ENS 491/492
GRADUATION PROJECT

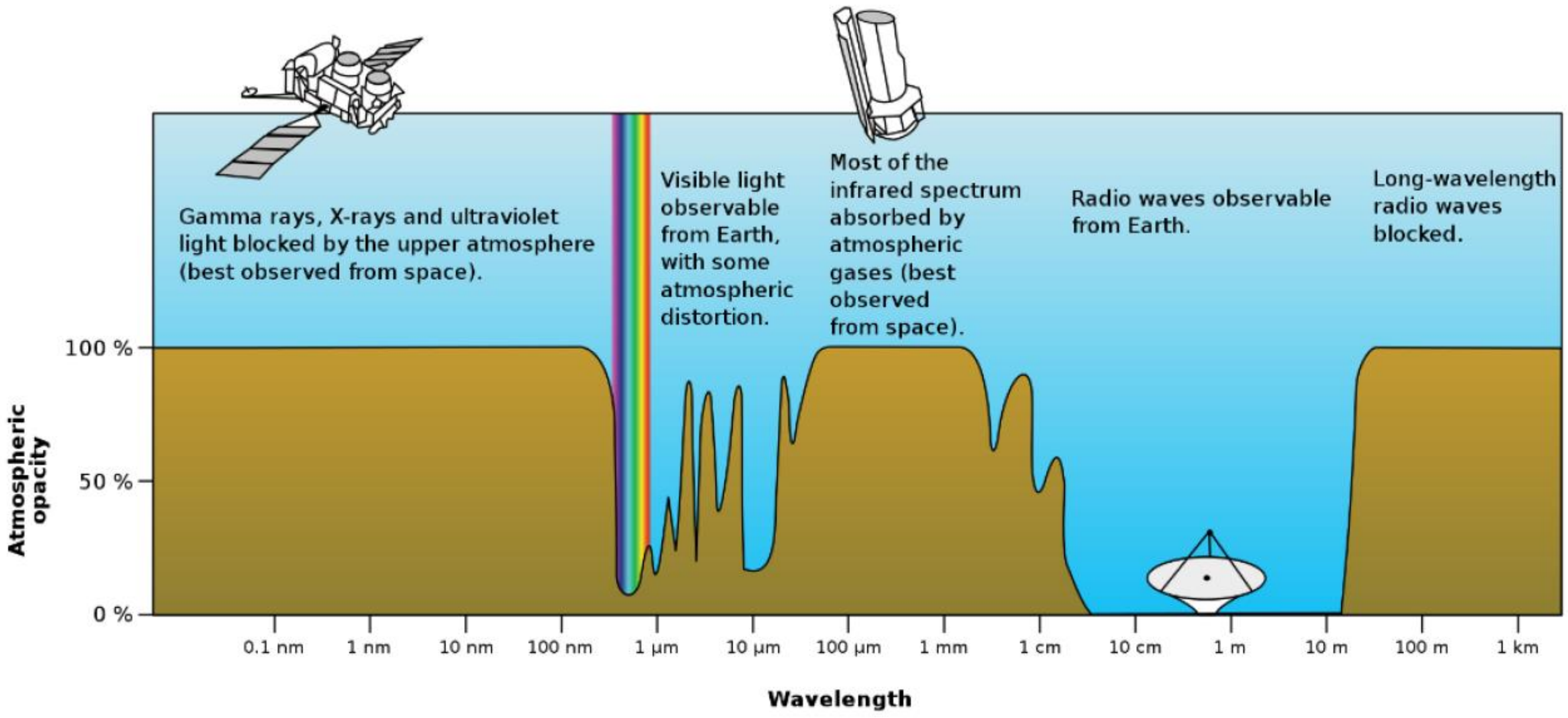
Classification of Background Events in X-ray Imaging Detectors with Machine Learning

Efe Öztaban, 25202
Kayra Bilgin, 25117

Supervised by
Emrah Kalemci
Hüseyin Özkan

Problem Statement

- In astrophysical research, X-ray astronomy is one of the leading fields.
- Difficulties related to the detection instruments and data acquisition systems make this field challenging.
- It is impossible to make X-ray observations from ground.

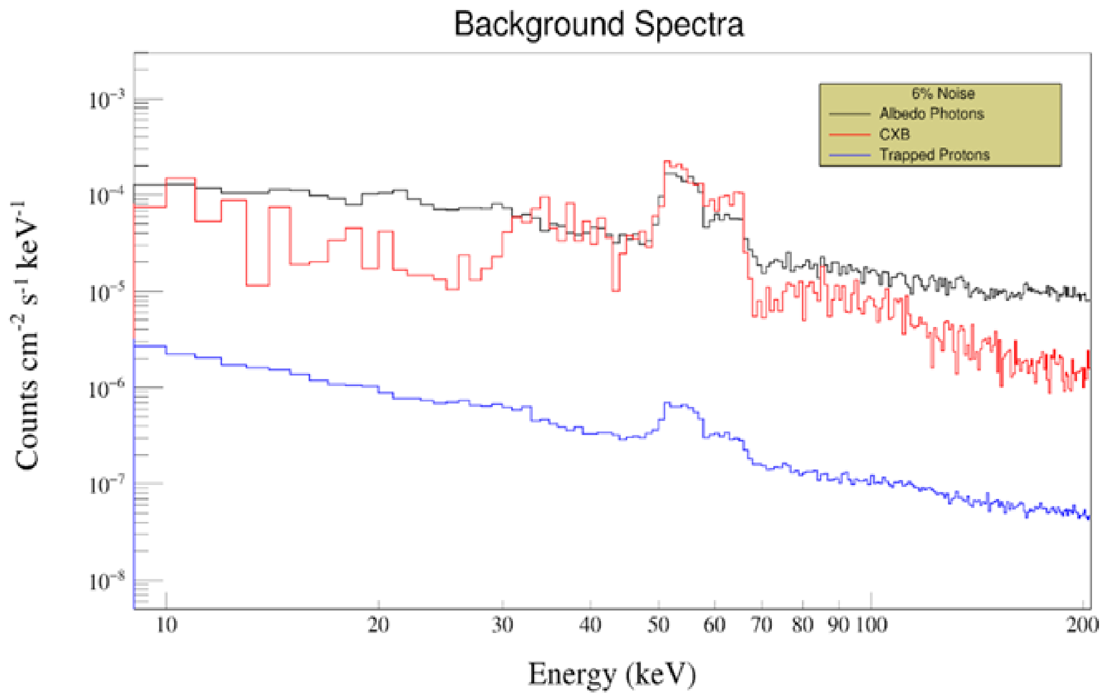


Problem Statement

- X-ray observations very challenging in terms of the cost and complexity of the observation instruments due to the harsh environment of space. One of the major undesirable effects of the space environment is the high background radiation.

