271	0.0004 0035	22	{'n_features_ to_select': 22}	0.472 62627	0.714 04319	- 0.231 84009	0.238 14418	0.150 17233	0.2686 2918	0.317 9431 7	4	0.6983 6737	0.6585 3651	0.8106 9877	0.7615 6494	0.7009 167	0.7260 1686	0.053 6335 7
2 2	0.0214 1962	0.000 4902 9	0.00280 228	0.0004 005	23	{'n_features_ to_select': 23}	0.472 62627	0.714 04319	- 0.231 84009	0.238 14418	0.150 17233	0.2686 2918	0.317 9431 7	3	0.6983 6737	0.6585 3651	0.8106 9877	0.7615 6494	0.7009 167	0.7260 1686	0.053 6335 7
2 3	0.0206 1868	0.000 4903 7	0.00240 221	0.0004 9035	24	{'n_features_ to_select': 24}	0.472 62627	0.714 04319	- 0.231 84009	0.238 14418	0.150 17233	0.2686 2918	0.317 9431 7	5	0.6983 6737	0.6585 3651	0.8106 9877	0.7615 6494	0.7009 167	0.7260 1686	0.053 6335 7
2 4	0.0188 2348	0.000 2372 3	0.00280 261	0.0004 0042	25	{'n_features_ to_select': 25}	0.472 62627	0.714 04319	- 0.231 84009	0.238 14418	0.150 17233	0.2686 2918	0.317 9431 7	7	0.6983 6737	0.6585 3651	0.8106 9877	0.7615 6494	0.7009 167	0.7260 1686	0.053 6335 7
2 5	0.0174 1571	0.000 4904 1	0.00280 261	0.0004 0031	26	{'n_features_ to_select': 26}	0.472 62627	0.714 04319	- 0.231 84009	0.238 14418	0.150 17233	0.2686 2918	0.317 9431 7	1	0.6983 6737	0.6585 3651	0.8106 9877	0.7615 6494	0.7009 167	0.7260 1686	0.053 6335 7
2 6	0.0165 9336	0.000 4705 5	0.00280 17	0.0003 9986	27	{'n_features_ to_select': 27}	0.472 62627	0.714 04319	- 0.231 84009	0.238 14418	0.150 17233	0.2686 2918	0.317 9431 7	6	0.6983 6737	0.6585 3651	0.8106 9877	0.7615 6494	0.7009 167	0.7260 1686	0.053 6335 7
2 7	0.0150 0392	2.28E -05	0.00297 78	4.84E- 05	28	{'n_features_ to_select': 28}	0.422 49463	0.714 04319	- 0.231 84009	0.238 14418	0.150 17233	0.2586 0285	0.312 0885 9	14	0.6986 7555	0.6585 3651	0.8106 9877	0.7615 6494	0.7009 167	0.7260 7849	0.053 6019 3

28	0.01360169	0.00051198	0.00260458	0.00049353	29	{'n_features_to_select': 29}	0.42249463	0.71404319	-0.23184009	0.23814418	0.15017233	0.25860285	0.31208859	10	0.69867555	0.65853651	0.81069877	0.76156494	0.7009167	0.72607849	0.05360193
29	0.01221304	0.00040018	0.00240111	0.00049144	30	{'n_features_to_select': 30}	0.42249463	0.71404319	-0.23184009	0.23814418	0.15017233	0.25860285	0.31208859	11	0.69867555	0.65853651	0.81069877	0.76156494	0.7009167	0.72607849	0.05360193
30	0.01041288	0.00048886	0.00260153	0.00048991	31	{'n_features_to_select': 31}	0.42249463	0.71404319	-0.23184009	0.23814418	0.15017233	0.25860285	0.31208859	12	0.69867555	0.65853651	0.81069877	0.76156494	0.7009167	0.72607849	0.05360193
31	0.00980907	0.00039962	0.00220099	0.00040083	32	{'n_features_to_select': 32}	0.42249463	0.71404319	-0.23184009	0.23814418	0.15017233	0.25860285	0.31208859	15	0.69867555	0.65853651	0.81069877	0.76156494	0.7009167	0.72607849	0.05360193
32	0.00800772	3.75E-06	0.00254841	0.00045866	33	{'n_features_to_select': 33}	0.42249463	0.71404319	-0.23184009	0.23814418	0.15017233	0.25860285	0.31208859	16	0.69867555	0.65853651	0.81069877	0.76156494	0.7009167	0.72607849	0.05360193
33	0.0066061	0.00049027	0.00260239	0.00049031	34	{'n_features_to_select': 34}	0.42249463	0.71404319	-0.23184009	0.23814418	0.15017233	0.25860285	0.31208859	13	0.69867555	0.65853651	0.81069877	0.76156494	0.7009167	0.72607849	0.05360193
34	0.00560513	0.00049029	0.0026022	0.00049025	35	{'n_features_to_select': 35}	0.42249463	0.71404319	-0.23184009	0.23814418	0.17624737	0.26381786	0.31044668	8	0.69867555	0.65853651	0.81069877	0.76156494	0.70094714	0.72608458	0.05359907

ZnO

	mean_fit_time	std_fit_time	mean_score_time	std_score_time	param_n_features_to_select	params	split0_test_r2	split1_test_r2	mean_test_r2	std_test_r2	rank_test_r2	split0_train_r2	split1_train_r2	mean_train_r2	std_train_r2
0	0.19766223	0.000857	0.0048902	0.00011349	1	{'n_features_to_select': 1}	-2.66891892	-28.4637097	-15.5663143	12.8973954	141	0.60483871	0.66216216	0.63350044	0.02866173
1	0.19502616	0.00015092	0.00500441	2.38E-07	2	{'n_features_to_select': 2}	-7.27117586	-9.34301241	-8.30709414	1.03591827	8	1	0.97838343	0.98919172	0.01080828
2	0.19424963	0.00170326	0.00491035	9.60E-05	3	{'n_features_to_select': 3}	-9.76684136	-10.733022	-10.2499317	0.48309032	19	1	1	1	0
3	0.19076061	0.00047946	0.00500596	1.19E-07	4	{'n_features_to_select': 4}	-9.4800264	-13.4409937	-11.4605101	1.98048367	134	1	1	1	0
4	0.18883145	0.0013417	0.00450444	0.00050044	5	{'n_features_to_select': 5}	-3.15946718	-17.2781005	-10.2187838	7.05931664	16	1	1	1	0
5	0.18786907	0.001302	0.00500512	2.38E-07	6	{'n_features_to_select': 6}	-3.15844905	-17.2893053	-10.2238772	7.06542814	17	1	1	1	0
6	0.18717015	1.19E-07	0.00500453	3.58E-07	7	{'n_features_to_select': 7}	-3.15664666	-17.2984046	-10.2275256	7.07087895	18	1	1	1	0
7	0.18466902	0.00050139	0.0045042	0.00050044	8	{'n_features_to_select': 8}	-3.15723848	-17.4613791	-10.3093088	7.1520703	20	1	1	1	0
8	0.18342841	0.00017321	0.00450456	0.00050056	9	{'n_features_to_select': 9}	-3.15779378	-17.4651761	-10.3114849	7.15369115	21	1	1	1	0
9	0.18216562	0.00100064	0.00500464	2.38E-07	10	{'n_features_to_select': 10}	-3.15932165	-17.4649087	-10.3121152	7.15279351	22	1	1	1	0
10	0.18166602	0.00050128	0.00500441	0	11	{'n_features_to_select': 11}	-3.15852632	-17.4671431	-10.3128347	7.15430838	23	1	1	1	0
11	0.18016422	1.19E-07	0.00500441	0	12	{'n_features_to_select': 12}	-3.15881715	-17.4689934	-10.3139053	7.15508813	24	1	1	1	0

