

# International Academy, Research, and Industry Association (IARIA) AIHealth 2024 Conference



---

## Cancer : Investigating the Impact of the Implementation Platform on Machine Learning Models

**Adedayo Olowolayemo, MSc | Amina Souag, PhD  
Konstantinos Sirlantzis, Prof.**

March 2024



# The Urgency of Accurate Cancer Diagnosis

- Accurate cancer diagnosis is imperative for timely treatment and improved patient outcomes.

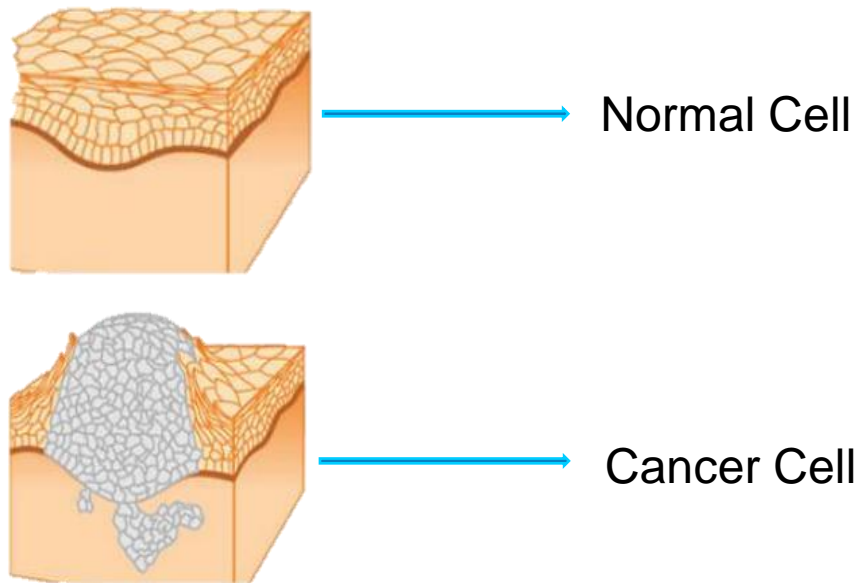


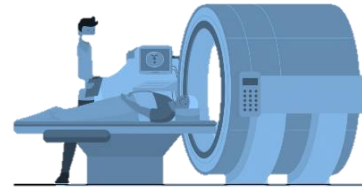
Figure 1 -Pictorial view of a Normal Cell and a Cancer Cell (NCI, 2021)

- Cancer arises when genetic and environmental factors disrupt the normal cell cycle, causing cells to proliferate uncontrollably, forming malignant (cancerous) and benign (non-cancerous) tumours.
- The rapid increase in cancer cases is expected to reach 30 million annually by 2040 (Zhitnyuk et al., 2022)
- Cancer is responsible for nearly 10,000,000 deaths globally in the year 2020 alone (Chhikara & Parang, 2023)

# Evolution of Cancer Diagnosis Platforms From Traditional Methods to Data-driven Machine Learning Models



Traditional cancer diagnosis relied on manual methods and basic tests, limiting accuracy and timely detection.



Technological advances introduced detailed imaging and molecular diagnostics, enabling personalized cancer treatment.



Machine learning algorithms utilize vast healthcare data to enhance cancer diagnosis and treatment, aiming to improve patient outcomes.

**Research  
Gap**

What is the impact of implementation platforms on algorithm performance in cancer classification?

# Research Aims

Evaluate and contrast the performance of the same algorithms when implemented on Python SciKit-Learn platform and Knime analytics .

