The Data Cards Playbook

pair-code.github.io/datacardsplaybook

Data Cards Template (Beta) | Updated November 2022

ABRIDGED VERSION PRODUCED FOR I-GUIDE SUMMER SCHOOL August 2024

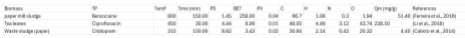
# Excel Dataset for Pharmaceutical Removal via Biochar Adsorption	The dataset is used for a group project for I-GUIDE Summer School 2024. The dataset contains information extracted from scholarly literature, searched and extracted from Google Scholar, Web of Science, Scopus and PubMed. 'Pharmaceutical Waste Removal' has been used as the major keyword while searching for articles. The collected information has been inserted into an excel file across and these information included - pore volume, pyrolysis temperature, types of pharmaceuticals, excess molar refraction, solute dipolarity/polarizability, Hydrogen bond acidity, Hydrogen bond basicity, McGowan's characteristics volume, gas-to-hexadecane partition, pH of the solution and point of zero charge of sorbents.
DATASET LINK	DATA CARD AUTHOR(S)
Provide a link to the dataset:	*Select **one role per** Data Card Author:* *(Usage Note: Select the most appropriate choice to describe the author's role in creating the Data Card.)*
https://platform.i-guide.io/datasets/f2ee85bd-1b94-4583-b50a-060afd9deddc	**Jarely Mendez, Team 1:** Data Card Manager

Authorship

Dataset Owners

TEAM(S)	CONTACT DETAIL(S)	AUTHOR(S)
Provide the names of the groups or team(s) that own the dataset:	*Provide pathways to contact dataset owners:*	*Provide the details of all authors associated with the dataset:*
Team 1 of I-GUIDE Summer School 2024	**Dataset Owner(s):** Jude A. Okolie **Affiliation:** University of Oklahoma **Contact:** jude.a.okolie@gmail.com	Jude A. Okolie, Title, University of Oklahoma, 2024

Dataset Overview

DATA SUBJECT(S)	DATASET SNAPSHOT	CONTENT DESCRIPTION
*Select **all applicable** subjects contained the dataset:*	*Provide a snapshot of the dataset:*	*Provide a short description of the content in a datapoint.*
Data about places and objects	Size of Dataset: 35 KB Number of Instances: 86 Number of Fields: 14 **Visualization:** 	The dataset contains information extracted from scholarly literature, searched and extracted from Google Scholar, Web of Science, Scopus and PubMed. 'Pharmaceutical Waste Removal' has been used as the major keyword while searching for articles. The collected information has been inserted into an Excel file.

Sensitivity of Data: NOT APPLICABLE

SENSITIVITY TYPE(S)	FIELD(S) WITH SENSITIVE DATA	SECURITY AND PRIVACY HANDLING
*Select **all applicable** data types present in the dataset:*	*List fields in the dataset that contain PII, and specify if their collection was intentional or unintentional.*	*Summarize the steps to handle sensitive data in this dataset, including steps taken to identify and mitigate risks from PII or sensitive information.*

Example of Data Points

PRIMARY DATA MODALITY		DATA FIELDS
*Select **one**:*		*List the fields in data points and their descriptions.* *(Usage Note: Describe each field in a datapoint. Optionally use this to show the example.)*
Text Data		
TYPICAL DATA POINT		ATYPICAL DATA POINT
Provide an example of a typical data point and describe what makes it typical. ***Use additional notes to capture any other relevant information or considerations.***		*Provide an example of an outlier data point and describe what makes it atypical.* ***Use additional notes to capture any other relevant information or considerations.***

Motivations & Intentions

Motivations

PURPOSE(S)	RESEARCH AND PROBLEM SPACE(S)	MOTIVATING FACTOR(S)
*Select **one**:*	*Provide a description of the specific problem space that this dataset intends to address.*	*List the primary motivations for creating or curating this dataset:* *(Usage Note: use this to describe the problem space and corresponding motivations for the dataset.)*