12	0.17825747	9.51E-05	0.00500441	0	13	{'n_features_to_select': 13}	- 3.15914116	- 17.4701391	- 10.3146401	7.155498 97	25	1	1	1	0
13	0.17966342	0.0015015 6	0.00457895	0.00057542	14	{'n_features_to_select': 14}	- 3.15903948	- 17.4711107	- 10.3150751	7.156035 61	27	1	1	1	0
14	0.1774888	0.0005612 4	0.0045011	0.00050354	15	{'n_features_to_select': 15}	- 3.15893904	- 17.4720043	- 10.3154716	7.156532 6	32	1	1	1	0
15	0.17601478	0.0001455 6	0.0040009	2.86E-06	16	{'n_features_to_select': 16}	- 3.15765552	- 17.4728727	- 10.3152641	7.157608 58	29	1	1	1	0
16	0.17491818	1.60E-05	0.00500119	3.46E-06	17	{'n_features_to_select': 17}	- 3.15811812	- 17.4736911	- 10.3159046	7.157786 5	35	1	1	1	0
17	0.17399752	6.14E-05	0.00500464	2.38E-07	18	{'n_features_to_select': 18}	- 3.15739465	- 17.4743644	- 10.3158795	7.158484 89	34	1	1	1	0
18	0.17310178	0.0010610 8	0.00500417	0	19	{'n_features_to_select': 19}	- 3.15677561	- 17.4728111	- 10.3147934	7.158017 76	26	1	1	1	0
19	0.17175889	0.0003974 4	0.00500464	0	20	{'n_features_to_select': 20}	- 3.15704546	- 17.4731931	- 10.3151193	7.158073 81	28	1	1	1	0
20	0.1705091	0.0003540 5	0.00500441	2.38E-07	21	{'n_features_to_select': 21}	- 3.15697843	- 17.4735598	- 10.3152691	7.158290 68	30	1	1	1	0
21	0.16965449	0.0005008	0.00450432	0.00050056	22	{'n_features_to_select': 22}	- 3.15691159	- 17.4738827	- 10.3153972	7.158485 58	31	1	1	1	0
22	0.16674089	0.0004105 6	0.00400364	1.19E-07	23	{'n_features_to_select': 23}	- 3.15714594	- 17.4741941	-10.31567	7.158524 08	33	1	1	1	0
23	0.16499698	0.0011540 7	0.005005	1.19E-07	24	{'n_features_to_select': 24}	- 3.15738015	- 17.4744668	- 10.3159235	7.158543 31	36	1	1	1	0
24	0.16574442	0.0024083 9	0.00500464	2.38E-07	25	{'n_features_to_select': 25}	- 3.15745891	- 17.4746332	-10.316046	7.158587 13	37	1	1	1	0
25	0.16314852	3.58E-07	0.00500464	2.38E-07	26	{'n_features_to_select': 26}	- 3.15748681	- 17.4747946	- 10.3161407	7.158653 9	38	1	1	1	0
26	0.16164672	0.0005008	0.00500453	1.19E-07	27	{'n_features_to_select': 27}	- 4.44373391	- 17.4750155	- 10.9593747	6.515640 78	132	1	1	1	0
27	0.16157591	0.0014306 3	0.00450385	0.00050104	28	{'n_features_to_select': 28}	- 3.15794394	- 17.4752299	- 10.3165869	7.158642 99	39	1	1	1	0
28	0.15904415	0.0001007 3	0.00450409	0.00050056	29	{'n_features_to_select': 29}	- 3.15800895	- 17.4754964	- 10.3167527	7.158743 71	40	1	1	1	0
29	0.15864432	0.0005005 6	0.00500464	0	30	{'n_features_to_select': 30}	- 3.15826389	-17.475649	- 10.3169565	7.158692 56	41	1	1	1	0
30	0.1566428	0.0005003 2	0.00500441	2.38E-07	31	{'n_features_to_select': 31}	- 3.15844002	- 17.4757907	- 10.3171154	7.158675 34	42	1	1	1	0
31	0.15569174	0.0004500 2	0.00500429	3.58E-07	32	{'n_features_to_select': 32}	- 3.15853698	- 15.5896447	- 9.37409083	6.215553 85	13	1	1	1	0
32	0.15388906	0.0011373 8	0.00500536	7.15E-07	33	{'n_features_to_select': 33}	- 3.15846237	- 17.4765973	- 10.3175298	7.159067 46	43	1	1	1	0
33	0.15406799	7.25E-05	0.00500488	2.38E-07	34	{'n_features_to_select': 34}	-3.1585459	- 17.4766954	- 10.3176207	7.159074 75	44	1	1	1	0
34	0.15113771	1.19E-07	0.00500453	1.19E-07	35	{'n_features_to_select': 35}	-3.1586294	- 17.4767804	- 10.3177049	7.159075 48	45	1	1	1	0
35	0.15043879	0.0003018 4	0.00500405	3.58E-07	36	{'n_features_to_select': 36}	- 3.15867004	- 16.1838077	- 9.67123885	6.512568 81	14	1	1	1	0
36	0.14915204	1.12E-05	0.00500107	3.34E-06	37	{'n_features_to_select': 37}	- 3.15871068	- 17.4769844	- 10.3178476	7.159136 87	46	1	1	1	0
37	0.14763439	0.0005005 6	0.00500488	2.38E-07	38	{'n_features_to_select': 38}	- 3.15870757	- 17.4770675	- 10.3178875	7.159179 96	47	1	1	1	0
38	0.14613318	0	0.00500417	2.38E-07	39	{'n_features_to_select': 39}	- 3.15869571	- 17.4772466	- 10.3179712	7.159275 46	48	1	1	1	0
39	0.14588416	0.0007520 9	0.00450432	0.00050032	40	{'n_features_to_select': 40}	- 3.15874129	- 16.4932189	- 9.82598008	6.667238 8	15	1	1	1	0

40	0.14373887	0.00039256	0.00500405	1.19E-07	41	{'n_features_to_select': 41}	-3.15880085	-17.4774133	-10.3181071	7.15930623	49	1	1	1	0
41	0.14313018	3.58E-07	0.0045042	0.0005002	42	{'n_features_to_select': 42}	-3.15886042	-17.4774811	-10.3181708	7.15931036	50	1	1	1	0
42	0.14162874	0.00050068	0.0045042	0.00050044	43	{'n_features_to_select': 43}	-3.158864	-17.4775512	-10.3182076	7.15934362	51	1	1	1	0
43	0.13948095	0.00064623	0.00500429	1.19E-07	44	{'n_features_to_select': 44}	-3.1588396	-17.4776206	-10.3182301	7.15939051	52	1	1	1	0
44	0.13880014	0.00032663	0.00400353	2.38E-07	45	{'n_features_to_select': 45}	-3.15879987	-15.028464	-9.09363196	5.93483209	12	1	1	1	0
45	0.13712525	2.38E-07	0.00450361	0.00050056	46	{'n_features_to_select': 46}	-3.15876016	-17.4777663	-10.3182632	7.15950307	54	1	1	1	0
46	0.13562334	0.00050056	0.00500453	1.19E-07	47	{'n_features_to_select': 47}	-3.35469278	-17.4778249	-10.4162589	7.06156607	116	1	1	1	0
47	0.13414645	0.00097704	0.00500417	0	48	{'n_features_to_select': 48}	-4.21798357	-17.4778713	-10.8479274	6.62994385	131	1	1	1	0
48	0.13362169	0.00050068	0.00500441	0	49	{'n_features_to_select': 49}	-3.15858541	-17.4779173	-10.3182513	7.15966592	53	1	1	1	0
49	0.13346493	0.00165784	0.00450385	0.00050056	50	{'n_features_to_select': 50}	-3.15629951	-17.5747084	-10.365504	7.20920444	60	1	1	1	0
50	0.13111937	1.19E-07	0.00500441	0	51	{'n_features_to_select': 51}	-3.15859645	-17.5748537	-10.3667251	7.2081286	64	1	1	1	0
51	0.12886906	0.00024796	0.00450492	0.00050139	52	{'n_features_to_select': 52}	-3.1586261	-17.5748041	-10.3667151	7.20808901	63	1	1	1	0
52	0.12861896	0.0005002	0.00500441	2.38E-07	53	{'n_features_to_select': 53}	-3.15865953	-17.5747553	-10.3667074	7.20804789	62	1	1	1	0
53	0.12635326	0.00046325	0.00500453	3.58E-07	54	{'n_features_to_select': 54}	-3.15865632	-17.5748659	-10.3667611	7.20810477	65	1	1	1	0
54	0.12518382	5.46E-05	0.00500894	4.05E-06	55	{'n_features_to_select': 55}	-3.11012088	-17.5750931	-10.342607	7.23248612	55	1	1	1	0
55	0.12463009	0.00050223	0.00451648	0.00051248	56	{'n_features_to_select': 56}	-3.15867084	-17.575318	-10.3669944	7.20832357	66	1	1	1	0
56	0.12330914	0.00080323	0.00450409	0.00050008	57	{'n_features_to_select': 57}	-3.13373873	-17.5754486	-10.3545937	7.22085493	57	1	1	1	0
57	0.12211096	0.001001	0.0045042	0.00050092	58	{'n_features_to_select': 58}	-3.15865252	-17.5755482	-10.3671004	7.20844783	67	1	1	1	0
58	0.12060976	0.00050044	0.00500453	1.19E-07	59	{'n_features_to_select': 59}	-3.15868593	-17.5756135	-10.3671497	7.20846378	68	1	1	1	0
59	0.1196568	0.00045228	0.00400376	2.38E-07	60	{'n_features_to_select': 60}	-3.1586618	-17.5756786	-10.3671702	7.20850841	69	1	1	1	0
60	0.11810768	0.001001	0.00500464	0	61	{'n_features_to_select': 61}	-3.15938261	-17.5757436	-10.3675631	7.20818048	84	1	1	1	0
61	0.1181078	4.77E-07	0.00400329	4.77E-07	62	{'n_features_to_select': 62}	-3.15863448	-17.5758083	-10.3672214	7.20858693	70	1	1	1	0
62	0.11560464	0.00050044	0.00450456	0.00050056	63	{'n_features_to_select': 63}	-3.16212577	-17.5758729	-10.3689993	7.20687357	104	1	1	1	0
63	0.11410368	3.58E-07	0.00450397	0.0005002	64	{'n_features_to_select': 64}	-3.15865719	-17.5759164	-10.3672868	7.20862959	71	1	1	1	0
64	0.11410368	0.001001	0.00600541	0.00200188	65	{'n_features_to_select': 65}	-3.15865809	-17.5759597	-10.3673089	7.20865082	72	1	1	1	0
65	0.11316431	0.00106251	0.00500453	1.19E-07	66	{'n_features_to_select': 66}	-3.15865202	-17.5761729	-10.3674125	7.20876045	75	1	1	1	0
66	0.11109996	1.19E-06	0.00500476	3.58E-07	67	{'n_features_to_select': 67}	-3.15864595	-17.5763794	-10.3675127	7.20886675	78	1	1	1	0
67	0.10909951	5.96E-07	0.00500453	3.58E-07	68	{'n_features_to_select': 68}	-3.15863988	-17.5764056	-10.3675227	7.20888286	79	1	1	1	0

