ENS 491-492 – Graduation Project

Final Report

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

Group Members:

Efe Öztaban

Kayra Bilgin

Supervisor(s):

Emrah Kalemci

Hüseyin Özkan

Date: 12.06.2022

1. EXECUTIVE SUMMARY

In astrophysical research, X-ray astronomy is one of the leading fields. X-ray astronomy allows scientists to collect a very wide range of data which cannot be created in lab environment on Earth. Collected data is used for research about many astronomical objects and phenomena from black holes to galaxy clusters. Because of that X-ray astronomy has a very essential effect on the different fields of astrophysics. Major contributions of the X-rays to astrophysical studies make X-ray astronomy and observations very important. However, difficulties related to the detection instruments and data acquisition systems make this field challenging.

Earth's atmosphere does not allow photons with wavelengths which are below 300-400 nm to pass. Therefore, it is impossible to make X-ray observations from ground. In addition to that, reflection of X-ray photons is very hard due to their high energy. That requires a larger surface area and focal length for X-ray detection telescopes. Because of these limitations, X-ray observations must be made in space using complicated detection instruments. This necessity makes X-ray observations very challenging in terms of the cost and complexity of the observation instruments due to the harsh environment of space [9] [10].

One of the major undesirable effects of the space environment is the high background radiation. The main elements which form the background noise are these four components:

1. Cosmic X-ray Background Radiation (CXB)

Cosmic X-ray Background Radiation is caused by the cosmic photons coming from outer space.

1. Albedo Photons

Albedo photons are formed because of the interaction between the cosmic rays and atoms or molecules in the Earth's atmosphere. These reflected photons cause a background mostly in the low Earth orbits [10].

2. Albedo Neutrons

Albedo neutrons are also created by the interactions of cosmic rays and the particles in the Earth's atmosphere [10].

3. Trapped Particles

Trapped particles are formed because of the particles (protons, electrons and ions) which come from the Sun or outer space. These particles interact with the magnetic field of the Earth and are trapped in the region around the magnetic fields. These trapped particles include trapped protons and trapped electrons. They are known as the Van Allen Belt particles as well [10].

In order to reduce the background, mechanical shielding systems such as collimators and computational methods are used in the satellites [4] [5] [7].

In recent years, X-ray observation missions have become more accessible, easier, and cheaper with the increase in the cubesat missions. Cubesats are miniature satellites which are build in standard dimensions (from 10cm x 10 cm x 10cm cubes). Even if the mission duration and mission capabilities are smaller in cubesat missions, low risk, low costs and high experimentality makes cubesat missions more advantageous [13] [14] [16]. In spite of all the advantages of cubesat missions, detector systems are affected by the space environment more due to their limitations. 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. As an example, for background radiation and sensitivity, simulated background, and sensitivity of the iXRD on Sharjah-Sat-1 are given in Figs. 1. and 2.

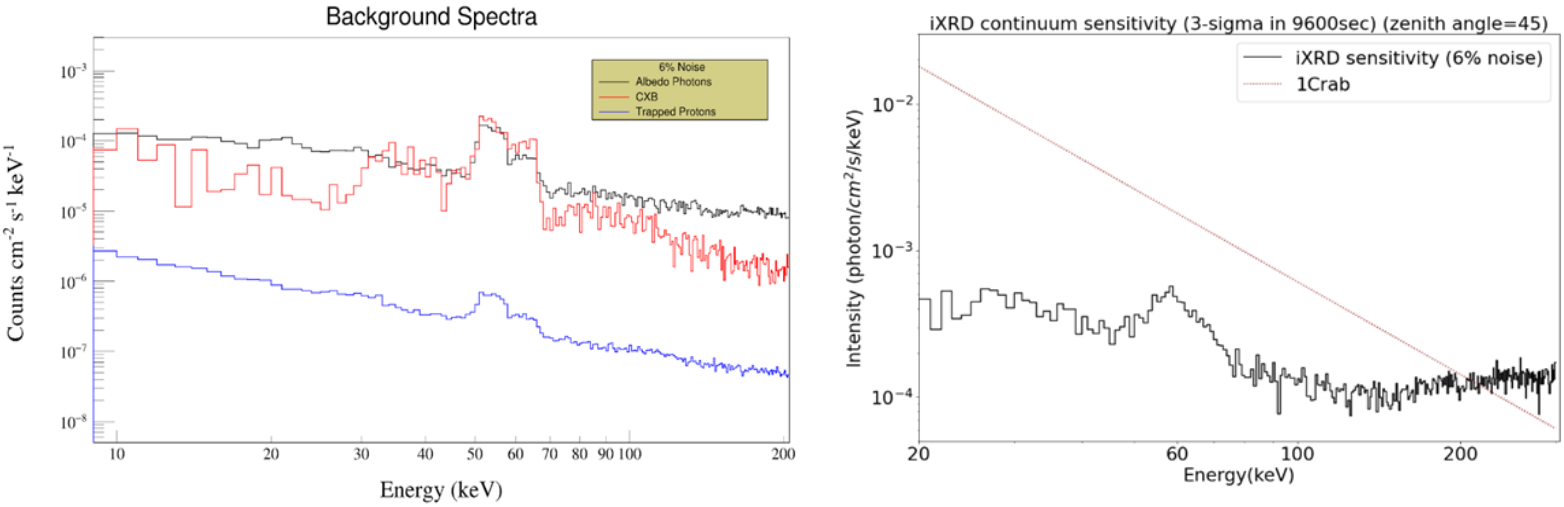


Figure 1 and 2. Spectra of the background components of the iXRD (right), sensitivity curve of the iXRD (left)[8] .