Monitoring **Research** Production Others (Please Specify)	The proposed project aims to address concerns related to the removal of pharmaceuticals pollutants from wastewater, a critical environmental challenge due to the increasing prevalence of these pollutants. Pharmaceuticals as emerging pollutants pose significant risks to both human health and the aquatic ecosystem. Their microscopic nature allows them to often go undetected, thereby complicating efforts to mitigate their impact. Biochar produced waste materials could be used as a promising adsorbent for the remediation of pharmaceutical pollutants. The specific questions to be answered include: How effective is biochar, compared to other adsorbents in removing various pharmaceutical micropollutants from wastewater? What are the optimal conditions under which biochar can maximize the adsorption of these pollutants? Additionally, the research aims to understand the relationship between the properties of biochar and its efficiency in removing different types of pharmaceuticals. Furthermore, publicly available software will be developed for predicting biochar effectiveness in removing pharmaceuticals from wastewater. The proposed problem is intrinsically linked to the theme of water security, a central focus of the Summer School 2024. Water security involves ensuring sustainable access to safe and clean water, a necessity that is compromised by the presence of pharmaceutical pollutants. The research into biochar as a cost-effective and eco-friendly adsorbent aligns with the goals of water security by providing a viable solution for clean water and waste valorization. It addresses the urgent need for innovative and sustainable technologies to mitigate water pollution, thereby safeguarding both environmental and public health. The proposed project contributes to the broader goal of maintaining and improving water	• *Pharmaceuticals can affect aquatic life even at low concentrations.* • *Hormonal drugs can disrupt reproduction and growth in fish and amphibians.* • *Antibiotics can promote antibiotic-resistant bacteria in water bodies.* • *Pharmaceuticals can accumulate in the tissues of aquatic organisms, leading to higher concentrations as you move up the food chain. This can impact predator species, including humans who consume contaminated fish and shellfish.* • *Trace amounts of pharmaceuticals can enter drinking water supplies leading to contamination and potential health risks for humans.* • *Long-term exposure may pose health risks, especially for vulnerable populations such as children and pregnant women.* • *Presence of antibiotics in wastewater contributes to the development of antibiotic-resistant bacteria, posing a public health challenge.* • *Many regions have strict regulations regarding wastewater treatment and pharmaceutical contaminants. Non-compliance can result in hefty fines and reputational damage.* • *Contaminated water can affect industries relying on clean water, such as tourism and fisheries.*

	quality, a key aspect of water security.	

Intended Use

DATASET USE(S)	SUITABLE USE CASE(S)	UNSUITABLE USE CASE(S)
*Select **one**:*	*Summarize known suitable and intended use cases of this dataset.*	*Summarize known unsuitable and unintended use cases of this dataset.* .
Safe for production use **Safe for research use** Conditional use-some unsafe applications Only approved use Others (Please specify)	**Research**: Where the dataset is fit-for-use and of good enough quality to use. Where it is ethically acceptable to use the dataset.	**N/A**

Provenance

Collection

METHOD(S) USED	METHODOLOGY DETAIL(S)	SOURCE DESCRIPTION(S)
*Select **all applicable** methods used to collect data:*	*Provide a description of each collection method used.*	*Provide a description of each upstream source of data.*
Taken from other existing datasets	The dataset contains information extracted from scholarly literature, searched and extracted from Google Scholar, Web of Science, Scopus and PubMed. **Dates of Collection:** [01 2024 - 04 2024] **Primary modality of collected data:** *Usage Note: Select one for this collection type.* **Text Data**	Sources available on Excel sheet.
COLLECTION CADENCE	DATA INTEGRATION	DATA PROCESSING
*Select **all applicable**:*	*List all fields collected from different sources, and specify if they were included or excluded from the dataset.* *(Usage Note: Duplicate and complete the following for each upstream source.)*	*Summarize how data from different sources or methods aggregated, processed, or connected.* *(Usage Note: Duplicate and complete the following for each source OR collection method.)*
Streamed (Data is continuously acquired from single or multiple sources.)		**Data aggregation done using Microsoft Excel.**