68	0.10909891	0.00100088	0.00500453	1.19E-07	69	{'n_features_to_select': 69}	-3.15864078	-17.5764317	-10.3675363	7.20889547	81	1	1	1	0
69	0.10795808	0.00014067	0.00500464	4.77E-07	70	{'n_features_to_select': 70}	-3.15863471	-17.5764578	-10.3675463	7.20891155	82	1	1	1	0
70	0.10656786	0.00052214	0.00450122	0.0004977	71	{'n_features_to_select': 71}	-3.15862865	-17.5764839	-10.3675563	7.20892761	83	1	1	1	0
71	0.10606623	0.00203204	0.00600517	0.001001	72	{'n_features_to_select': 72}	-3.20916012	-17.5765099	-10.392835	7.18367488	112	1	1	1	0
72	0.10359478	0.00050068	0.00500464	4.77E-07	73	{'n_features_to_select': 73}	-3.15861651	-17.5765359	-10.3675762	7.20895968	85	1	1	1	0
73	0.10239387	0.00020361	0.00500476	1.19E-07	74	{'n_features_to_select': 74}	-3.15861045	-17.5765618	-10.3675861	7.20897568	86	1	1	1	0
74	0.10036743	0.00025857	0.00500786	3.22E-06	75	{'n_features_to_select': 75}	-3.15859045	-17.5765877	-10.3675891	7.20899864	87	1	1	1	0
75	0.09959054	0.00050044	0.00450409	0.00050056	76	{'n_features_to_select': 76}	-3.15859226	-17.5766136	-10.3676029	7.20901067	88	1	1	1	0
76	0.0978024	0.0001924	0.00550079	0.00049663	77	{'n_features_to_select': 77}	-3.20759116	-17.5766394	-10.3921153	7.18452414	111	1	1	1	0
77	0.10188663	0.00242078	0.00500739	2.98E-06	78	{'n_features_to_select': 78}	-3.15280694	-17.5766652	-10.3647361	7.21192915	59	1	1	1	0
78	0.09708834	2.38E-07	0.00450373	0.0005002	79	{'n_features_to_select': 79}	-3.1585977	-17.576691	-10.3676444	7.20904665	95	1	1	1	0
79	0.09512913	4.23E-05	0.00450492	0.0005002	80	{'n_features_to_select': 80}	-3.35333136	-17.5767167	-10.4650241	7.11169269	129	1	1	1	0
80	0.09658778	0.00350344	0.00500464	4.77E-07	81	{'n_features_to_select': 81}	-3.15859626	-17.5767425	-10.3676694	7.20907309	101	1	1	1	0
81	0.09208453	0.00100195	0.00450444	0.00050044	82	{'n_features_to_select': 82}	-3.12357114	-17.5767681	-10.3501696	7.22659849	56	1	1	1	0
82	0.09058261	0.0005002	0.00450409	0.00050056	83	{'n_features_to_select': 83}	-3.15861412	-27.4794034	-15.3190088	12.1603946	140	1	1	1	0
83	0.09018445	0.000103	0.00500441	2.38E-07	84	{'n_features_to_select': 84}	-3.15860912	-17.5760948	-10.367352	7.20874283	73	1	1	1	0
84	0.08908093	1.19E-07	0.00400388	1.19E-07	85	{'n_features_to_select': 85}	-3.16653925	-17.5761234	-10.3713313	7.20479206	105	1	1	1	0
85	0.08583427	0.00024438	0.00500441	2.38E-07	86	{'n_features_to_select': 86}	-3.15780895	-54.3419533	-28.7498811	25.5920722	144	1	1	1	0
86	0.08432865	0.00025177	0.00500476	1.19E-07	87	{'n_features_to_select': 87}	-3.15862908	-17.5761531	-10.3673911	7.20876199	74	1	1	1	0
87	0.08257568	0.00050104	0.00500476	5.96E-07	88	{'n_features_to_select': 88}	-3.16854386	-17.5761815	-10.3723627	7.20381881	106	1	1	1	0
88	0.08257496	0.00150144	0.00500476	1.19E-07	89	{'n_features_to_select': 89}	-3.15870898	-17.5762118	-10.3674604	7.20875143	76	1	1	1	0
89	0.08067513	0.0003984	0.00500453	3.58E-07	90	{'n_features_to_select': 90}	-3.21849001	-17.5762422	-10.3973661	7.17887607	113	1	1	1	0
90	0.07957232	0.00050008	0.00500441	4.77E-07	91	{'n_features_to_select': 91}	-3.15870421	-17.5762724	-10.3674883	7.20878411	77	1	1	1	0
91	0.07807088	2.38E-07	0.00500488	2.38E-07	92	{'n_features_to_select': 92}	-3.14651688	-17.5763027	-10.3614098	7.21489289	58	1	1	1	0
92	0.07707071	4.77E-07	0.00500464	0	93	{'n_features_to_select': 93}	-3.15871857	-17.5763329	-10.3675257	7.20880716	80	1	1	1	0
93	0.075822	0.00024688	0.00450456	0.0005008	94	{'n_features_to_select': 94}	-3.15872053	-89.9374354	-46.548078	43.3893574	145	1	1	1	0
94	0.07545686	0.00038958	0.0045042	0.00050092	95	{'n_features_to_select': 95}	-3.15871554	-17.5765414	-10.3676285	7.20891294	91	1	1	1	0
95	0.07357335	0.00049377	0.00450385	0.00050032	96	{'n_features_to_select': 96}	-3.15871055	-17.5765713	-10.3676409	7.2089304	94	1	1	1	0