In this project, firstly background components are examined. Then computational tools are developed for classification of the background and source events with machine learning algorithms.

2. PROBLEM STATEMENT

The main purpose of this project is developing computational methods to improve the sensitivity and tools to aid spectral fitting to be used in iXRD (Improved X-ray Detector) on the Sharjah-Sat-1 cubesat mission. It is aimed to identify and classify the background data of the iXRD with these computational tools in order to improve the detection sensitivity. Statistical methods and machine learning techniques/models are intended to be used for clustering and classification of the data.

Novel machine learning models and algorithms based on convolutional neural networks were previously developed to be used for background classifications in space missions such as Athena, Lynx and AXIS. However, in this project these approaches are aimed to be applied to a cubesat mission which is more critical due to shielding limitations and detection limitations (such as detected energy range and triggered event location on the detector). [15]

2.1. Objectives/Tasks

General objectives of the project and intended results of each objective are listed below:

a. Searching for a suitable statistical method for classification:

Firstly, simulation data will be studied in order to have a better understanding of the features of the source and background components of the data. For that purpose, data obtained from the GEANT4 simulations will be used. The obtained data from the GEANT4 will be processed to convert to the related format and form the expected type of data. Then this data will be visualized and studied to observe the features of each component of the data. With these steps, special features of the background and source components of the data will be determined. These features will be used in the statistical methods for classification.

b. Design and implementation of statistical methods to classify the background and source data:

With the determination of the special features of the different components of the data and suitable statistical method, implementation of the statistical method will be started. Data obtained from the GEANT4 simulations will be used to train and test the implemented statistical method-based machine learning model to classify the detector's inputs as source data or different background effects. With the optimization of the implemented statistical model, classification of the data into background and source components will be done efficiently.

c. Calculation of the sensitivity curve of the detector with all factors:

With the calculations made before, the sensitivity curve of the detector will be calculated again. Because the cleaner version of source data is achieved by adding detailed parameters (such as system response and electrical noise), the resulting sensitivity curve will be closer to the real sensitivity.

d. Calculation of the system response and converting into a form to be used:

With the GEANT4 simulation results, system response will be calculated. After the calculation of the response matrix of the system, results will be converted into a specific format to be used in a tool which will reduce the effect of system response on the result data. More specifically, OGIP library, which is developed by HEASARC, NASA will be used during the format conversion process. Thus, the resulting system response file will be converted into FITS format to be used in standard fitting programs such as ISIS and XSPEC. This will make it easier to use developed tools in the project and share the files with the community in a known format.

e. Developing and forming a toolbox for processing of iXRD data:

The tools developed and used during the project will be gathered to obtain a detailed toolbox. With this toolbox, real data from iXRD on the Sharjah Sat-1 will be processed to obtain data which is closest to the real source data.

2.2. Realistic Constraints

In this project the data which is the training data generated by the GEANT4 simulation is used. For this reason, the significance of the data is quite high in terms of the reliability of the results. Machine learning models were trained on the data to be obtained from the simulation. For this, the size of the training data is useful for the model to be created to classify the inputs, since as the sample size increases, the reliability of the model also increases. However, the amount of real data to be received from the satellite, which is going to be launched into space, will be limited. Since limited data will decrease the accuracy of the model, simulation data is very crucial. Therefore, the realistic constraint of this project is different aspects related to the simulation data.

3. METHODOLOGY

a. GEANT4 simulation data:

Firstly, data received from the GEANT4 simulations have been studied. These simulations are conducted for different components of the photon sources. These components are:

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

For each data set from the simulations, special features exist in the data. These data sets consist of every electron cloud formed by the events on the detector. For every electron cloud these features exist in the data:

i. Electron Cloud number

This stands for the electron cloud number for every event. Electron clouds formed by every event are numbered from 0 to the last e-cloud number. With the electron cloud number, event number can be generated.

ii. X position (in mm)

This stands for the position of the electron cloud in the detector in the x axis. This data will be formed by the triggered pixel locations on the detector.

iii. Y position (in mm)

This stands for the position of the electron cloud in the detector in the y axis. This data will be formed by the triggered pixel locations on the detector.

iv. Z position (in mm)

This stands for the position of the electron cloud in the detector in the z axis. This data will be formed by the cathode/anode ratio in the real case.

v. Energy (in keV)

This stands for the energy of the electron cloud formed in the detector.

b. Pre-processing of the raw GEANT4 simulation data:

After the simulation data is received, this data is converted from binary format into a related format to be processed and visualized. After the conversion process is done, event features are created by using every electron cloud of each event.