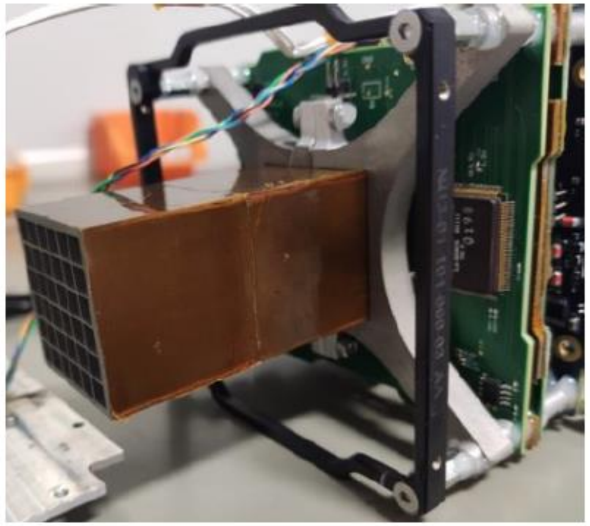
- Main elements which form the background noise are these four components:

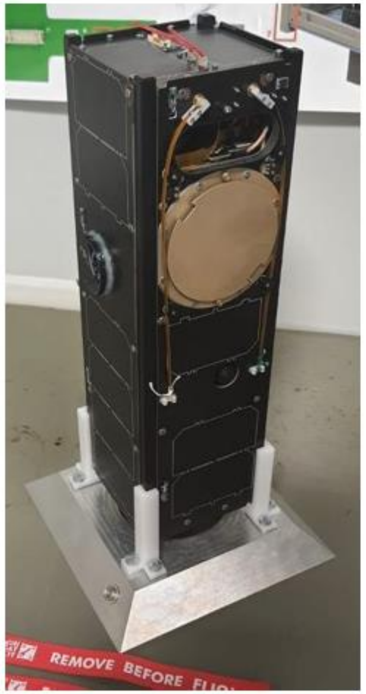
  1. Cosmic X-ray Background Radiation (CXB)
  2. Albedo Photons
  3. Albedo Neutrons
  4. Trapped Particles

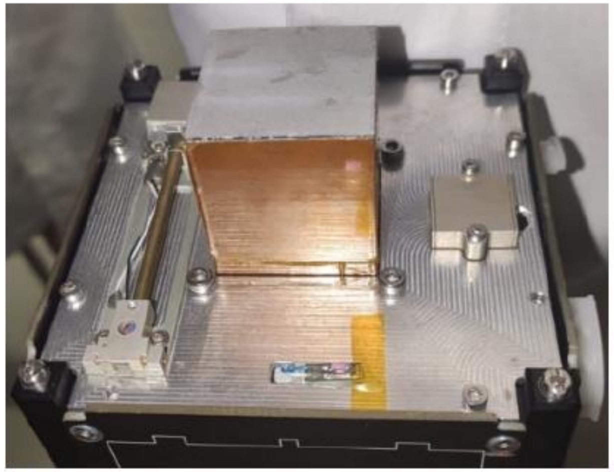


Problem Statement

- In order to reduce the background, mechanical shielding systems such as collimators and computational methods are used in the satellites.







Problem Statement

- In the recent years, X-ray observation missions have become more accessible, easier, and cheaper with the increase in the cubesat missions.

- Because of the size and cost limitations of the cubesat missions, shielding systems which are as complex as the systems used in bigger space missions cannot be used. Having a large background component requires having more effective computational methods for reducing it in order to improve the detection sensitivity.

Objectives and Tasks

1. Searching for a suitable statistical method for classification
    a) Learning the dataset and effect of features
    b) Visulization of the data

2. Design and implementation of statistical methods to classify the background and source data
    a) Learning the effect of resolution of the features

3. Improving the method and prepare it to use in the mission
4. Calculation of the system response and converting into a form to be used

Methodology

- For data preparation:
  - Geant4 simulation data is used (created by Ali Altingün)
  - Data preprocessing is conducted with C++ and Python programming
  - Data visualization is done with Python libraries such as *Matplotlib* and *Seaborn*

- For data classification:
  - Machine learning models are trained and tested with using Python libraries such as *Scikit-learn*

- For response matrix:
  - Calculation of response matrix is done with Python
  - Results are transformed in FITS file with Python libraries such as *AstroPy*

Results
Data Preprocessing

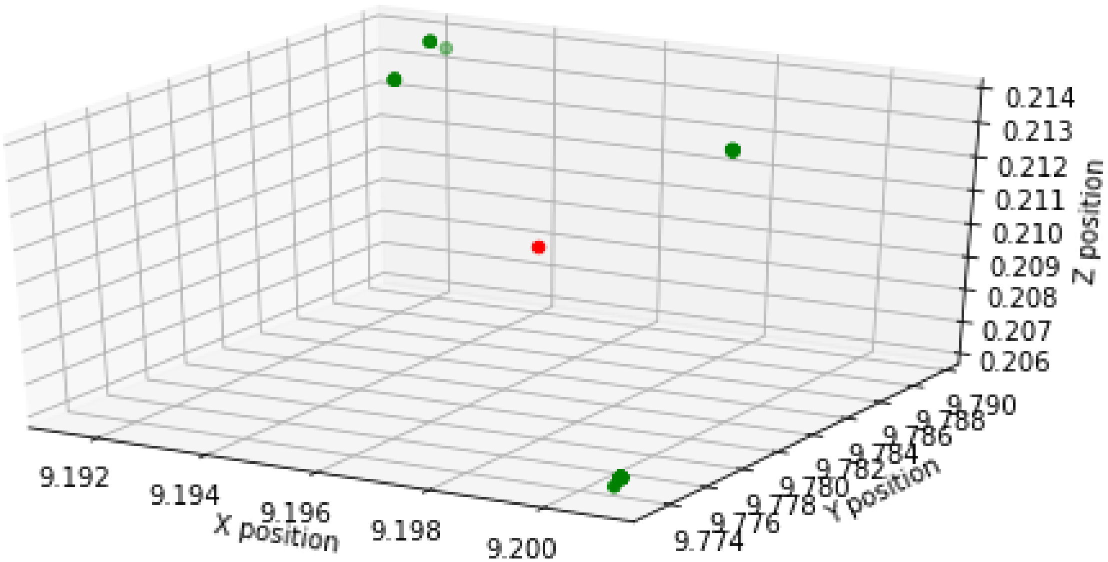
GEANT 4 Data

Simulations:

- Source data (as Crab source)
- Cosmic X-ray Background (CXB) data
- Albedo photons data

Simulation features:

- Electron Cloud Number
- Energy (keV)
- X position (mm)
- Y position (mm)
- Z position (mm)



Electron cloud positions and calculated weighted average event position of a Crab event with 51.9988 keV energy

Results
Data Preprocessing

Preparing the main dataset:

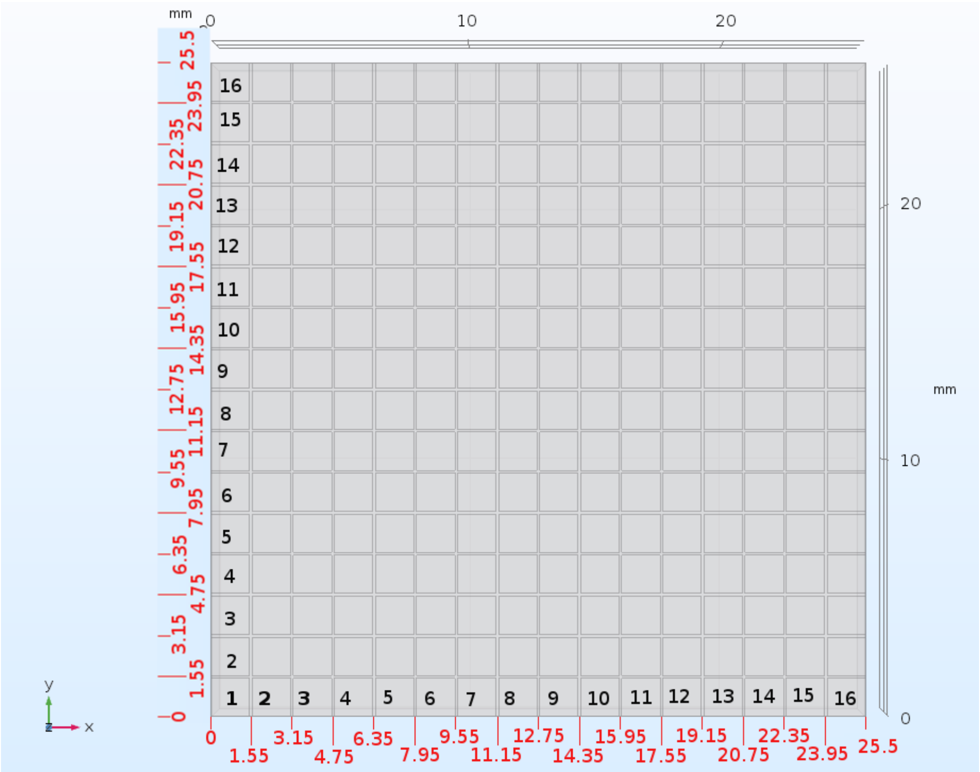
- 1 count per second for Crab source (40.000 events)
- 0.5 count per second for albedo photons (20.000 events)
- 0.4 count per second for CXB (16.000 events)

Different Datasets:

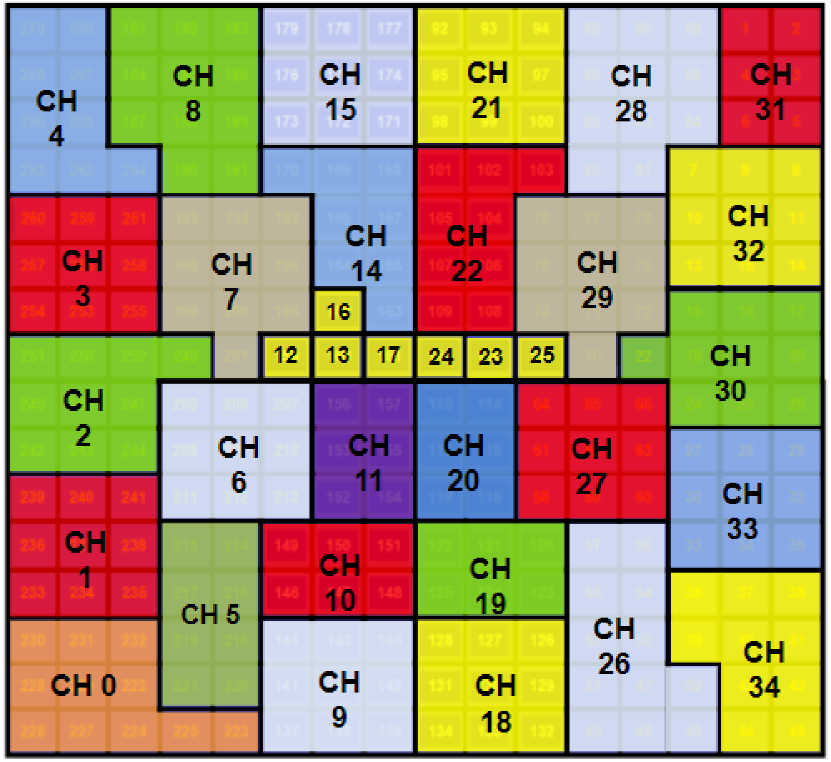
- Dataset from original GEANT4 simulations
- Datasets by changing the precision of the simulation data (for position and energy features)
- Datasets by thresholding energy
- Dataset by using the pixel map to simulate 256 pixels of the detector
- Dataset by using channel map to simulate 35 channels of the detector