01

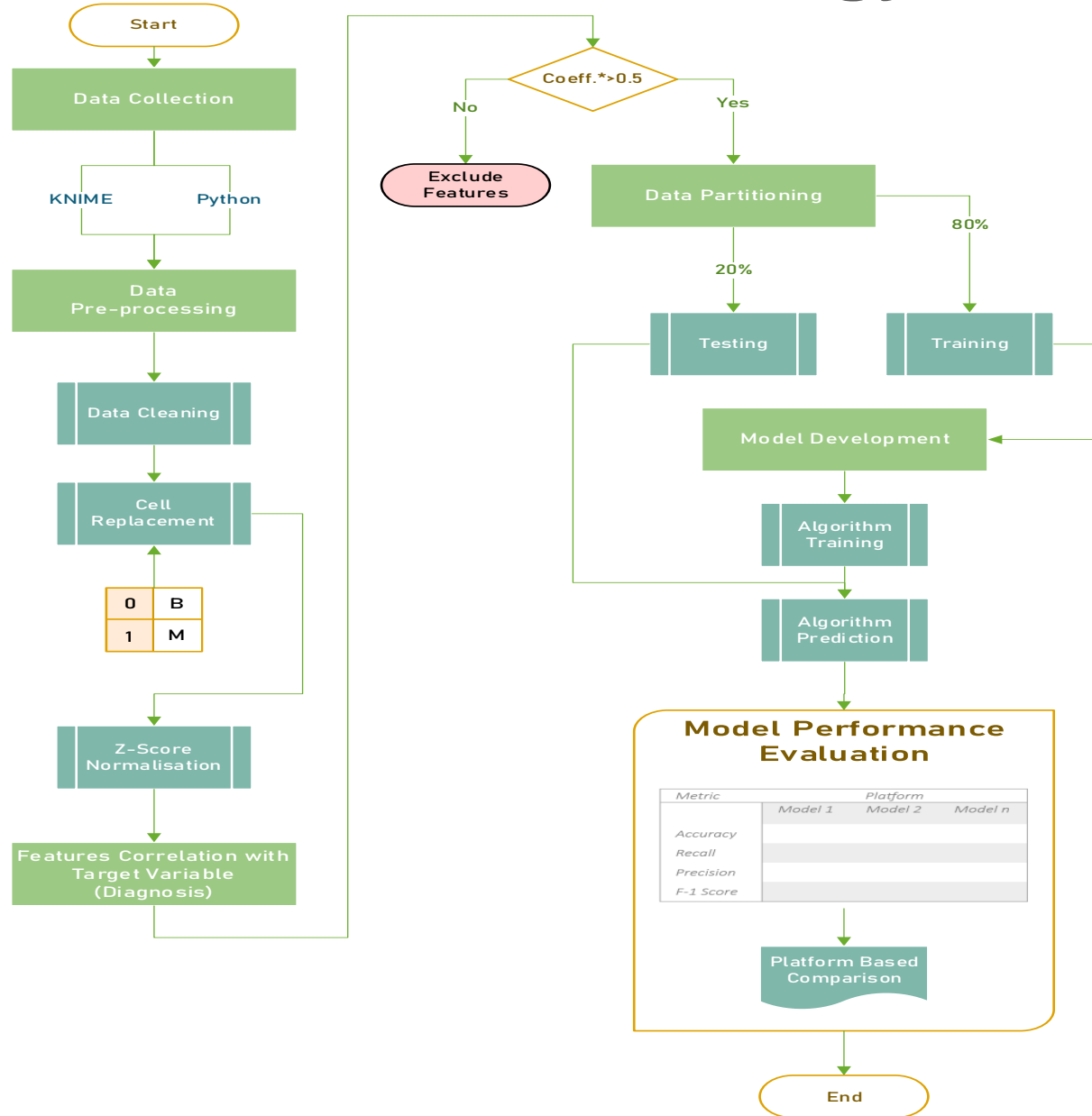
Investigate whether the selection of an implementation platform influences the performance of selected algorithms in the classification of cancer data.

02

Assess and compare performance of classification algorithms including Logistic Regression, Decision Trees, Random Forests & Gradient.

03

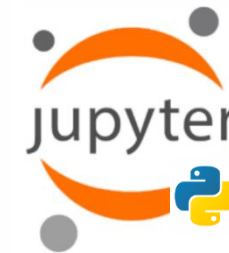
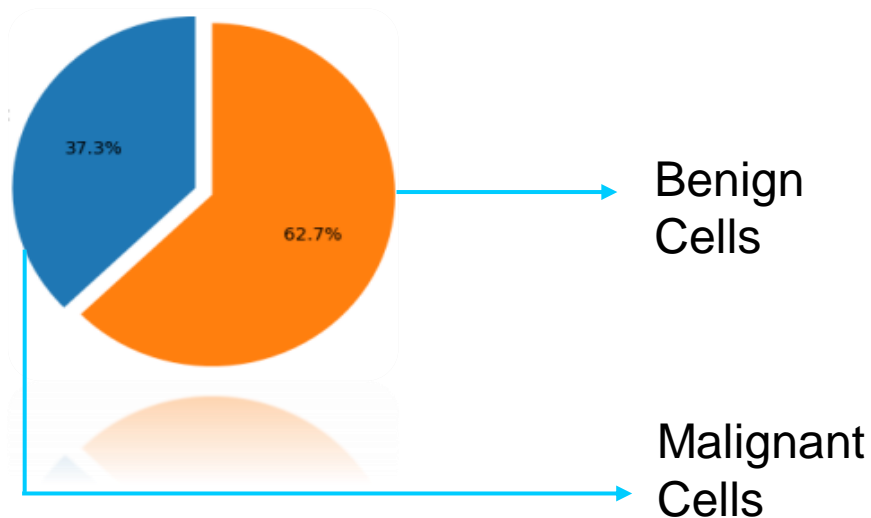
# Research Methodology



# Dataset and Implementation Platforms

## Wisconsin Diagnostic Breast Cancer (WDBC)

- An open-source multivariate dataset publicly available on UCI Machine Learning Repository (<https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>)
- Has 569 instances and 30 features.
- It is a useful resource in medical informatics and machine learning, aiding in the creation and evaluation of classification algorithms.