96	0.07211697	1.07E-06	0.00501561	1.10E-05	97	{'n_features_to_select': 97}	-3.15871251	-17.5766012	-10.3676569	7.20894436	98	1	1	1	0
97	0.07069337	0.00037134	0.00450444	0.00050044	98	{'n_features_to_select': 98}	-3.15871099	-17.5766009	-10.367656	7.20894497	97	1	1	1	0
98	0.06856227	0.00050044	0.00400341	1.19E-07	99	{'n_features_to_select': 99}	-3.15868975	-17.5766056	-10.3676477	7.20895794	96	1	1	1	0
99	0.06706083	3.58E-07	0.00500464	0	100	{'n_features_to_select': 100}	-3.47556876	17.5766103	-10.5260895	7.05052078	130	1	1	1	0
100	0.06656075	0.00050092	0.00500429	1.19E-07	101	{'n_features_to_select': 101}	-3.15870803	-17.576615	-10.3676615	7.20895348	99	1	1	1	0
101	0.06405807	2.38E-07	0.00450409	0.00050032	102	{'n_features_to_select': 102}	-3.15870652	-17.5766197	-10.3676631	7.20895658	100	1	1	1	0
102	0.06355739	0.00050116	0.00450397	0.00050092	103	{'n_features_to_select': 103}	-3.22983064	-17.576626	-10.4032283	7.17339769	114	1	1	1	0
103	0.06265521	0.00040722	0.00450385	0.00050032	104	{'n_features_to_select': 104}	-3.20290151	-17.5766324	-10.3897669	7.18686543	110	1	1	1	0
104	0.06111801	0.00012791	0.00492454	8.03E-05	105	{'n_features_to_select': 105}	-3.15859842	-17.5766387	-10.3676186	7.20902015	89	1	1	1	0
105	0.0595628	0.00049174	0.00500417	0	106	{'n_features_to_select': 106}	-3.16840004	-17.5766462	-10.3725231	7.20412306	107	1	1	1	0
106	0.05783129	0.00022125	0.00500453	1.19E-07	107	{'n_features_to_select': 107}	-3.15860234	-17.5766536	-10.367628	7.20902564	90	1	1	1	0
107	0.05655122	0.00050068	0.00450397	0.0005002	108	{'n_features_to_select': 108}	-3.15859909	-17.5766611	-10.3676301	7.20903099	92	1	1	1	0
108	0.05504978	3.58E-07	0.0045048	0.00050032	109	{'n_features_to_select': 109}	-3.15860105	-17.5766685	-10.3676348	7.20903373	93	1	1	1	0
109	0.05404878	0	0.00450444	0.00050068	110	{'n_features_to_select': 110}	-3.23751755	-17.576676	-10.4070968	7.16957921	115	1	1	1	0
110	0.05354834	0.00050044	0.0045042	0.0005002	111	{'n_features_to_select': 111}	-3.19504002	-17.5766834	-10.3858617	7.19082169	109	1	1	1	0
111	0.05154681	0.00050044	0.00500476	3.58E-07	112	{'n_features_to_select': 112}	-3.41419602	-4.14071798	-3.777457	0.36326098	2	1	1	1	0
112	0.05054641	0.00050068	0.00450373	0.00050044	113	{'n_features_to_select': 113}	-3.16064751	-17.5766824	-10.368665	7.20801744	103	1	1	1	0
113	0.04904425	1.19E-07	0.00500488	2.38E-07	114	{'n_features_to_select': 114}	-3.15878699	-6.75235772	-4.95557235	1.79678537	4	1	1	1	0
114	0.04804337	1.19E-07	0.00400353	0	115	{'n_features_to_select': 115}	-3.14801857	-6.54704044	-4.84752951	1.69951094	3	1	1	1	0
115	0.04609299	5.08E-05	0.00500464	2.38E-07	116	{'n_features_to_select': 116}	-3.15883947	-4.05312504	-3.60598226	0.44714279	1	1	1	1	0
116	0.044541	0.0005008	0.0045042	0.00050044	117	{'n_features_to_select': 117}	-3.15883736	-17.5773444	-10.3680909	7.20925351	102	1	1	1	0
117	0.04362869	0.00041103	0.00500464	2.38E-07	118	{'n_features_to_select': 118}	-3.15883784	-19.6643692	-11.4116035	8.2527657	133	1	1	1	0
118	0.04276729	0.0002718	0.00450397	0.00050044	119	{'n_features_to_select': 119}	-3.15458744	-17.5773521	-10.3659698	7.21138234	61	1	1	1	0
119	0.0415374	0.00050008	0.00450432	0.00050032	120	{'n_features_to_select': 120}	-3.21811757	-8.61398332	-5.91605045	2.69793288	6	1	1	1	0
120	0.04003632	1.19E-07	0.00500464	2.38E-07	121	{'n_features_to_select': 121}	-3.81686589	-10.44692	-7.13189296	3.31502707	7	1	1	1	0
121	0.03903639	8.34E-07	0.004004	2.38E-07	122	{'n_features_to_select': 122}	-3.15872921	-30.2687977	-16.7137635	13.5550342	142	1	1	1	0
122	0.03753436	0.00050008	0.00450361	0.00050008	123	{'n_features_to_select': 123}	-3.18738651	-13.7687373	-8.47806192	5.29067541	9	1	1	1	0
123	0.03553224	0.00050044	0.00500488	2.38E-07	124	{'n_features_to_select': 124}	-3.15733737	-23.4433619	-13.3003497	10.1430123	137	1	1	1	0

124	0.03503192	3.58E-07	0.00400388	1.19E-07	125	{'n_features_to_select': 125}	-3.15686307	-6.8288761	-4.99286959	1.83600652	5	1	1	1	0
125	0.03353035	0.00050008	0.00500488	7.15E-07	126	{'n_features_to_select': 126}	-3.17000268	-17.5775584	-10.3737806	7.20377788	108	1	1	1	0
126	0.03264725	0.00038302	0.00500453	1.19E-07	127	{'n_features_to_select': 127}	-3.30785763	-31.6405961	-17.4742269	14.1663692	143	1	1	1	0
127	0.03102863	3.58E-07	0.00500381	1.19E-07	128	{'n_features_to_select': 128}	-3.30785763	-25.1475567	-14.2277071	10.9198495	139	1	1	1	0
128	0.02952707	0.00050032	0.00500453	1.19E-07	129	{'n_features_to_select': 129}	-3.31388828	-23.1799841	-13.2469362	9.93304791	136	1	1	1	0
129	0.02852595	0.00050032	0.00450409	0.00050008	130	{'n_features_to_select': 130}	-3.31388828	-17.5775622	-10.4457252	7.13183696	127	1	1	1	0
130	0.02702475	7.15E-07	0.00500453	3.58E-07	131	{'n_features_to_select': 131}	-3.26547019	-17.5775643	-10.4215172	7.15604704	117	1	1	1	0
131	0.02602398	5.96E-07	0.00500476	1.19E-07	132	{'n_features_to_select': 132}	-3.30785763	-17.5775686	-10.4427131	7.13485547	118	1	1	1	0
132	0.02552307	0.00050032	0.00450409	0.00050056	133	{'n_features_to_select': 133}	-3.30785763	-17.5775729	-10.4427153	7.13485762	119	1	1	1	0
133	0.02315855	0.00013757	0.00483358	0.00017202	134	{'n_features_to_select': 134}	-3.31388828	-17.5775747	-10.4457315	7.13184318	128	1	1	1	0
134	0.02201998	1.19E-07	0.00450444	0.00050044	135	{'n_features_to_select': 135}	-3.31388828	-24.9625091	-14.1381987	10.8243104	138	1	1	1	0
135	0.02077365	0.00024962	0.00450432	0.00050008	136	{'n_features_to_select': 136}	-3.30785763	-17.5775782	-10.4427179	7.13486028	120	1	1	1	0
136	0.01902127	9.54E-07	0.00500929	3.22E-06	137	{'n_features_to_select': 137}	-3.30785763	-17.57758	-10.4427188	7.13486117	121	1	1	1	0
137	0.01757562	0.00058448	0.00500488	0	138	{'n_features_to_select': 138}	-3.30785763	-93.3725299	-48.3401938	45.0323361	146	1	1	1	0
138	0.01657283	0.00055659	0.00500906	3.22E-06	139	{'n_features_to_select': 139}	-3.30785763	-17.5775816	-10.4427196	7.134862	122	1	1	1	0
139	0.01501894	3.34E-06	0.00497282	2.56E-05	140	{'n_features_to_select': 140}	-3.30785763	-20.0928938	-11.7003757	8.39251806	135	1	1	1	0
140	0.01392162	0.00091445	0.00450122	0.00050342	141	{'n_features_to_select': 141}	-3.30785763	-17.5775845	-10.4427211	7.13486345	126	1	1	1	0
141	0.01351297	0.00050128	0.00500488	0.00100184	142	{'n_features_to_select': 142}	-3.30785763	-17.5775843	-10.4427209	7.13486331	125	1	1	1	0
142	0.01145017	0.00043881	0.00450397	0.0005002	143	{'n_features_to_select': 143}	-3.31388828	-14.1364557	-8.72517199	5.41128371	11	1	1	1	0
143	0.01050055	0.00049007	0.00400412	1.19E-07	144	{'n_features_to_select': 144}	-3.30785763	-17.5775841	-10.4427208	7.13486322	124	1	1	1	0
144	0.00900948	1.19E-07	0.00400376	2.38E-07	145	{'n_features_to_select': 145}	-3.30785763	-17.577584	-10.4427208	7.1348632	123	1	1	1	0
145	0.00750768	0.00050128	0.00450575	0.00050008	146	{'n_features_to_select': 146}	-3.30785763	-13.6710506	-8.48945412	5.18159649	10	1	1	1	0