Results

Data Preprocessing

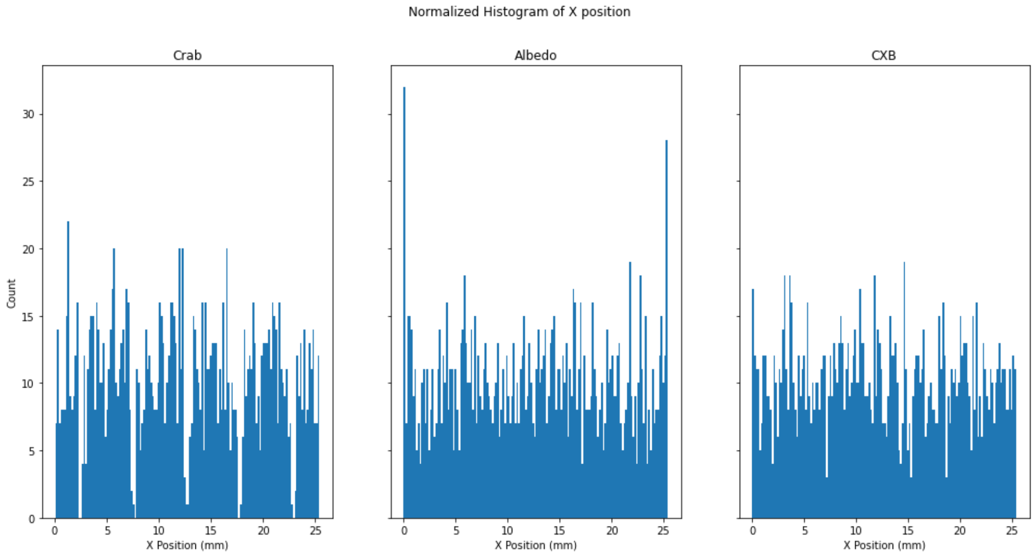


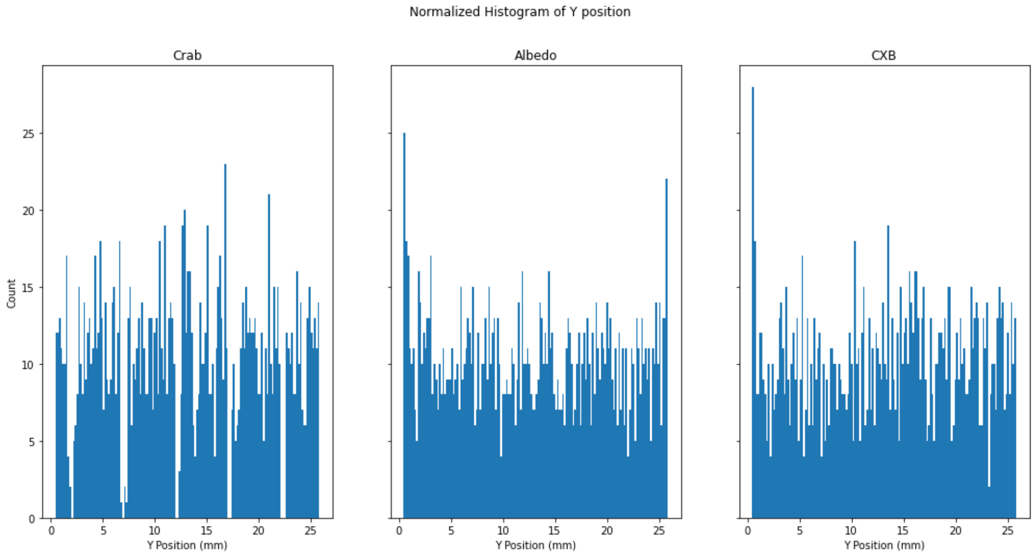
Pixel Map of iXRD

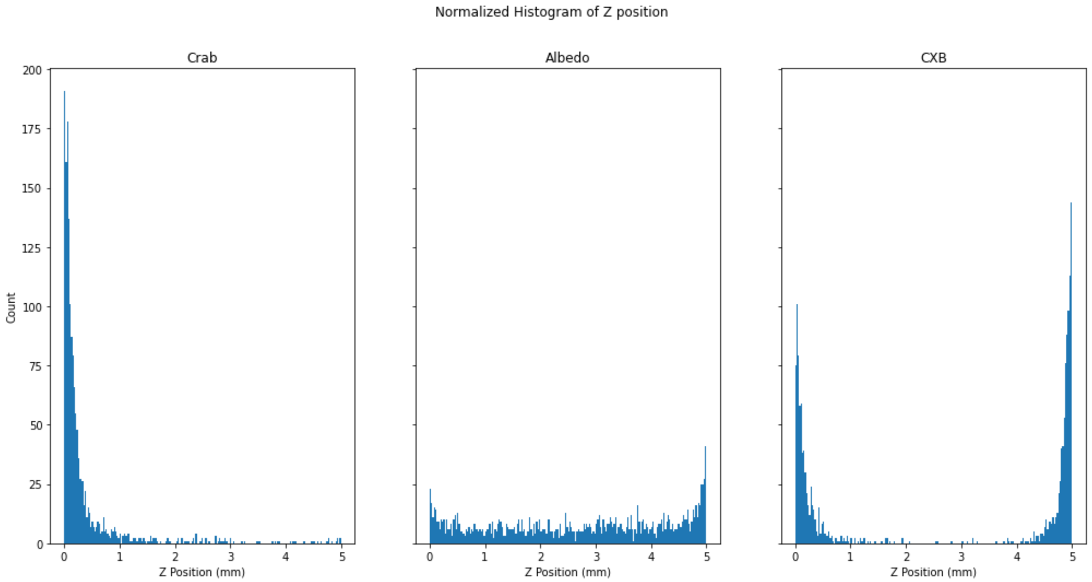


Channel Map of iXRD

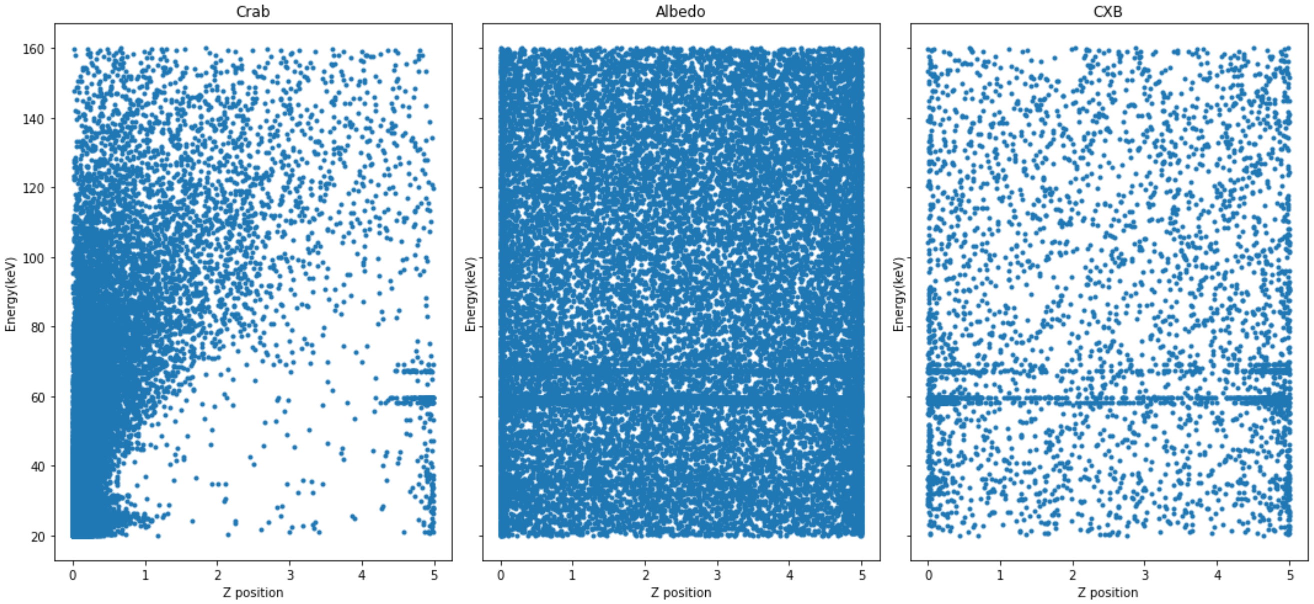
DataVisualization

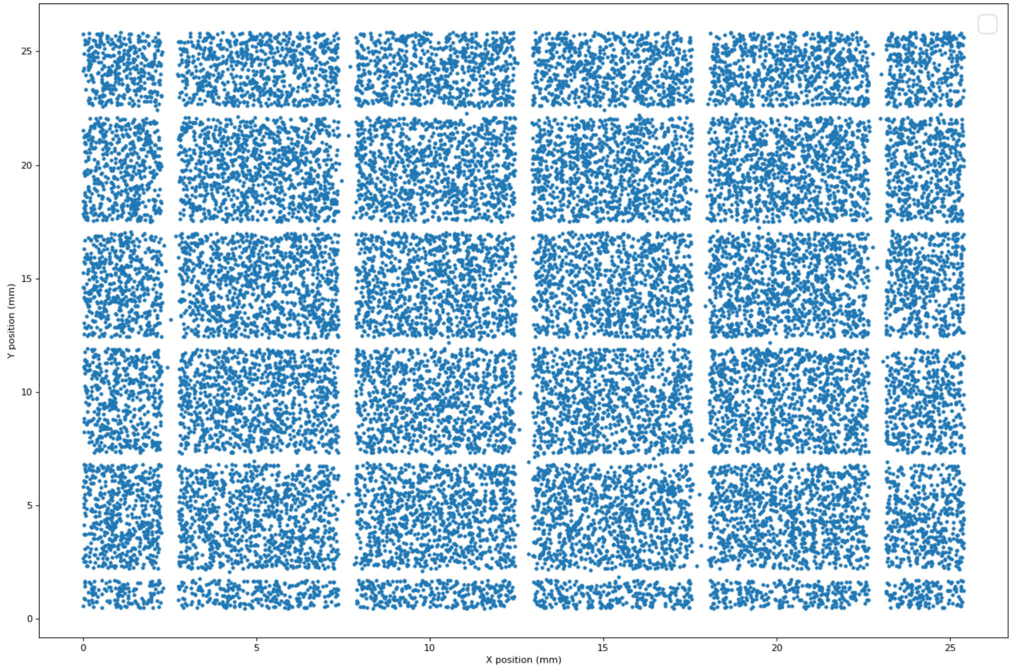




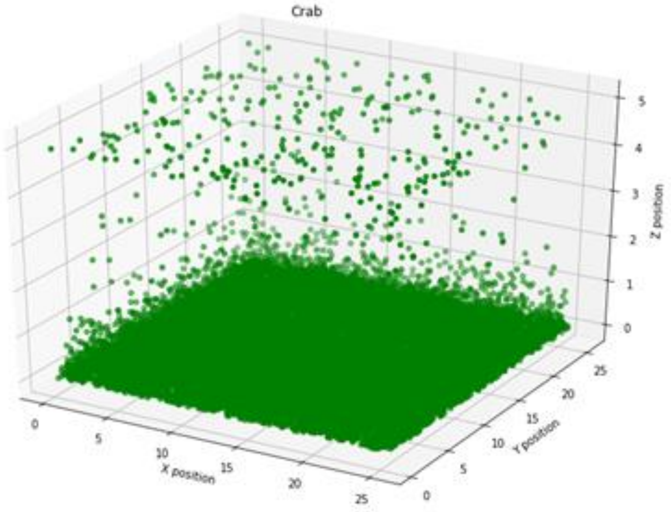


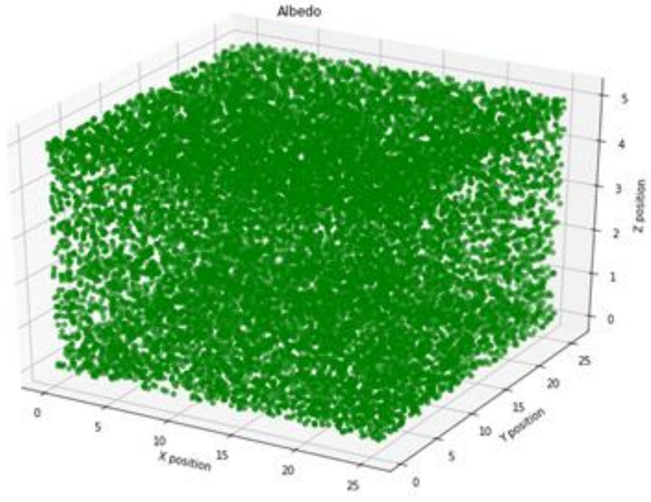