For every event, x position, y position, z position and energy information are calculated from every electron cloud of the event. x, y and z positions of the events are calculated as the weighted average of the x, y and z positions of every electron clouds of that event. Weights are taken as the energies of each electron cloud and weighted average of the x, y and z position of the electron clouds are calculated. In addition, the energy of each event is calculated by the sum of the energies of every electron cloud of that event. In fig. 3 below, all the electron clouds of one event are shown with green dots. Also, the calculated weighted average of the electron clouds is shown with a red dot.

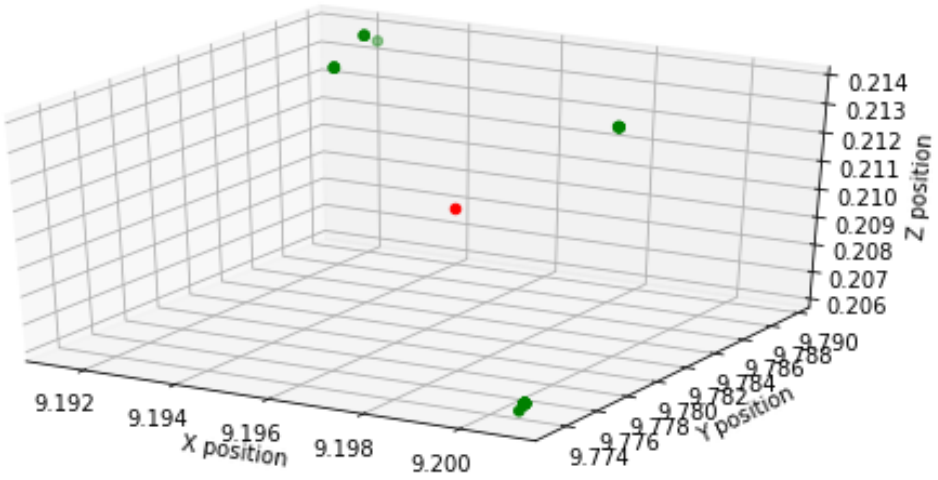


Figure 3. Electron cloud positions and calculated weighted average event position of a Crab event with 51.9988 keV energy.

After the processed data is prepared, datasets are prepared for visualization and analysis of the data. In order to prepare these datasets, some processes are applied on the data. Firstly, data is labelled as source (1) and background (0) data. Events from crab data are labelled as source data. Events from albedo photons and CXB are labelled as background data. After the labelling process is done, data is filtered by the energy range. Because of the detection limits of the detector the events in the 45-800 keV energy range are selected. Then, a dataset is formed by merging the different types of events which are Crab, albedo and CXB. In this merging process data is taken from different sources in different ratios. These ratios are determined by the estimated count rates of events from

different sources on the X-ray detector. The count rates for different sources:

- 1 count per second for Crab source
- 0.5 count per second for albedo photons
- 0.4 count per second for CXB

Using these ratios 40.000 Crab events, 20.000 albedo events and 16.000 CXB events are merged to form a dataset with a total 76.000 events.

After the main dataset is formed using the GEANT4 simulation results directly, different datasets are also formed in order to analyse the data. For these datasets, precision of the is decreased by rounding the x, y and z positions to 5 decimal places, 3 decimal places and integers. In addition to these datasets, 2 different datasets are formed in order to simulate the detector pixels and readout system on the data by discretizing the data. Firstly, using the pixel map below (in fig. 4) x and y positions of the events are changed with the centre points of the 256 pixels. Also, z position is discretized in the range of (0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5).

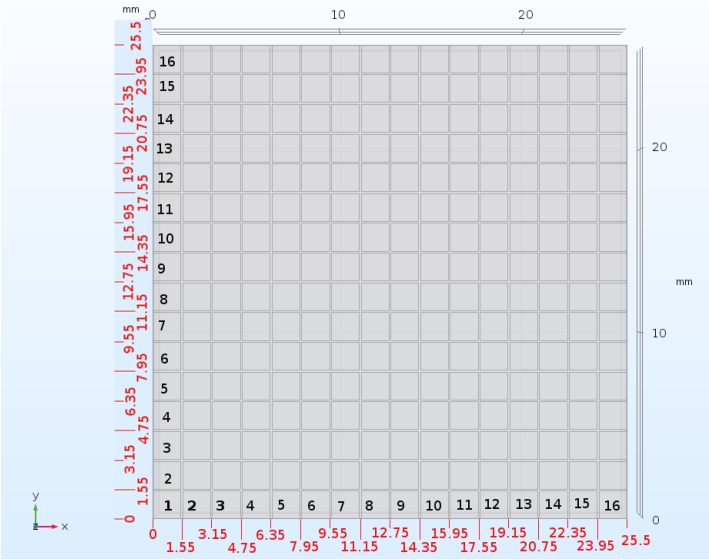


Figure 4. Pixel map (x and y positions) of the iXRD.