TiO2+ZnO

	mean_fit_time	std_fit_time	mean_score_time	std_score_time	param_n_features_to_select	params	split0_test_r2	split1_test_r2	split2_test_r2	split3_test_r2	split4_test_r2	mean_test_r2	std_test_r2	rank_test_r2	split0_train_r2	split1_train_r2	split2_train_r2	split3_train_r2	split4_train_r2	mean_train_r2	std_train_r2
0	0.243890572	0.002120067	0.005316353	0.000406627	1	{'n_features_to_select': 1}	0.002194267	0.063868588	0.20572061	-0.362434142	-3.16517589	-0.651165313	1.270906808	135	0.092149804	0.088686557	0.004229603	0.00205504	0.186721209	0.074768443	0.068252787
1	0.241086054	0.001531413	0.005805063	0.000400209	2	{'n_features_to_select': 2}	0.00325219	0.150594031	0.215379221	-0.161319306	-3.121739333	-0.58276664	1.27610916	122	0.100250945	0.096860741	0.004902291	0.174885014	0.19484913	0.114349624	0.067272664
2	0.238798475	0.002143931	0.005404806	0.000490388	3	{'n_features_to_select': 3}	0.00325219	0.117428132	-0.861141059	-0.124891825	-3.440265664	-0.861123645	1.334070464	165	0.100250945	0.13196121	0.030816936	0.195455945	0.214788456	0.134654698	0.06647009
3	0.23697052	0.001380391	0.005004692	3.81E-07	4	{'n_features_to_select': 4}	0.025275114	0.14858371	-0.861141059	-0.125595399	-5.088711067	-1.18031774	1.985411018	166	0.172055916	0.132721231	0.030816936	0.195457474	0.305568196	0.16732395	0.089170479
4	0.235214329	0.001899039	0.005204535	0.000400019	5	{'n_features_to_select': 5}	0.443501962	0.14858371	-2.384375948	-0.126406496	-5.088711067	-1.401481568	2.097342403	168	0.491179623	0.132721231	0.279077486	0.19548231	0.305568196	0.280805769	0.121723002
5	0.234183073	0.001426241	0.005204391	0.000400448	6	{'n_features_to_select': 6}	0.445816329	0.14858371	-2.139513878	-0.126406496	-2.207851111	-0.775874289	1.155770697	163	0.493828241	0.132721231	0.293499863	0.19548231	0.534618149	0.330029959	0.15940794
6	0.241946554	0.015087192	0.005404902	0.000490408	7	{'n_features_to_select': 7}	0.445816329	0.14858371	-2.139513878	-0.126406496	-1.275561018	-0.589416271	0.969982037	124	0.493828241	0.132721231	0.293499863	0.19548231	0.701543767	0.363415082	0.208637392
7	0.23091507	0.001553096	0.005102015	0.000490278	8	{'n_features_to_select': 8}	0.817638667	0.14858371	-2.139513878	-0.126406496	-1.275561018	-0.515051803	1.056860677	11	0.572547524	0.132721231	0.293499863	0.19548231	0.701543767	0.379158939	0.220516325
8	0.230392122	0.001487383	0.005600405	0.000486385	9	{'n_features_to_select': 9}	0.817638667	0.14858371	-2.139513878	-0.126406496	-1.956143271	-0.651168254	1.182413157	136	0.572547524	0.132721231	0.293499863	0.19548231	0.819045	0.402659186	0.256882739
9	0.229604912	0.00175004	0.005398846	0.000483647	10	{'n_features_to_select': 10}	0.801631883	0.14858371	-2.139513878	-0.126406496	-1.956143271	-0.65436961	1.178447079	139	0.665542308	0.132721231	0.293499863	0.19548231	0.819045	0.421258142	0.271462554
10	0.228408813	0.002155364	0.005197811	0.000399025	11	{'n_features_to_select': 11}	0.801631883	0.14858371	-2.139513878	-0.126406496	-1.956143271	-0.65436961	1.178447079	141	0.665542308	0.132721231	0.293499863	0.19548231	0.819045	0.421258142	0.271462554
11	0.227801418	0.001740942	0.005186224	0.000412283	12	{'n_features_to_select': 12}	0.801631883	0.14858371	-2.139513878	-0.126406496	-1.956143271	-0.65436961	1.178447079	140	0.665542308	0.132721231	0.293499863	0.19548231	0.819045	0.421258142	0.271462554
12	0.22440877	0.001863506	0.005200672	0.000402366	13	{'n_features_to_select': 13}	0.801631883	0.391510774	-2.139513878	-0.126406496	-1.956143271	-0.605784198	1.21499108	128	0.665542308	0.634911647	0.293499863	0.19548231	0.819045	0.521696226	0.23681842
13	0.223620987	0.00180867	0.005404854	0.000490155	14	{'n_features_to_select': 14}	0.801631883	0.391510774	-2.139513878	-0.126406496	-1.956143271	-0.605784198	1.21499108	127	0.665542308	0.634911647	0.293499863	0.19548231	0.819045	0.521696226	0.23681842
14	0.22234025	0.00186502	0.00580554	0.000399971	15	{'n_features_to_select': 15}	0.801631883	0.391510774	-2.139513878	-0.126406496	-1.956143271	-0.605784198	1.21499108	126	0.665542308	0.634911647	0.293499863	0.19548231	0.819045	0.521696226	0.23681842
15	0.221669674	0.002689131	0.005404758	0.000490524	16	{'n_features_to_select': 16}	0.801631883	0.393018825	-2.139513878	-0.126406496	-1.956143271	-0.605482587	1.215238774	125	0.665542308	0.634933769	0.293499863	0.19548231	0.819045	0.52170065	0.236820535
16	0.219265842	0.001595432	0.005605984	0.000490214	17	{'n_features_to_select': 17}	0.801631883	0.562081732	-2.139513878	-0.126406496	-1.956143271	-0.571670006	1.244549019	80	0.665542308	0.758719199	0.293499863	0.19548231	0.819045	0.546457736	0.253263229

17	0.218876171	0.001072636	0.005204725	0.00040052	18	{'n_features_to_select': 18}	0.801631883	0.562081732	-2.139513878	-0.126406496	-1.956143271	-0.571670006	1.244549019	81	0.665542308	0.758719199	0.293499863	0.19548231	0.819045	0.546457736	0.253263229
18	0.218009281	0.000672077	0.005605173	0.000490719	19	{'n_features_to_select': 19}	0.799820689	0.562081732	-2.139513878	-0.126406496	-1.956143271	-0.572032245	1.244149452	83	0.681388302	0.758719199	0.293499863	0.19548231	0.819045	0.549626935	0.254827871
19	0.215817308	0.001752659	0.005004692	6.47E-07	20	{'n_features_to_select': 20}	0.799820689	0.562081732	-2.139513878	-0.126406496	-1.956143271	-0.572032245	1.244149452	82	0.681388302	0.758719199	0.293499863	0.19548231	0.819045	0.549626935	0.254827871
20	0.214087582	0.001164034	0.00538168	0.00047272	21	{'n_features_to_select': 21}	0.799820689	0.562081732	-2.139513878	-0.126406496	-1.956143271	-0.572032245	1.244149452	84	0.681388302	0.758719199	0.293499863	0.19548231	0.819045	0.549626935	0.254827871
21	0.212900305	0.000899589	0.005186892	0.00041046	22	{'n_features_to_select': 22}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	92	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
22	0.21222949	0.001632762	0.005204868	0.000400091	23	{'n_features_to_select': 23}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	117	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
23	0.21022706	0.001772008	0.005204391	0.000400329	24	{'n_features_to_select': 24}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	94	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
24	0.208927679	0.002394664	0.005204868	0.000399971	25	{'n_features_to_select': 25}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	98	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
25	0.207389021	0.00116685	0.005004025	4.91E-07	26	{'n_features_to_select': 26}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	113	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
26	0.207226181	0.000860619	0.005404758	0.000490135	27	{'n_features_to_select': 27}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	105	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
27	0.205826807	0.000806444	0.005203009	0.000396014	28	{'n_features_to_select': 28}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	100	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
28	0.203443623	0.00184807	0.005352879	0.000542042	29	{'n_features_to_select': 29}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	108	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
29	0.203114319	0.0013214	0.005204344	0.000400472	30	{'n_features_to_select': 30}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	106	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
30	0.200273561	0.001219602	0.005397177	0.000492862	31	{'n_features_to_select': 31}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	103	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
31	0.199601221	0.001630645	0.005605125	0.000490193	32	{'n_features_to_select': 32}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	106	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
32	0.198125553	0.001401243	0.005204725	0.00040052	33	{'n_features_to_select': 33}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	104	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
33	0.196611691	0.001397521	0.005204582	0.000400233	34	{'n_features_to_select': 34}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	97	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
34	0.195796013	0.001350354	0.005204725	0.0004004	35	{'n_features_to_select': 35}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	99	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
35	0.19419136	0.001329599	0.00500474	6.84E-07	36	{'n_features_to_select': 36}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	95	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257