DataVisualization

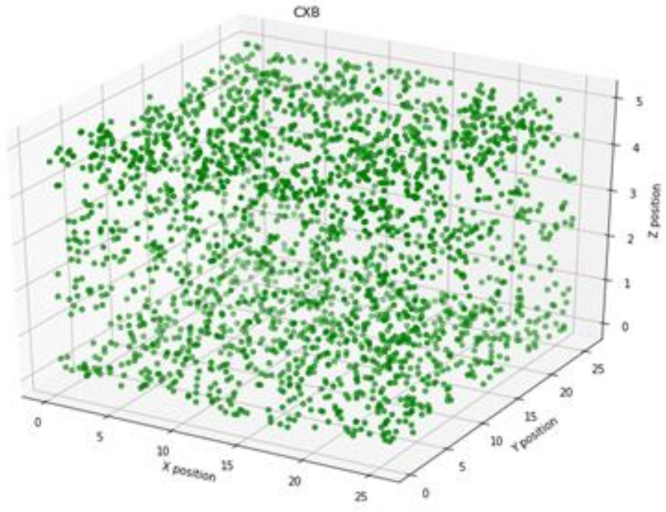




DataVisualization







Results

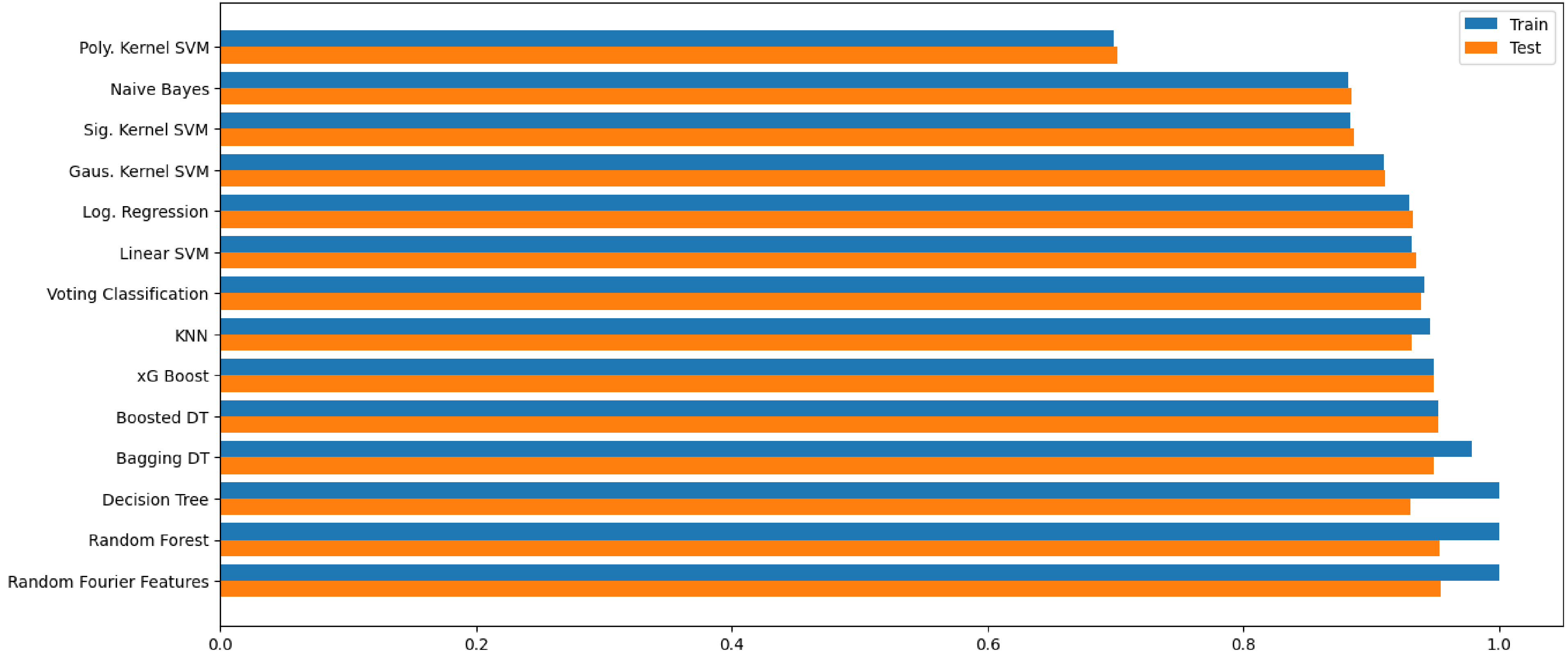
Classification by Machine Learning Models

- Prepared datasets are merged and labelled (source as 1 and background as 0).
- Test and train data is splitted from this dataset.
- Different Machine Learning models are trained and tuned for binary classification.