Human and Other Sensitive Attributes — NOT APPLICABLE

SENSITIVE HUMAN ATTRIBUTE(S)	INTENTIONALITY	METHODOLOGY DETAIL(S)
*Select **all attributes** that are represented (directly or indirectly) in the dataset.* *[Note: there are demonstrative examples of potentially relevant signals for a given dataset analysis]*	*List fields in the dataset that contain human attributes, and specify if their collection was intentional or unintentional.* *Use additional notes to capture any other relevant information or considerations.*	*Describe the methods used to collect human attributes in the dataset.*

Extended Use

Use with Other Data

SAFETY LEVEL	KNOWN SAFE DATASET(S) OR DATA TYPE(S)	BEST PRACTICES
*Select **one**:*	*List the known datasets or data types and corresponding transformations that **are safe to join or aggregate** this dataset with.*	*Summarize best practices for using this dataset with other datasets or data types.* *Use additional notes to capture any other relevant information or considerations.*
Safe to use with other data	**Dataset or Data Type:** <Summarize here. Include visualizations, metrics, or links where necessary.>	This Excel datasheet can be used with other datasheets or used to support literature.
	KNOWN UNSAFE DATASET(S) OR DATA TYPE(S)	LIMITATION(S) AND RECOMMENDATION(S)
Fill out this row if you selected "Conditionally safe to use with other datasets" or "Should not be used with other datasets":	*List the known datasets or data types and corresponding transformations that are **unsafe to join or aggregate** with this dataset.*	*Summarize limitations of the dataset that introduce foreseeable risks when the dataset is conjoined with other datasets.*

Transformations

Fill this section if any transformations were applied in the creation of your dataset.

Synopsis

TRANSFORMATION(S) APPLIED	FIELD(S) TRANSFORMED	LIBRARY(IES) AND METHOD(S) USED
*Select **all applicable** transformations that were applied to the dataset.*	*Provide the fields in the dataset that were transformed.* *(Usage Note: Duplicate and complete the following for each transformation type applied. Include the data types to which fields were transformed.)*	*Provide a description of the methods used to transform or process the dataset.* *(Usage Note: Duplicate and complete the following for each transformation type applied.)*
Cleaning Missing Values **Data Aggregation**		

Breakdown of Transformations

Fill out relevant rows.

CLEANING MISSING VALUE(S)	METHOD(S) USED	COMPARATIVE SUMMARY
Which fields in the data were missing values? How many?	*How were missing values cleaned?*	*Why were missing values cleaned using this method (over others)? Provide comparative charts showing before and after missing values were cleaned.*
There are 18 blank cells.		

CLEANING MISMATCHED VALUE(S)	METHOD(S) USED	COMPARATIVE SUMMARY
Which fields in the data were corrected for mismatched values?	*How were incorrect or mismatched values cleaned?*	*Why were incorrect or mismatched values cleaned using this method (over others)? Provide a comparative analysis demonstrating before and after values were cleaned.*

ANOMALIES	METHOD(S) USED	COMPARATIVE SUMMARY
How many anomalies or outliers were detected? *If at all, how were detected anomalies or outliers handled? Why or why not?*	*What methods were used to detect anomalies or outliers?*	*Provide a comparative analysis demonstrating before and after anomaly handling measures.*
DIMENSIONALITY REDUCTION	METHOD(S) USED	COMPARATIVE SUMMARY
How many original features were collected and how many dimensions were reduced?	*What methods were used to reduce the dimensionality of the data?*	*Why were features reduced using this method (over others)? Provide comparative charts showing before and after dimensionality reduction processes.*
JOINING INPUT SOURCES	METHOD(S) USED	COMPARATIVE SUMMARY
What were the distinct input sources that were joined?	*What are the shared columns of fields used to join these sources?*	*Why were features joined using this method over others?* *Provide comparative charts showing before and after dimensionality reduction processes.*