36	0.192909908	0.001312842	0.005004549	3.57E-07	37	{'n_features_to_select': 37}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	96	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
37	0.190773773	0.00149788	0.004904318	0.00020125	38	{'n_features_to_select': 38}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	110	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
38	0.189615059	0.000798382	0.005204105	0.000400831	39	{'n_features_to_select': 39}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	112	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
39	0.187843704	0.001443359	0.00536871	0.000519135	40	{'n_features_to_select': 40}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	111	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
40	0.186954021	0.00113695	0.00540247	0.000487638	41	{'n_features_to_select': 41}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	100	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
41	0.185754824	0.001218351	0.005204582	0.000400591	42	{'n_features_to_select': 42}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	119	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
42	0.184067965	0.001497198	0.005204725	0.000400162	43	{'n_features_to_select': 43}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	109	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
43	0.183305979	0.001535293	0.005404997	0.000490427	44	{'n_features_to_select': 44}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	115	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
44	0.181558371	0.000866641	0.005204773	0.000400496	45	{'n_features_to_select': 45}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	118	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
45	0.180271006	0.001201815	0.005204868	0.00040009	46	{'n_features_to_select': 46}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	114	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
46	0.178162527	0.001096889	0.005404806	0.000490291	47	{'n_features_to_select': 47}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	91	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
47	0.177361488	0.001167897	0.005004549	6.14E-07	48	{'n_features_to_select': 48}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	121	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
48	0.176205683	0.00126942	0.005004358	4.62E-07	49	{'n_features_to_select': 49}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	115	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
49	0.174749994	0.000815453	0.005205202	0.000401235	50	{'n_features_to_select': 50}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	93	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
50	0.17363863	0.001801251	0.005204964	0.000401354	51	{'n_features_to_select': 51}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	100	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
51	0.172646618	0.001418058	0.005403423	0.000488714	52	{'n_features_to_select': 52}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	120	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
52	0.171419621	0.001388339	0.005201435	0.000398365	53	{'n_features_to_select': 53}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	89	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
53	0.169568014	0.001606471	0.005004597	4.10E-07	54	{'n_features_to_select': 54}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.956143271	-0.581808574	1.235360361	90	0.681388302	0.759871147	0.293499863	0.19548231	0.819045	0.549857324	0.255017257
54	0.168089104	0.001369845	0.005004501	5.35E-07	55	{'n_features_to_select': 55}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	38	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605

55	0.166369295	0.001179259	0.005404758	0.000490524	56	{'n_features_to_select': 56}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	43	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
56	0.165010977	0.001276992	0.005406332	0.000491286	57	{'n_features_to_select': 57}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	26	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
57	0.163891935	0.001279131	0.005205011	0.000400377	58	{'n_features_to_select': 58}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	29	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
58	0.162578678	0.001303581	0.005004883	3.99E-07	59	{'n_features_to_select': 59}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	34	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
59	0.161290455	0.000609181	0.005386686	0.000468327	60	{'n_features_to_select': 60}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	44	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
60	0.1596241	0.000986587	0.005004358	5.09E-07	61	{'n_features_to_select': 61}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	41	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
61	0.157979012	0.001382644	0.005204535	0.000400257	62	{'n_features_to_select': 62}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	52	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
62	0.15694313	0.000748915	0.00500412	4.10E-07	63	{'n_features_to_select': 63}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	15	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
63	0.155242014	0.000799889	0.005204678	0.000400663	64	{'n_features_to_select': 64}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	48	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
64	0.155101871	0.001836149	0.005205154	0.000400186	65	{'n_features_to_select': 65}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	42	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
65	0.152545357	0.001307751	0.005204821	0.000400353	66	{'n_features_to_select': 66}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	51	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
66	0.151673841	0.001003711	0.00500493	7.60E-07	67	{'n_features_to_select': 67}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	46	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
67	0.150584316	0.001076685	0.005202961	0.00040178	68	{'n_features_to_select': 68}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	16	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
68	0.148784018	0.000991365	0.00500412	5.31E-07	69	{'n_features_to_select': 69}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	38	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
69	0.147004652	0.001100942	0.004996777	9.29E-06	70	{'n_features_to_select': 70}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	24	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
70	0.146217203	0.001109486	0.005204582	0.000400472	71	{'n_features_to_select': 71}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	27	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
71	0.144661236	0.001279155	0.005204487	0.000400162	72	{'n_features_to_select': 72}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	18	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
72	0.143268013	0.001001458	0.00500412	5.52E-07	73	{'n_features_to_select': 73}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	35	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
73	0.142263365	0.000934489	0.005204725	0.00040052	74	{'n_features_to_select': 74}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	49	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605

74	0.140365362	0.001227116	0.005204773	0.000400377	75	{'n_features_to_select': 75}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	37	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
75	0.139527273	0.00135776	0.005004358	5.09E-07	76	{'n_features_to_select': 76}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	31	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
76	0.137438869	0.000911311	0.005004787	3.23E-07	77	{'n_features_to_select': 77}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	50	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
77	0.136096144	0.001396102	0.005204964	0.000400281	78	{'n_features_to_select': 78}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	30	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
78	0.134357405	0.000743417	0.005004597	4.62E-07	79	{'n_features_to_select': 79}	0.799820689	0.513200085	-2.139513878	-0.100088592	-1.781610848	-0.541638509	1.199825043	12	0.681388302	0.759871147	0.293499863	0.189710061	0.822275454	0.549348965	0.257309832
79	0.132934141	0.001113603	0.005586624	0.000476908	80	{'n_features_to_select': 80}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	14	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
80	0.131574249	0.000773531	0.005204487	0.000400281	81	{'n_features_to_select': 81}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	47	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
81	0.130499458	0.001324608	0.00540514	0.000490116	82	{'n_features_to_select': 82}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	27	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
82	0.12915082	0.001636326	0.005404425	0.00049031	83	{'n_features_to_select': 83}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	33	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
83	0.127718401	0.001021462	0.005204821	0.000400472	84	{'n_features_to_select': 84}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	25	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
84	0.126335239	0.000744442	0.005204535	0.000400257	85	{'n_features_to_select': 85}	0.799820689	0.513200085	-2.139513878	-0.151918112	-1.781610848	-0.552004413	1.196183854	53	0.681388302	0.759871147	0.293499863	0.185412465	0.822275454	0.548489446	0.258514096
85	0.124429464	0.000579572	0.005204296	0.000400257	86	{'n_features_to_select': 86}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	17	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
86	0.12316041	0.000736486	0.005404997	0.000490524	87	{'n_features_to_select': 87}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	40	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
87	0.1221982	0.000752704	0.005329514	0.00041554	88	{'n_features_to_select': 88}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	22	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
88	0.120949459	0.001110172	0.005004787	5.56E-07	89	{'n_features_to_select': 89}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	21	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
89	0.118945503	0.000634036	0.004973125	3.87E-05	90	{'n_features_to_select': 90}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	20	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
90	0.117777109	0.000760122	0.005004311	7.48E-07	91	{'n_features_to_select': 91}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	35	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
91	0.116155148	0.000954623	0.005204821	0.000400472	92	{'n_features_to_select': 92}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	45	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
92	0.114938068	0.001145686	0.005004597	4.10E-07	93	{'n_features_to_select': 93}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	19	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605

93	0.113703346	0.000800658	0.005004358	7.89E-07	94	{'n_features_to_select': 94}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	31	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
94	0.112102318	0.000632938	0.005611849	0.000482368	95	{'n_features_to_select': 95}	0.799820689	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.54690209	1.197932672	23	0.681388302	0.759871147	0.293499863	0.19548231	0.822275454	0.550503415	0.255701605
95	0.110631752	0.000660218	0.005004883	4.77E-07	96	{'n_features_to_select': 96}	0.799820689	0.513200085	-2.139513878	-0.119489101	-1.781610848	-0.54551861	1.198421394	13	0.681388302	0.759871147	0.293499863	0.195368263	0.822275454	0.550480606	0.255733276
96	0.109787846	0.001195224	0.005010462	4.91E-06	97	{'n_features_to_select': 97}	0.799820689	0.513200085	-2.9146811	-0.126628488	-1.781610848	-0.701979932	1.423007114	160	0.681388302	0.759871147	0.268150487	0.195481107	0.822275454	0.545433299	0.260944909
97	0.10812149	0.001077886	0.005205059	0.000400949	98	{'n_features_to_select': 98}	0.666526226	0.513200085	-2.139513878	-0.149162152	-1.781610848	-0.578112113	1.167087799	88	0.732426253	0.759871147	0.293499863	0.190836372	0.822275454	0.559781818	0.262972079
98	0.106496525	0.001020273	0.005004358	6.47E-07	99	{'n_features_to_select': 99}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	71	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
99	0.105895567	0.00072768	0.005186272	0.000411177	100	{'n_features_to_select': 100}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	73	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
100	0.103927231	0.000760396	0.005205107	0.000400021	101	{'n_features_to_select': 101}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	75	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
101	0.103174353	0.001360942	0.005004025	2.43E-07	102	{'n_features_to_select': 102}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	65	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
102	0.100934029	0.000764899	0.005004263	4.91E-07	103	{'n_features_to_select': 103}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	76	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
103	0.098890162	0.000400353	0.005204725	0.00040052	104	{'n_features_to_select': 104}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	62	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
104	0.098549414	0.000571351	0.005004597	3.50E-07	105	{'n_features_to_select': 105}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	67	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
105	0.096460485	0.000536721	0.005204582	0.000400472	106	{'n_features_to_select': 106}	0.691353479	0.513200085	-2.139513878	-0.124099777	-1.781610848	-0.568134188	1.17426737	58	0.733109286	0.759871147	0.293499863	0.195465827	0.822275454	0.560844315	0.261766169
106	0.097250175	0.001143182	0.005351019	0.000382707	107	{'n_features_to_select': 107}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	66	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
107	0.094273329	0.00075307	0.005204439	0.000400543	108	{'n_features_to_select': 108}	0.691353479	0.513200085	-2.139513878	-0.158630772	-1.781610848	-0.575040387	1.171734381	87	0.733109286	0.759871147	0.293499863	0.180348303	0.822275454	0.557820811	0.266021692
108	0.110839462	0.020440907	0.006000376	0.001101459	109	{'n_features_to_select': 109}	0.691353479	0.513200085	-2.139513878	-0.115499391	-1.781610848	-0.566414111	1.174922652	57	0.733109286	0.759871147	0.293499863	0.195273239	0.822275454	0.560805798	0.261819938
109	0.093939352	0.002328522	0.005604982	0.000490466	110	{'n_features_to_select': 110}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	74	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
110	0.091269922	0.0015599	0.005404663	0.000490213	111	{'n_features_to_select': 111}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	69	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
111	0.089032364	0.000853024	0.005112743	0.00021596	112	{'n_features_to_select': 112}	0.691353479	0.513200085	-2.139513878	-0.095630306	-1.781610848	-0.562440294	1.17647359	56	0.733109286	0.759871147	0.293499863	0.184874216	0.822275454	0.558725993	0.264740354