Secondly, these pixel locations are grouped. Because of the channel limit of the detector, only 35 channels are used to read data from the detector. Therefore, pixels are grouped

using the channel map below (in fig. 5) to simulate the channel readout system on the data, and dataset is formed with this information.

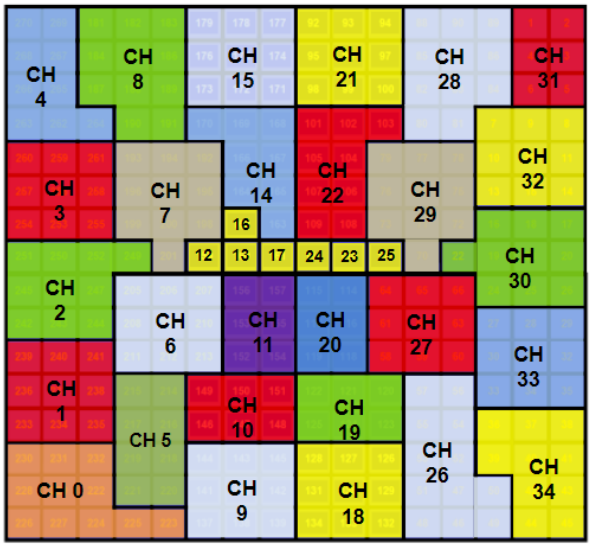


Figure 5. Channel map of the iXRD.

In brief, 6 different datasets are formed:

- 1 dataset from original GEANT4 simulations
- 3 datasets by changing the precision of the simulation data
- 1 dataset by using the pixel map to simulate 256 pixels of the detector
- 1 dataset by using channel map to simulate 35 channels of the detector

c. Visualization of the source and background data:

Charts are plotted to compare the data and to have a better understanding of the features of different sources. Phyton's libraries such as *Matplotlib* and *Seaborn*, were used for this visualization process. These charts consist of different types of visualization such as 3d scatter, 2d scatter and histogram.

Histograms show the number of photons in the different ranges of the x, y and z positions. The histograms below (in figs. 6,7 and 8) are plotted in the energy range of 20-200 keV. Looking at fig. 3, the z-position values in the source data are concentrated between 0 and 1. However, this interpretation is not suitable for Albedo and CXB data since concentration of the z-position value of Albedo and CXB is between 0 and 5. In addition, the concentration increases between 4 and 5. This is because of the source location of the Crab events. Crab events are coming from the upper edge (0 mm position) of the detector. However, for figure 2 and 1, x and y position are not a significant feature

to distinguish the Crab and background events. Still, some decreases in specific locations (both in x and y position) are seen on the Crab data. This is because of the grid shape of the collimator which is located on the upper edge of the detector.

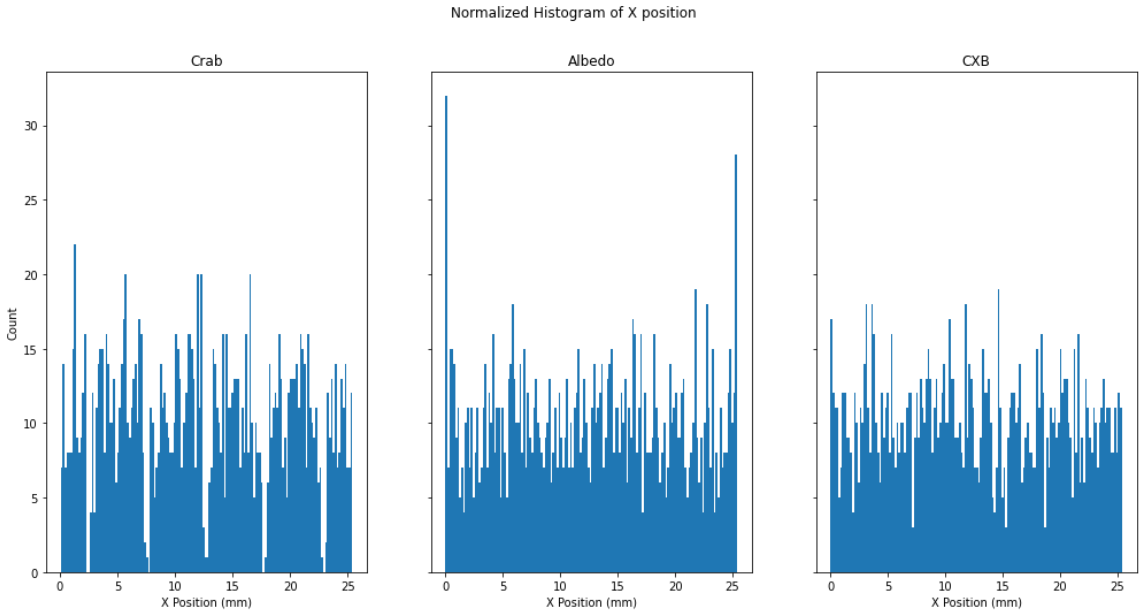


Figure 6. X position histograms for Crab, Albedo and CXB events.