Models	Train Score	Test Score
Naive Bayes	0.8816958410107923	0.885108903072975
Log. Regression	0.9303270597525665	0.9322234651575969
KNN	0.9462687549355093	0.9320918602355728
Decision Tree	1.0	0.930907415937356
Bagging DT	0.9788595683074494	0.9488714877936435
Boosted DT	0.9521255593577257	0.9521616108442456
Random Forest	1.0	0.9536092649865104
xG Boost	0.9490819952619111	0.9484766730275712
Linear SVM	0.9320380363253488	0.9350529709811147
Poly. Kernel SVM	0.6989503816793893	0.7014542343883661
Gaus. Kernel SVM	0.9103382469070808	0.9105744554846351
Sig. Kernel SVM	0.8834726243748355	0.8865565572152398
Random Fourier Features	1.0	0.9541356846746069
Voting Classification	0.9418926033166622	0.9388037112588011

Results
Classification by Machine Learning Models

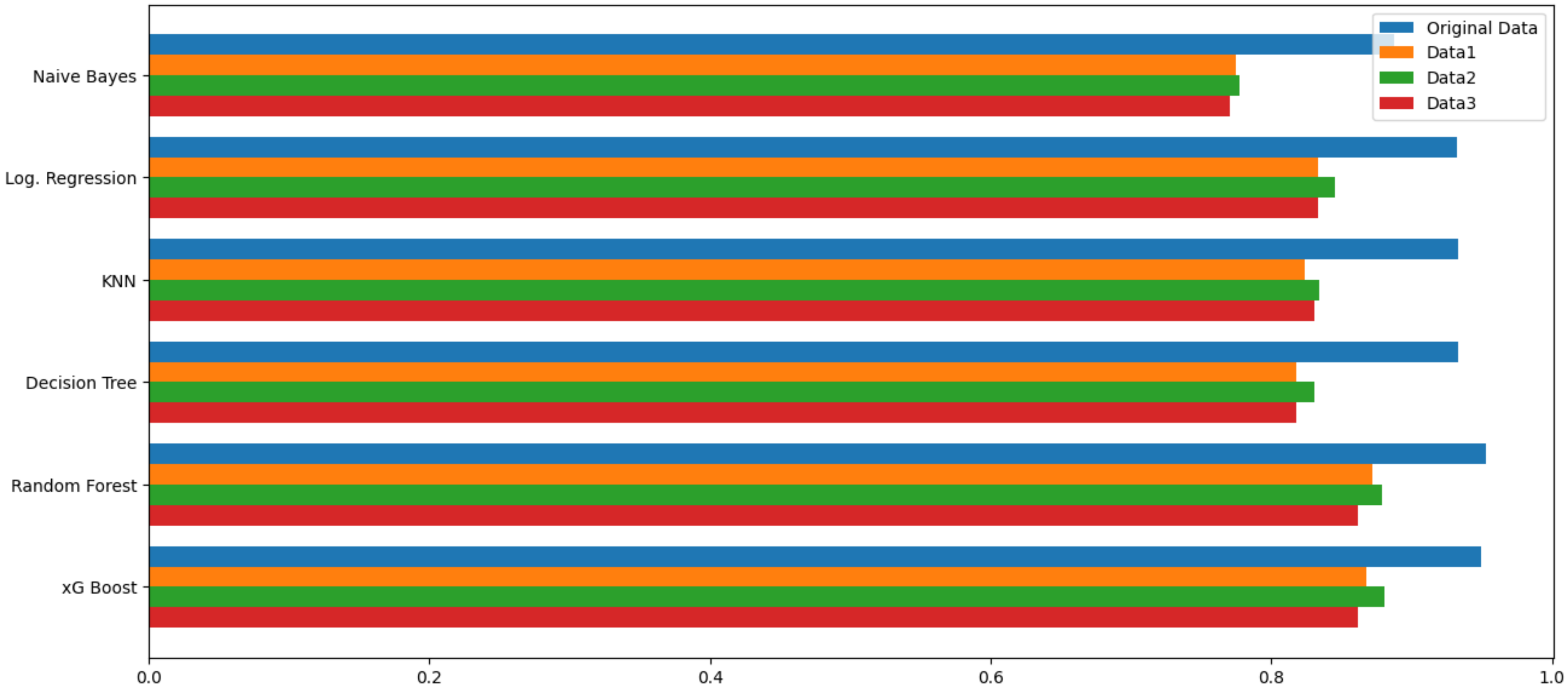
Train and test score comparisons for every ML model trained with the simulation data:



Results

Classification by Machine Learning Models

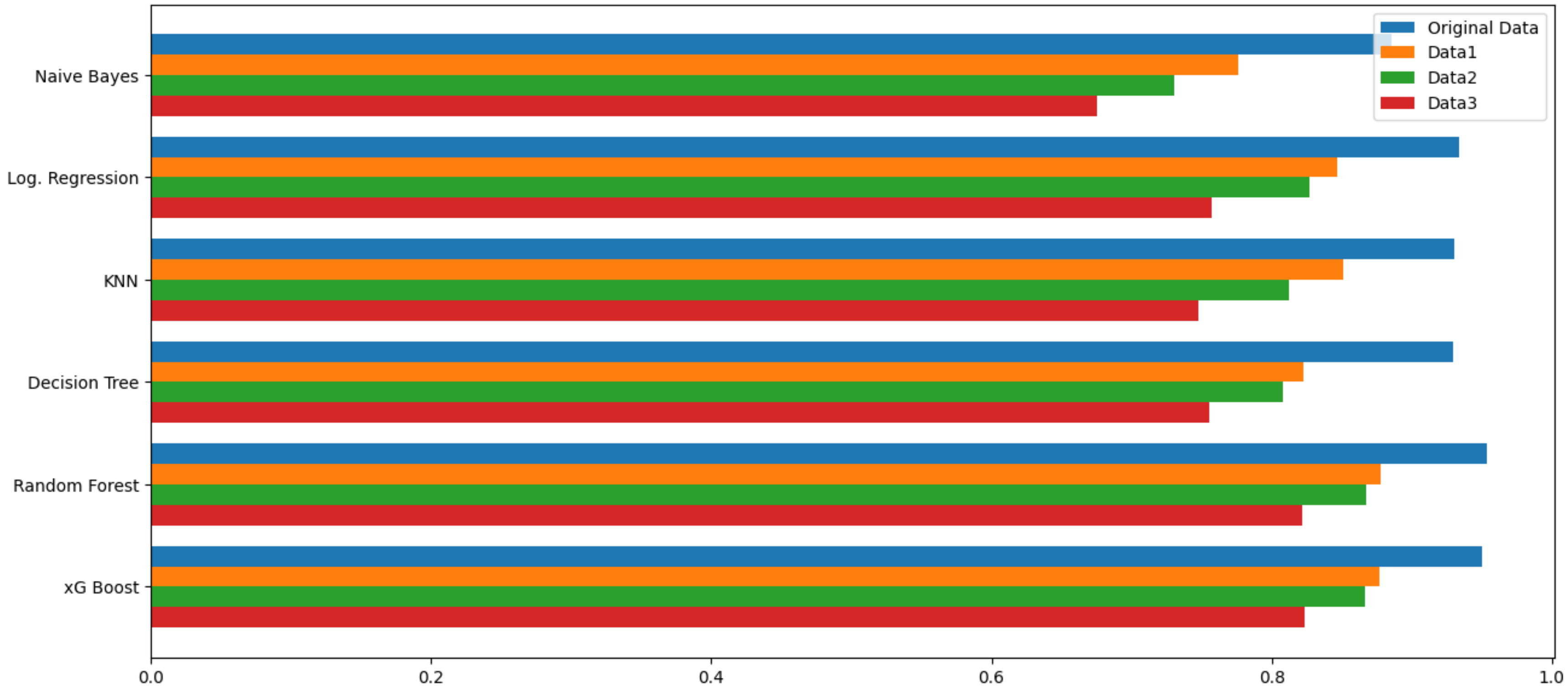
Test score comparisons of each ML model trained with 4 different datasets (with different position precisions):