112	0.086717606	0.000527127	0.005404806	0.000490485	113	{'n_features_to_select': 113}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	67	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
113	0.088481522	0.00162526	0.005204678	0.000400543	114	{'n_features_to_select': 114}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	64	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
114	0.084111214	0.000858739	0.005204582	0.000400233	115	{'n_features_to_select': 115}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	63	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
115	0.082275009	0.000401569	0.005004501	7.63E-07	116	{'n_features_to_select': 116}	0.691353479	0.513200085	-3.158823141	-0.126406496	-1.781610848	-0.772457384	1.478253791	162	0.733109286	0.759871147	0.275034023	0.19548231	0.822275454	0.557154444	0.265609481
116	0.08087368	0.000749208	0.005004454	4.86E-07	117	{'n_features_to_select': 117}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	72	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
117	0.079706573	0.00077789	0.005004835	2.78E-07	118	{'n_features_to_select': 118}	0.691353479	0.513200085	-2.139513878	-0.088563449	-1.781610848	-0.561026922	1.177037659	54	0.733109286	0.759871147	0.293499863	0.188760967	0.822275454	0.559503343	0.263644921
118	0.078471375	0.000490602	0.005004454	4.10E-07	119	{'n_features_to_select': 119}	0.691353479	0.513200085	-2.139513878	-0.09363158	-1.781610848	-0.562040548	1.176632465	55	0.733109286	0.759871147	0.293499863	0.193835761	0.822275454	0.560518302	0.262221641
119	0.076972628	0.000915893	0.005205107	0.000400209	120	{'n_features_to_select': 120}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	78	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
120	0.075593853	0.000531697	0.005003977	1.53E-06	121	{'n_features_to_select': 121}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	61	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
121	0.074073792	0.001097465	0.005805874	0.00074894	122	{'n_features_to_select': 122}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	70	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
122	0.072265959	0.000400138	0.005204821	0.000400234	123	{'n_features_to_select': 123}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	59	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
123	0.071029663	0.000540116	0.005005169	6.64E-07	124	{'n_features_to_select': 124}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	79	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
124	0.069346094	0.000487141	0.005004072	3.87E-07	125	{'n_features_to_select': 125}	0.691353479	0.513200085	-3.470949092	-0.185235087	-1.781610848	-0.846648293	1.575640451	164	0.733109286	0.759871147	0.227903609	0.188111097	0.822275454	0.546254119	0.277974221
125	0.067806292	0.000741961	0.00559988	0.000489624	126	{'n_features_to_select': 126}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	77	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
126	0.067004585	0.000631744	0.005393124	0.000488294	127	{'n_features_to_select': 127}	0.691353479	0.513200085	-2.832074056	0.004481397	-1.781610848	-0.680929989	1.386925843	158	0.733109286	0.759871147	0.254210506	0.097777089	0.822275454	0.533448697	0.297434392
127	0.065375853	0.000507429	0.005601358	0.000487317	128	{'n_features_to_select': 128}	0.691353479	0.513200085	-2.139513878	-0.222701733	-1.781610848	-0.587854579	1.16745294	123	0.733109286	0.759871147	0.293499863	0.184695639	0.822275454	0.558690278	0.264790794
128	0.064003897	0.000887794	0.005197668	0.000400662	129	{'n_features_to_select': 129}	0.691353479	0.513200085	-2.139513878	-0.126406496	-1.781610848	-0.568595532	1.174093269	60	0.733109286	0.759871147	0.293499863	0.19548231	0.822275454	0.560847612	0.261761568
129	0.063197422	0.000976051	0.005002308	5.27E-06	130	{'n_features_to_select': 130}	0.691353479	0.513200085	-2.139513878	-0.147739465	-1.781610848	-0.572862125	1.172516327	85	0.733109286	0.759871147	0.293499863	0.194558101	0.822275454	0.56066277	0.262019703
130	0.061404467	0.000484288	0.005003119	3.42E-06	131	{'n_features_to_select': 131}	0.691353479	0.513200085	-2.139513878	-0.149028959	-1.781610848	-0.573120024	1.172422929	86	0.733109286	0.759871147	0.293499863	0.187200242	0.822275454	0.559191198	0.264084239

131	0.059792328	0.000424689	0.005004311	2.43E-07	132	{'n_features_to_select': 132}	0.691353479	0.513200085	-2.139513878	-0.471293243	-1.781610848	-0.637572881	1.156081505	130	0.733109286	0.759871147	0.293499863	-0.230211421	0.822275454	0.475708866	0.399658642
132	0.058249187	0.000381325	0.005387068	0.000469618	133	{'n_features_to_select': 133}	0.691353479	0.513200085	-1.737848398	-0.134498868	-1.781610848	-0.48988091	1.072732179	10	0.733109286	0.759871147	0.273621152	0.190528317	0.822275454	0.555881071	0.267260568
133	0.056851912	0.000400448	0.005307102	0.000401591	134	{'n_features_to_select': 134}	0.691353479	0.513200085	-2.139513878	-0.126406496	-2.452089185	-0.702691199	1.33256962	161	0.733109286	0.759871147	0.293499863	0.19548231	0.820919898	0.560576501	0.261491224
134	0.055650473	0.000490271	0.005004644	4.77E-07	135	{'n_features_to_select': 135}	0.691353479	0.513200085	-2.147719746	-0.197990273	-2.215999484	-0.671431188	1.268841891	153	0.733109286	0.759871147	0.293446912	0.182124266	0.824138198	0.558537962	0.265900064
135	0.054083109	0.000580828	0.005004501	5.76E-07	136	{'n_features_to_select': 136}	0.691353479	0.513200085	-2.139513878	0.02477272	-2.215999484	-0.625237416	1.286508598	129	0.733109286	0.759871147	0.293499863	-0.043762496	0.824138198	0.513371199	0.335965635
136	0.052682352	0.000452294	0.005203915	0.000400569	137	{'n_features_to_select': 137}	0.691353479	0.513200085	-2.139513878	-0.113876948	-2.215999484	-0.652967349	1.273628947	137	0.733109286	0.759871147	0.293499863	0.19051444	0.824138198	0.560226587	0.263524547
137	0.051246643	0.000400543	0.005004549	4.42E-07	138	{'n_features_to_select': 138}	0.691353479	0.513200085	-2.139513878	-0.126406496	-2.215999484	-0.655473259	1.272577695	146	0.733109286	0.759871147	0.293499863	0.19548231	0.824138198	0.561220161	0.262134436
138	0.050245953	0.000400448	0.005004549	5.35E-07	139	{'n_features_to_select': 139}	0.691353479	0.513200085	-2.139513878	-0.126406496	-2.215999484	-0.655473259	1.272577695	143	0.733109286	0.759871147	0.293499863	0.19548231	0.824138198	0.561220161	0.262134436
139	0.048160458	0.000518277	0.005404854	0.000490447	140	{'n_features_to_select': 140}	0.691353479	0.513200085	-2.139513878	-0.237641479	-2.215999484	-0.677720255	1.264078087	156	0.733109286	0.759871147	0.293499863	0.159484444	0.824138198	0.554020588	0.272374946
140	0.046642733	0.000490914	0.005404615	0.000490544	141	{'n_features_to_select': 141}	0.691353479	0.513200085	-2.139513878	-0.050925825	-2.215999484	-0.640377125	1.279194781	131	0.733109286	0.759871147	0.293499863	0.179209783	0.824138198	0.557965655	0.266715994
141	0.045441675	0.00049033	0.005204582	0.000400829	142	{'n_features_to_select': 142}	0.691353479	0.513200085	-2.139513878	-0.070329806	-2.215999484	-0.644257921	1.277428843	133	0.733109286	0.759871147	0.293499863	0.189478839	0.824138198	0.560019467	0.263815291
142	0.044440746	0.000490271	0.005017614	2.51E-05	143	{'n_features_to_select': 143}	0.691353479	0.513200085	-2.139513878	-0.126406496	-2.215999484	-0.655473259	1.272577695	142	0.733109286	0.759871147	0.293499863	0.19548231	0.824138198	0.561220161	0.262134436
143	0.042839193	0.000400496	0.005204725	0.0004004	144	{'n_features_to_select': 144}	0.691353479	0.513200085	-2.139513878	-0.059610639	-2.215999484	-0.642114087	1.27839886	132	0.733109286	0.759871147	0.293499863	0.152980787	0.824138198	0.552719856	0.274264926
144	0.04137826	0.000424385	0.005404997	0.000489552	145	{'n_features_to_select': 145}	0.691353479	0.513200085	-2.139513878	-0.220691869	-2.215999484	-0.674330333	1.265275876	155	0.733109286	0.759871147	0.293499863	0.167345558	0.824138198	0.55559281	0.270106277
145	0.039959764	0.000157901	0.005204296	0.000400376	146	{'n_features_to_select': 146}	0.691353479	0.513200085	-2.139513878	-0.126406496	-2.215999484	-0.655473259	1.272577695	149	0.733109286	0.759871147	0.293499863	0.19548231	0.824138198	0.561220161	0.262134436
146	0.038435078	0.000490583	0.005004692	3.81E-07	147	{'n_features_to_select': 147}	0.691353479	0.513200085	-2.139513878	-0.151641829	-2.215999484	-0.660520325	1.270517772	151	0.733109286	0.759871147	0.293499863	0.190855064	0.824138198	0.560294712	0.263428989
147	0.036923218	0.000392779	0.005238676	0.000396222	148	{'n_features_to_select': 148}	0.691353479	0.513200085	-2.139513878	-0.163619181	-2.215999484	-0.662915796	1.269566996	152	0.733109286	0.759871147	0.293499863	0.180925488	0.824138198	0.558308796	0.266229148
148	0.035832977	0.00040083	0.005204535	0.000400734	149	{'n_features_to_select': 149}	0.691353479	0.513200085	-2.139513878	-0.126406496	-2.215999484	-0.655473259	1.272577695	148	0.733109286	0.759871147	0.293499863	0.19548231	0.824138198	0.561220161	0.262134436
149	0.034030962	1.78E-07	0.005004454	2.78E-07	150	{'n_features_to_select': 150}	0.691353479	0.513200085	-2.139513878	-0.250228711	-2.215999484	-0.680237702	1.263211388	157	0.733109286	0.759871147	0.293499863	0.167738232	0.824138198	0.555671345	0.269993414