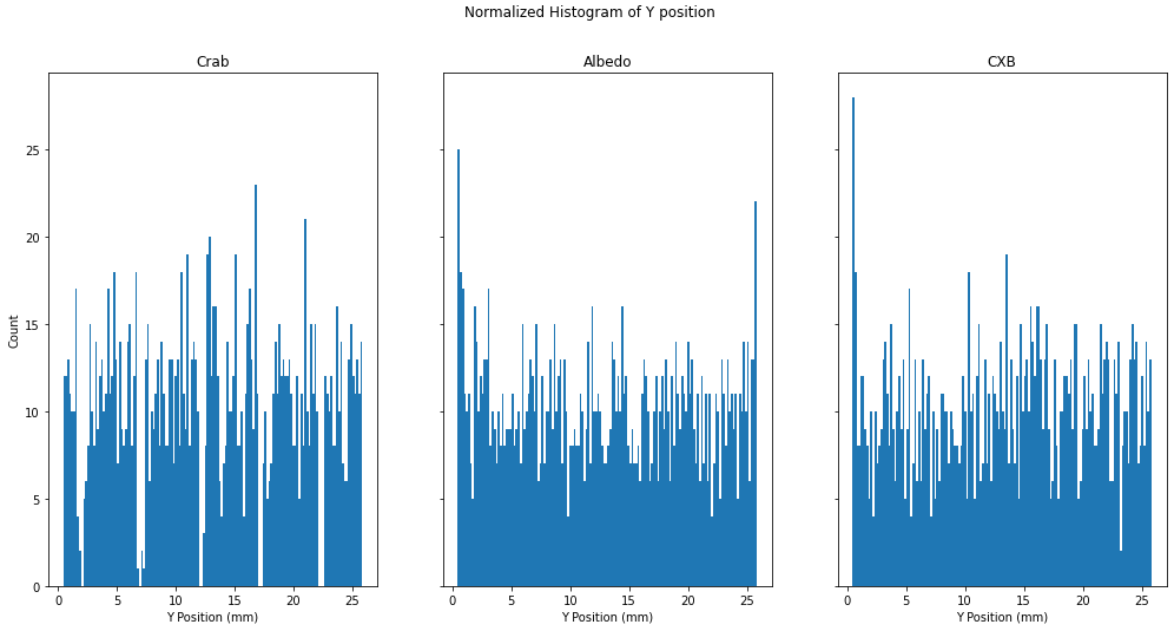


Figure 7. Y position histograms for Crab, Albedo and CXB events.

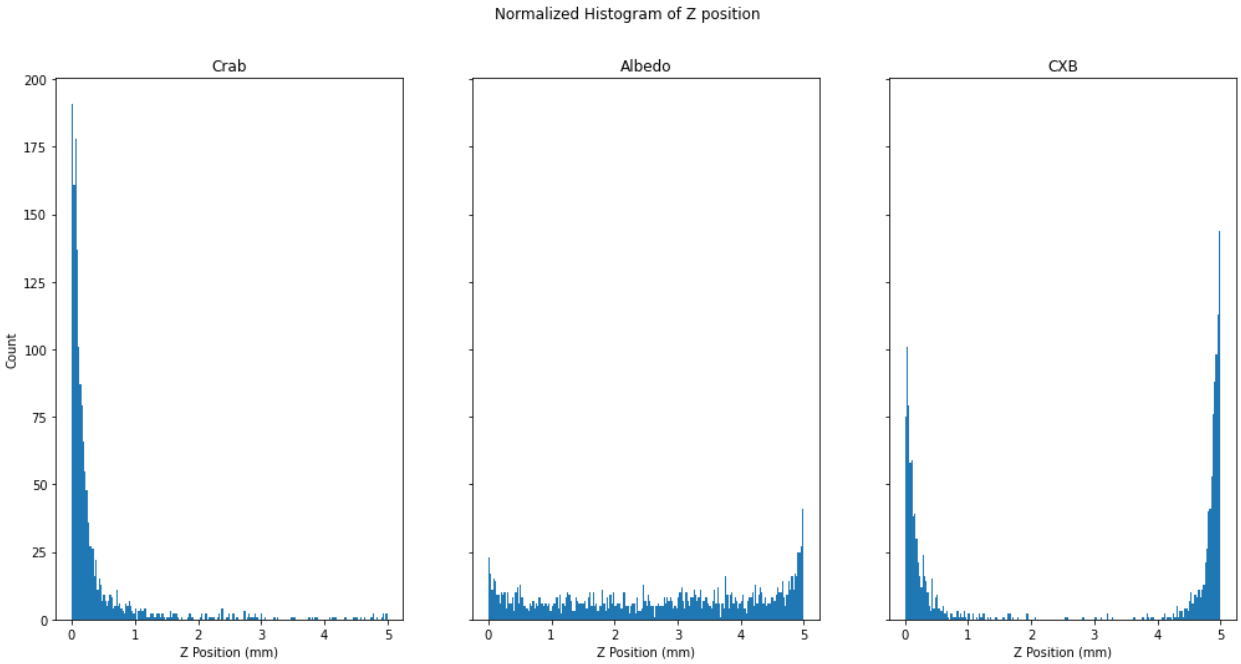


Figure 8. Z position histograms for Crab, Albedo and CXB events.