Results

Classification by Machine Learning Models

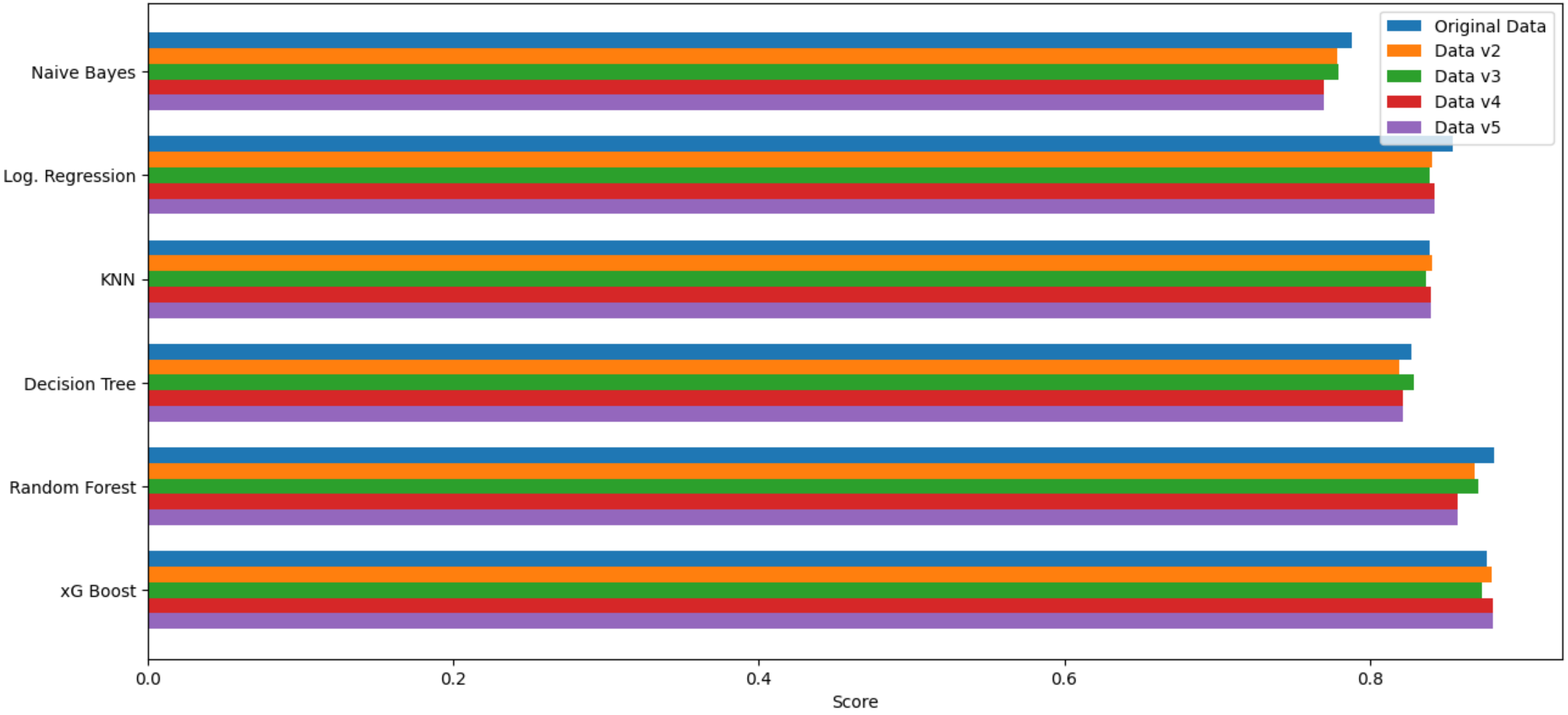
Test score comparisons of each ML model trained with 4 different datasets (with different energy thresholds):



Results

Classification by Machine Learning Models

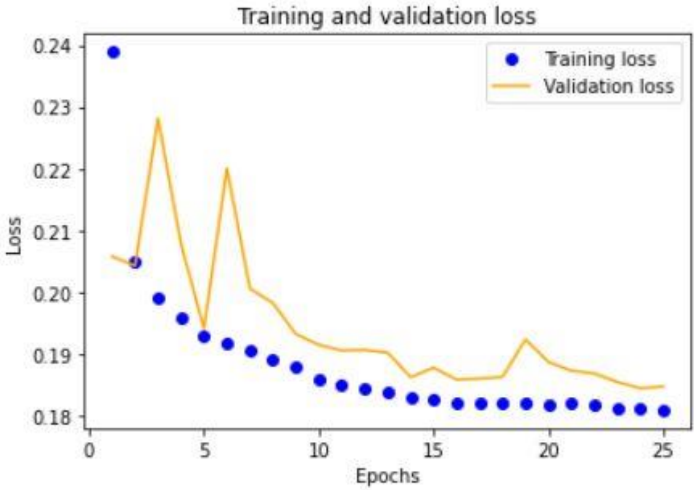
Test score comparisons of each ML model trained with 6 different datasets (with pixel and channel mappings):

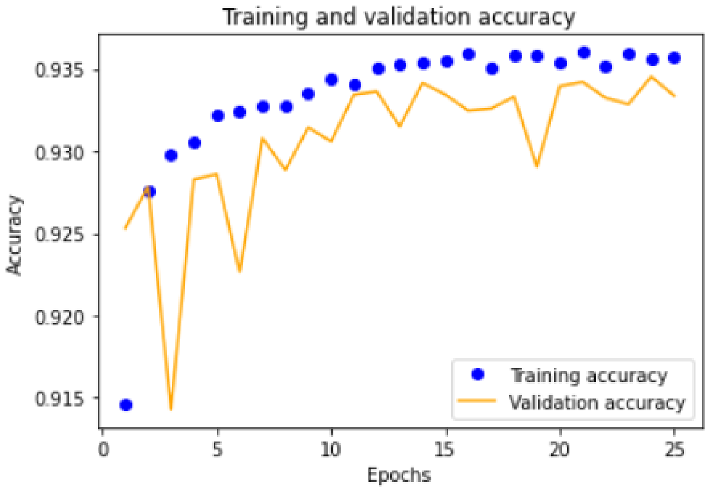


Results

Classification by Machine Learning Models

Also Neural Network is trained for binary classification: (with Binary crossentropy loss function)