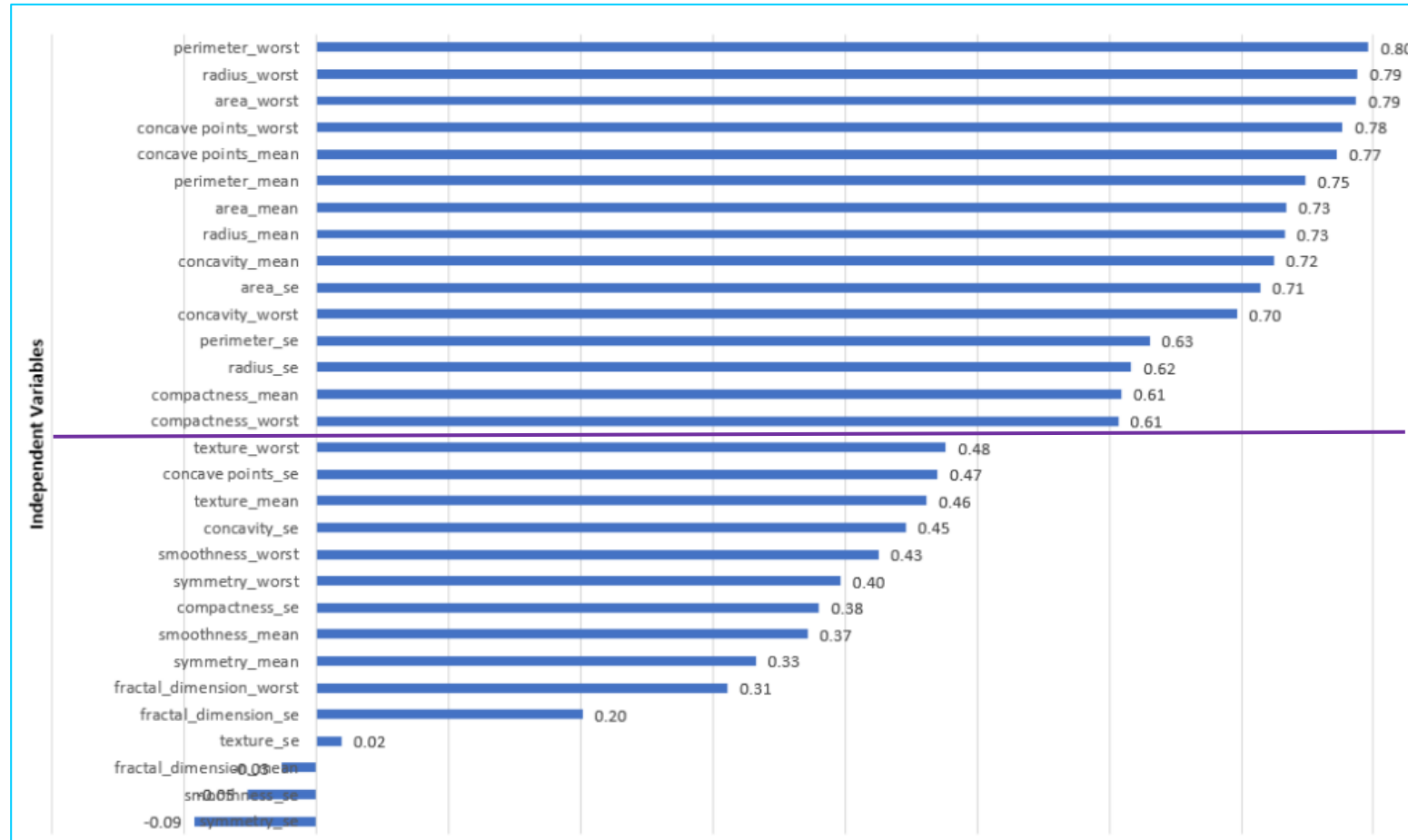
(Version 3.11.4)