2d scatter plots are plotted as energy vs. position. There is an energy vs. z position plot below (in fig. 9) in the range of 20-80 keV. From this plot, emission lines of the cubesat components can be seen in the detector data as straight lines in specific energies for all different source events.

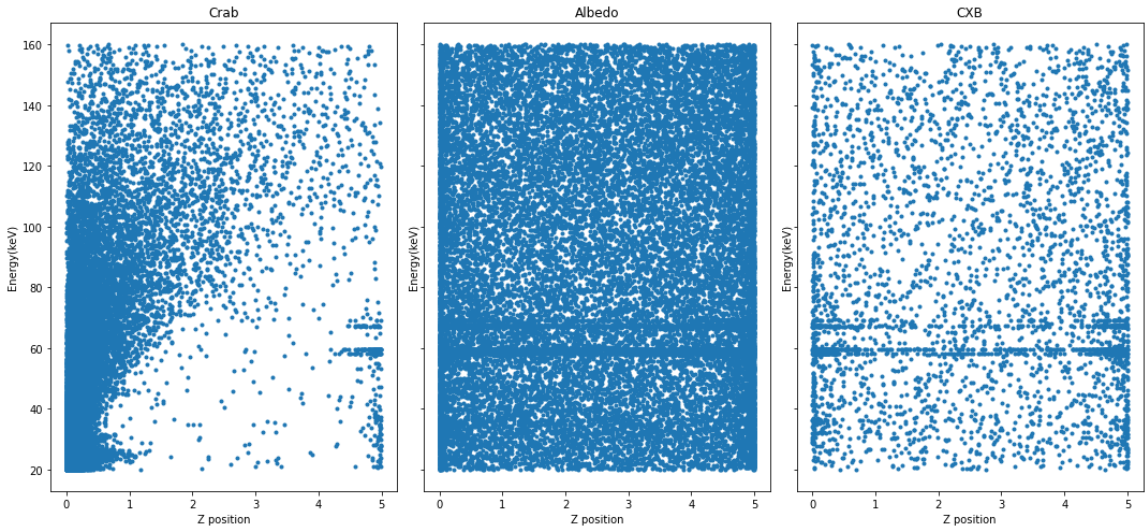


Figure 9. 2-d scatter plots for z position vs. energy of the Crab, albedo and CXB events.

The plots given in fig. 10 are the 3d scatter plots of the detector which shows the position of the events. By looking at the figures, it can be seen that x and y positions do not play a significant role in distinguishing between the source and background events. However, it is seen that the Crab events are concentrated on the upper edge (0 mm) of the detector.