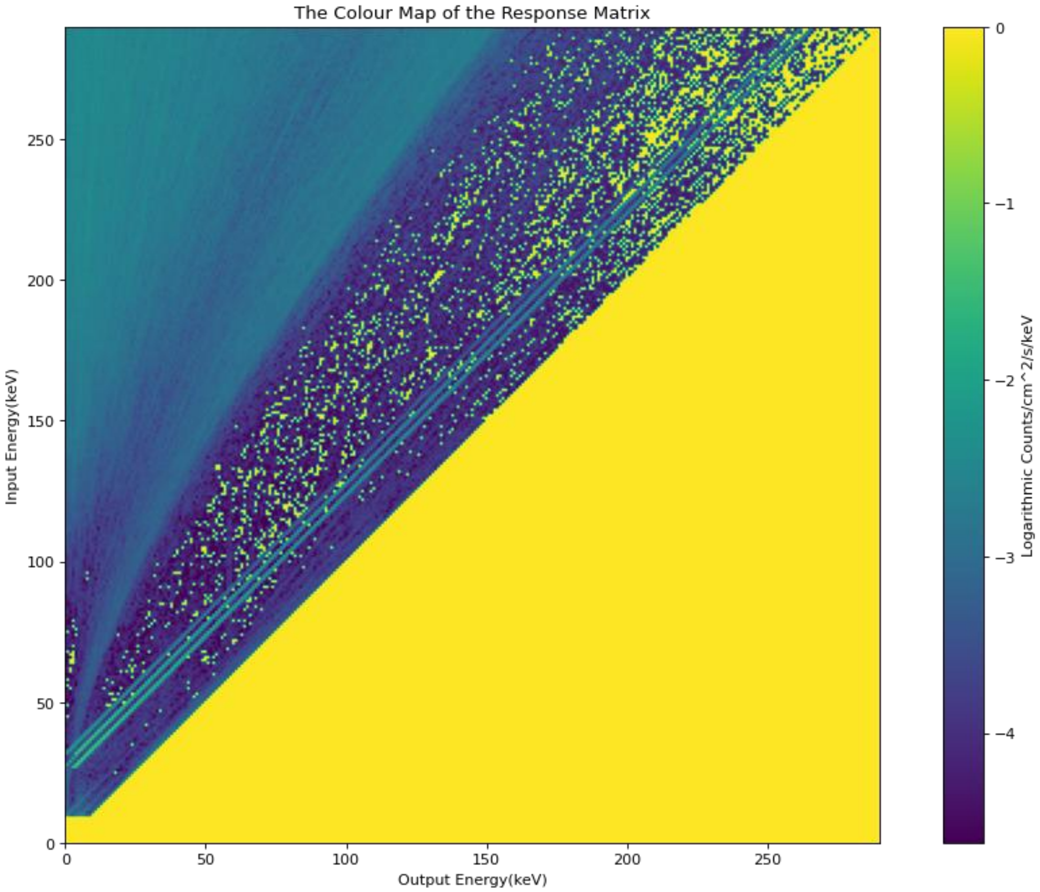




	precision	recall	f1-score
0	0.94	0.92	0.93
1	0.93	0.95	0.94
accuracy			0.94
macro avg	0.94	0.93	0.94
weighted avg	0.94	0.94	0.94

Results

Response Matrix




```
Filename: xrt_pc20.arf
No.    Name      Ver    Type      Cards   Dimensions   Format
  0  PRIMARY       1 PrimaryHDU      11   ()
  1  SPECRESP      1 BinTableHDU     69   2400R x 3C   [1E, 1E, 1E]
```

```
Filename: swxpc0to12s6_20090101v014.rmf
No.    Name      Ver    Type      Cards   Dimensions   Format
  0  PRIMARY       1 PrimaryHDU      10   ()
  1  MATRIX        1 BinTableHDU    101   2400R x 6C   [E, E, I, I, I, 1024E]
```

Results and Discussion

- Different Machine Learning models are tried for classification task.
- These models are tried with different datasets.
- The effect of energy and position resolution of detector on the classification accuracy is investigated.
- The effect of pixel and channel mappings on the classification accuracy is investigated.
- Even in low resolution cases, machine learning models are found efficient to classify events.

- The response matrix of iXRD was calculated and plotted with the simulation data.
- It will be converted into FITS format to form RMF and ARF files to be used in standard fitting programs such as ISIS and XSPEC.

Results and Discussion

- As future work:
  - Studies on the ML models and comparisons on the different datasets can be continued.
  - The trained models can be prepared to be used in Sharjah Sat-1 mission.

References

[1] Bishop, C. M. (2016). *Pattern recognition and machine learning*. Springer.

[2] Blanco, P. R., the HEXTE team (1997). Appendix F: The XTE Technical Appendix, HEXTE Feasibility of Observations.

[3] Boldt, E. (1987). The cosmic X-ray background. Physics Reports, 146(4), 215–257. https://doi.org/10.1016/0370-1573(87)90108-6

[4] Butler, R. C., Dean, A. J., Dipper, N. A., & Ramsden, D. (1978). An actively shielded hard X-ray telescope. Nuclear Instruments and Methods, 154(1), 117–119. https://doi.org/10.1016/0029-554x(78)90669-9

[5] Dean, A. J., & Dipper, N. A. (1981). An evaluation of the background effects in actively shielded hard X-ray telescopes at balloon altitudes. Monthly Notices of the Royal Astronomical Society, 194(2), 219–227. https://doi.org/10.1093/mnras/194.2.219

[6] Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and Artificial Neural Network Classification models: A methodology review. *Journal of Biomedical Informatics*, *35*(5-6), 352–359. https://doi.org/10.1016/s1532-0464(03)00034-0

[7] Kalemci, E., Aslan, A. R., Bozkurt, A., Baş, M. E., Diba, M., Veziroğlu, K., ... & Ümit, M. E. (2019, June). X-Ray Detector XRD on BeEagleSat and the Development of the Improved X-Ray Detector iXRD. In 2019 9th International Conference on Recent Advances in Space Technologies (RAST) (pp. 565-570). IEEE.

[8] Kalemci, E., Manousakis, A., Fernini, I., Naimiy, H., Bozkurt, A., Aslan, A., Altingun, A. Veziroglu, K., Yalçın, R., Gökalp, K., Diba, M., Yaşar, A., Oztaban, E., ... & Madara, S. (2021). Scientific Contribution of Sharjah-Sat-1 to X-ray Observations. In 2021 72nd International Astronautical Congress.

[9] Kalemci, E. (2018). Summary of the past, present and future of the X-ray astronomy. The European Physical Journal Plus, 133(10), 407.

[10] Kalemci, E. (2020). Uydularda Kullanılabilecek Özgün Yüksek Enerji Algılayıcı Sistemlerinin Milli İmkanlarla Geliştirilmesi.

[11] Knoll, G. F. (2010). Radiation detection and measurement. John Wiley & Sons.

[12] Likas, A., Vlassis, N., & J. Verbeek, J. (2003). The global K-means clustering algorithm. *Pattern Recognition*, *36*(2), 451–461. https://doi.org/10.1016/s0031-3203(02)00060-2

[13] Shkolnik, E. L. (2018). On the verge of an astronomy CubeSat Revolution. Nature Astronomy, 2(5), 374–378. https://doi.org/10.1038/s41550-018-0438-8

[14] Toorian, A., Diaz, K., & Lee, S. (2008). The CubeSat approach to Space Access. 2008 IEEE Aerospace Conference. https://doi.org/10.1109/aero.2008.4526293

[15] Wilkins, D., Allen, S., Miller, E. D., Bautz, M., Chattopadhyay, T., Fort, S., Grant, C. E., Herrmann, S., Kraft, R., Morris, G., & Nulsen, P. (2020). Identifying charged particle background events in X-ray imaging detectors with novel machine learning algorithms. Space Telescopes and Instrumentation 2020: Ultraviolet to Gamma Ray. https://doi.org/10.1117/12.2562354

[16] Woellert, K., Ehrenfreund, P., Ricco, A. J., & Hertzfeld, H. (2011). Cubesats: Cost-effective science and technology platforms for emerging and developing nations. Advances in Space Research, 47(4), 663–684. https://doi.org/10.1016/j.asr.2010.10.009