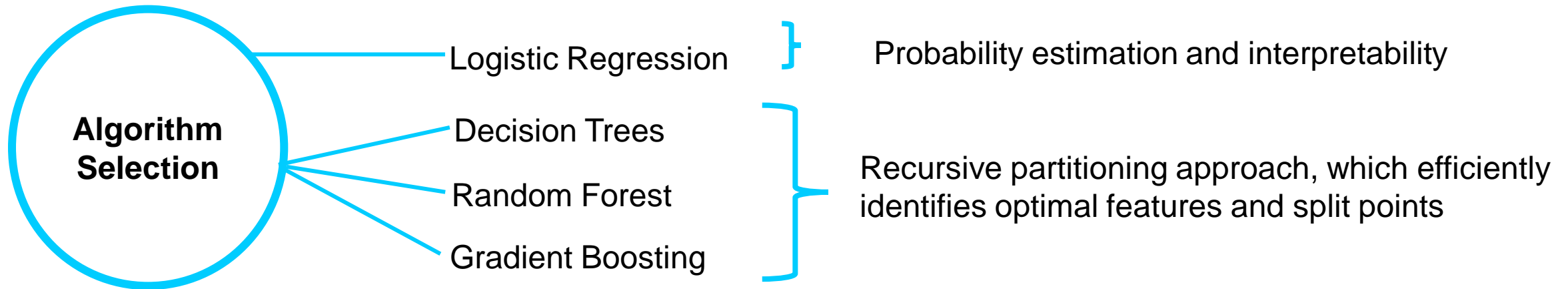
(Version 4.7.6)

# Feature selection

The study employed both the Filter Method and the Tree-Based Method to select essential features: initially, the Filter Method assessed features based on correlation scores with the target variable, retaining 15 out of 30 features.



# Model Selection



- Logistic Regression, a powerful tool in binary classification, stands out for its effectiveness
- Tree-based algorithms like Decision Trees, Random Forests, and Gradient Boosting which offer transparency, ensemble learning, and iterative refinement, respectively, enhancing predictive accuracy across various domains. These algorithms excel in modeling complex relationships and accommodating non-linearities, making them indispensable in this study.