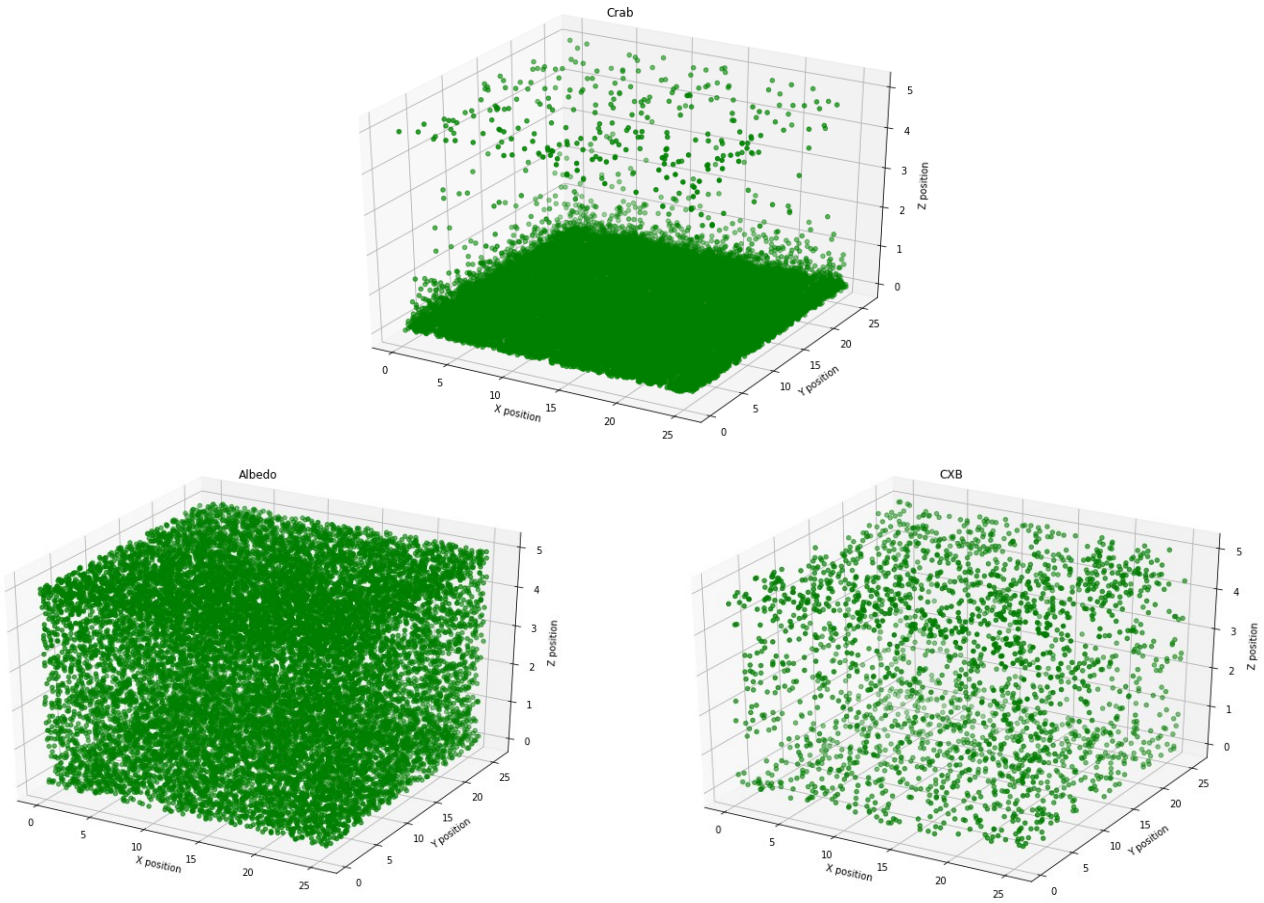


Figure 10. 3-d scatter plots for x, y and z positions of the Crab, albedo and CXB events on the iXRD.

The graph below is a 2d scatter plot which shows the x position vs. y position on the upper edge of the detector. To investigate the events on the upper edge of the detector, events between 0 and 0.1 mm on the z axis are selected. Effect of the collimator on the upper edge of the detector is clearly seen by the grid shapes on the Crab data below (in fig. 11).

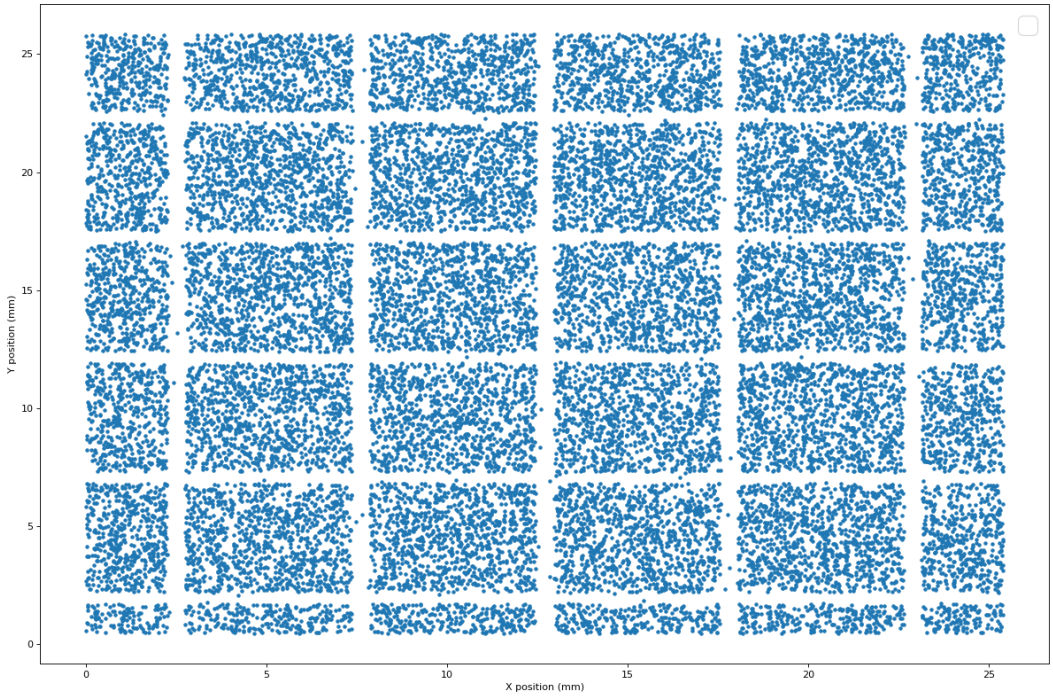


Figure 11. 2-d scatter plot of the upper edge of the iXRD for x position vs. y position.

d. Clustering of the data:

Clustering is a set of techniques for categorizing data into groups or clusters. Clusters are roughly described as collections of data objects that are more comparable to each other than to data objects from other clusters [12]. The data is clustered using the k-means clustering approach based on the features. Unsupervised machine learning technique k-means clustering is used to locate groups of data objects in a dataset. There are several clustering methods available, however k-means is one of the most prominent and widely used. As a result, the k-means clustering algorithm is applied to source, Albedo, and CXB data features independently.