150	0.032390928	0.000479158	0.005204535	0.000398708	151	{'n_features_to_select': 151}	0.691353479	0.513200085	-2.139513878	-0.113970186	-2.215999484	-0.652985997	1.273621054	138	0.733109286	0.759871147	0.293499863	0.193768276	0.824138198	0.560877354	0.26261319
151	0.031200743	0.000400496	0.005398083	0.000488198	152	{'n_features_to_select': 152}	0.691353479	0.513200085	-2.139513878	-0.092679283	-2.215999484	-0.648727816	1.275450341	134	0.733109286	0.759871147	0.293499863	0.192512555	0.824138198	0.56062621	0.262964512
152	0.029601908	0.000488598	0.005200958	0.000397042	153	{'n_features_to_select': 153}	0.691353479	0.513200085	-2.139513878	-0.210307354	-2.215999484	-0.67225343	1.266027102	154	0.733109286	0.759871147	0.293499863	0.173157236	0.824138198	0.556755146	0.268440418
153	0.028199291	0.000401767	0.00500021	3.95E-06	154	{'n_features_to_select': 154}	0.691353479	0.513200085	-2.139513878	-0.144202344	-2.215999484	-0.659032428	1.271117061	150	0.733109286	0.759871147	0.293499863	0.193733164	0.824138198	0.560870332	0.262623007
154	0.02701931	0.000624602	0.005004501	3.23E-07	155	{'n_features_to_select': 155}	0.691353479	0.513200085	-2.139513878	-0.126406496	-2.215999484	-0.655473259	1.272577695	147	0.733109286	0.759871147	0.293499863	0.19548231	0.824138198	0.561220161	0.262134436
155	0.02558713	0.000499809	0.00520463	0.000400448	156	{'n_features_to_select': 156}	0.691353479	0.513200085	-2.139513878	-0.126406496	-2.215999484	-0.655473259	1.272577695	144	0.733109286	0.759871147	0.293499863	0.19548231	0.824138198	0.561220161	0.262134436
156	0.023785162	0.000496059	0.00500412	1.33E-06	157	{'n_features_to_select': 157}	0.691353479	0.513200085	-2.139513878	-0.126406496	-2.215999484	-0.655473259	1.272577695	145	0.733109286	0.759871147	0.293499863	0.19548231	0.824138198	0.561220161	0.262134436
157	0.022511864	0.000448157	0.00500412	3.50E-07	158	{'n_features_to_select': 158}	0.691353479	0.372911888	-2.139513878	-0.198451579	-2.215999484	-0.697939915	1.24169148	159	0.733109286	0.824269021	0.293499863	0.157426273	0.824138198	0.566488528	0.283708072
158	0.020999765	3.65E-05	0.005004883	5.84E-07	159	{'n_features_to_select': 159}	0.691353479	0.383968513	-4.32204091	-0.469411097	-2.215999484	-1.1864259	1.865390441	167	0.733109286	0.82799071	0.039678892	0.0713203	0.824138198	0.499247477	0.36404461
159	0.019781113	0.000384462	0.005004454	6.47E-07	160	{'n_features_to_select': 160}	0.691353479	0.383968513	-0.495754112	-0.081431643	-2.215999484	-0.34357265	1.01963205	8	0.733109286	0.82799071	0.3305236	0.145041184	0.824138198	0.572160595	0.281306121
160	0.017992973	4.38E-05	0.004998636	7.07E-06	161	{'n_features_to_select': 161}	0.691353479	0.383968513	0.512979333	0.079087608	-2.215999484	-0.10972211	1.07193872	1	0.733109286	0.82799071	0.72936741	0.252039836	0.824138198	0.673329088	0.214877735
161	0.016620493	0.000499842	0.004998302	5.83E-06	162	{'n_features_to_select': 162}	0.691353479	0.383968513	0.40146702	0.005388609	-2.215999484	-0.146764373	1.057313678	3	0.733109286	0.82799071	0.735189951	0.269516542	0.824138198	0.677988937	0.208336288
162	0.015000105	4.03E-06	0.005005121	6.35E-06	163	{'n_features_to_select': 163}	0.691353479	0.383968513	0.40146702	0.005388609	-2.215999484	-0.146764373	1.057313678	2	0.733109286	0.82799071	0.735189951	0.269516542	0.824138198	0.677988937	0.208336288
163	0.013783026	0.000391122	0.005200338	0.000397474	164	{'n_features_to_select': 164}	0.691353479	0.383968513	-0.872078511	0.005388609	-2.215999484	-0.401473479	1.047921787	9	0.733109286	0.82799071	0.759765023	0.269516542	0.824138198	0.682903952	0.20991171
164	0.012206602	0.00039793	0.005204535	0.000400496	165	{'n_features_to_select': 165}	0.691353479	0.383968513	-0.576857897	0.005388609	-2.215999484	-0.342429356	1.027867687	7	0.733109286	0.82799071	0.79535663	0.269516542	0.824138198	0.690022273	0.212978524
165	0.011002398	3.78E-06	0.004983234	3.48E-05	166	{'n_features_to_select': 166}	0.691353479	0.383968513	-0.301314237	0.370244663	-2.215999484	-0.214349413	1.052064465	6	0.733109286	0.82799071	0.798045025	0.762248305	0.824138198	0.789106305	0.036540798
166	0.009517813	0.000445819	0.005085135	0.000163951	167	{'n_features_to_select': 167}	0.691353479	0.383968513	-0.301314237	0.426852863	-2.215999484	-0.203027773	1.058578992	4	0.733109286	0.82799071	0.798045025	0.764626551	0.824138198	0.789581954	0.036202001
167	0.008200359	0.00039987	0.005002975	2.78E-06	168	{'n_features_to_select': 168}	0.691353479	0.383968513	-0.301314237	0.392341307	-2.215999484	-0.209930085	1.054554307	5	0.733109286	0.82799071	0.798045025	0.772713573	0.824138198	0.791199358	0.035218219