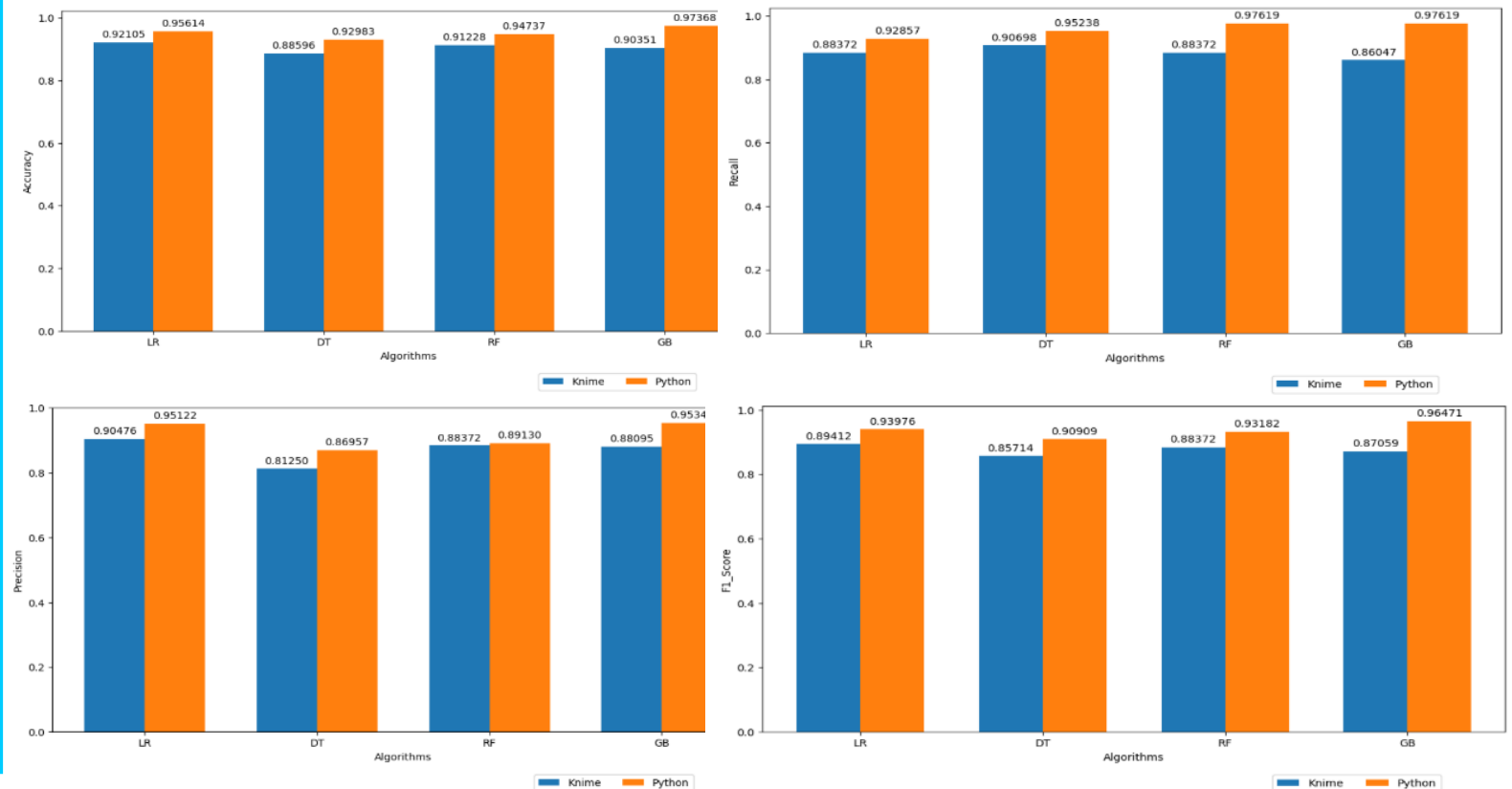
# Result

A comparative view of different algorithms performances across on the two platforms across the four metrics assessed.

Result Table

Algorithm	Tool	Accuracy	Recall	Precision	F1-Score
LR	SciKit-Learn	0.956	0.929	0.951	0.940
	Knime	0.921	0.884	0.905	0.894
DT	SciKit-Learn	0.930	0.952	0.870	0.909
	Knime	0.886	0.907	0.813	0.857
RF	SciKit-Learn	0.947	0.976	0.891	0.932
	Knime	0.912	0.884	0.884	0.884
GB	SciKit-Learn	0.974	0.976	0.953	0.965
	Knime	0.904	0.861	0.881	0.871

Result Visualization



# Key Findings

The choice of implementation platform (Python SciKit-Learn vs. Knime Analytics) impacts the performance of ML models.

The Gradient Boosting (GB) algorithm performed exceptionally well on the Python platform.

Python SciKit-Learn generally produced better results across all performance metrics (accuracy, precision, recall, and F1-Score) compared to Knime Analytics.

# Key Recommendations

Careful consideration should be given to the choice of implementation platform when developing predictive models for cancer diagnosis, as it has implications for model efficacy.

The study recommends using Python SciKit-Learn for enhanced predictive accuracy, especially when employing the Gradient Boosting algorithm.

# Conclusions

Implementation platforms can impact the performance of ML algorithms in cancer classification.

Python SciKit-Learn outperformed Knime Analytics in terms of predictive accuracy, with specific algorithms like Decision Trees (DT) and Gradient Boosting (GB) showing high recall values, which are crucial in cancer diagnosis.

# Future Work



Evaluate the performance of the algorithms on both platforms using other datasets.

Implement on different platforms including R and Weka or test multiple datasets.

Drill down to identify factors responsible for the observed differences by examining the platforms' architecture.

# References

- Ara, S., Das, A., & Dey, A. (2021). Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms. 2021 International Conference on Artificial Intelligence, ICAI 2021, 97–101.
- “*CANCER FACT SHEETS* - Global Cancer Observatory.” Accessed: Feb. 07, 2024. [Online]. Available:<https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet>
- NCI (2021) National Cancer Institute. Available at: <https://www.cancer.gov/about-cancer/> (Accessed: 18 July 2023).
- Zhitnyuk, Y. V., Koval, A. P., Alferov, A. A., Shtykova, Y. A., Mamedov, I. Z., Kushlinskii, N. E., Chudakov, D. M., & Shcherbo, D. S. (2022). Deep cfDNA fragment end profiling enables cancer detection. In *Molecular Cancer* (Vol. 21, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s12943-021-01491-8>



adedayoolowolayemo@gmail.com



amina.souag@canterbury.ac.uk

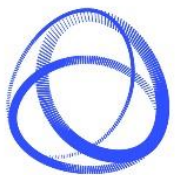


Konstantinos.sirlantzis@canterbury.ac.uk

# Thank you for listening.

Q & A

---



Canterbury  
Christ Church  
University