The k-means approach identifies the data clusters. We can better examine the properties of the data and act as a result of this. Significant differences in clustering of different types of photon sources cannot be discovered using the available data from simulation results. Clustering methods will be used again with a larger amount of data and new attributes in order to find changes in clustering characteristics between the source, albedo, and CXB photons. Furthermore, the effect of the atomic emission spectra of the iXRD (CdZnTe) is detected using the results of k-means clustering. The concentrated straight lines in fig. 12 below demonstrate this point.

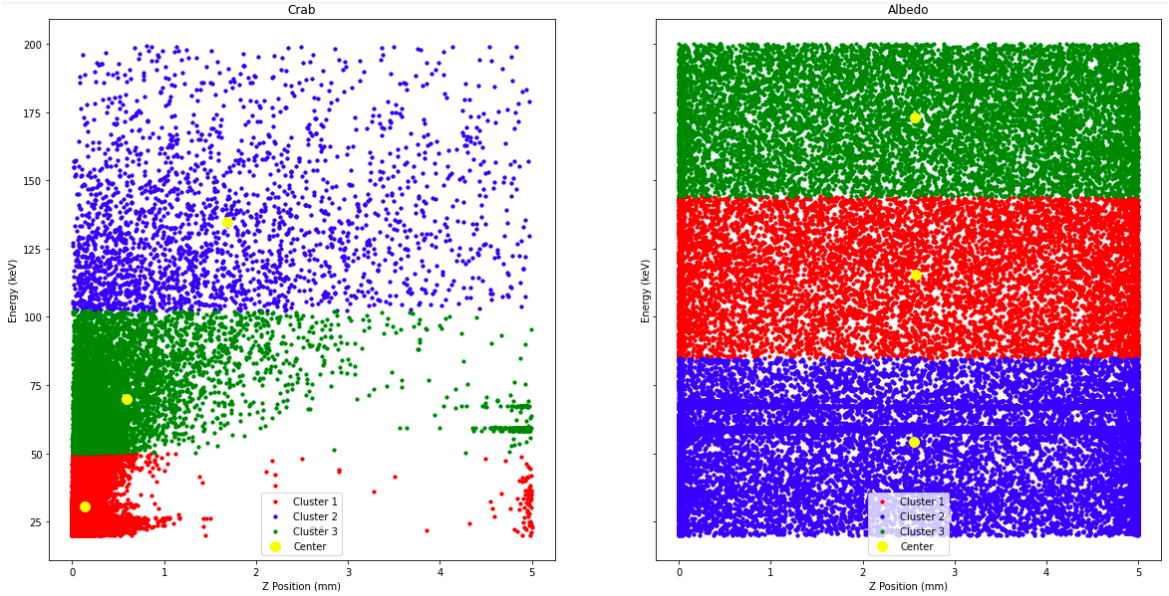


Figure 12. Clustering of the energy and z position features for the Crab and albedo events.

e. Classification of the data by Machine Learning models:

First, the features are separated from the target variable, and these labels are used for training and tests. Python's libraries such as scikit-learn were used to perform this process. After that, linear and nonlinear machine learning models were applied to the training data and the models were tested with separated test data, 20% of the main data. The scores of the models and their run times can be seen from the fig. 13.

Models	Train Score	Test Score	Time
Naive Bayes	0.8816958410107923	0.885108903072975	0.030114173889160156
Log. Regression	0.9303270597525665	0.9322234651575969	0.609877347946167
KNN	0.9462687549355093	0.9320918602355728	0.01018524169921875
Decision Tree	1.0	0.930907415937356	0.6356639862060547
Bagging DT	0.9788595683074494	0.9488714877936435	2.106945514678955
Boosted DT	0.9521255593577257	0.9521616108442456	2.00909161567688
Random Forest	1.0	0.9536092649865104	96.17297124862671
xG Boost	0.9490819952619111	0.9484766730275712	3.490297555923462
Linear SVM	0.9320380363253488	0.9350529709811147	9960.422051429749
Poly. Kernel SVM	0.6989503816793893	0.7014542343883661	75.95121312141418
Gaus. Kernel SVM	0.9103382469070808	0.9105744554846351	64.2294659614563
Sig. Kernel SVM	0.8834726243748355	0.8865565572152398	72.01771068572998
Random Fourier Features	1.0	0.9541356846746069	93.84549951553345
Voting Classification	0.9418926033166622	0.9388037112588011	10074.35827088356

Figure 13. Training scores, test scores and run-times of the ML models trained with the simulation data.

The train scores of each ML model for binary classification is plotted in the graph below (in fig. 14). According to the results, the one which uses Random Fourier Features has provided the best accuracy for the classification. In this model, features are mapped into another space using Random Fourier Features technique. After the feature mapping, Linear SVM model is trained with the new generated features.

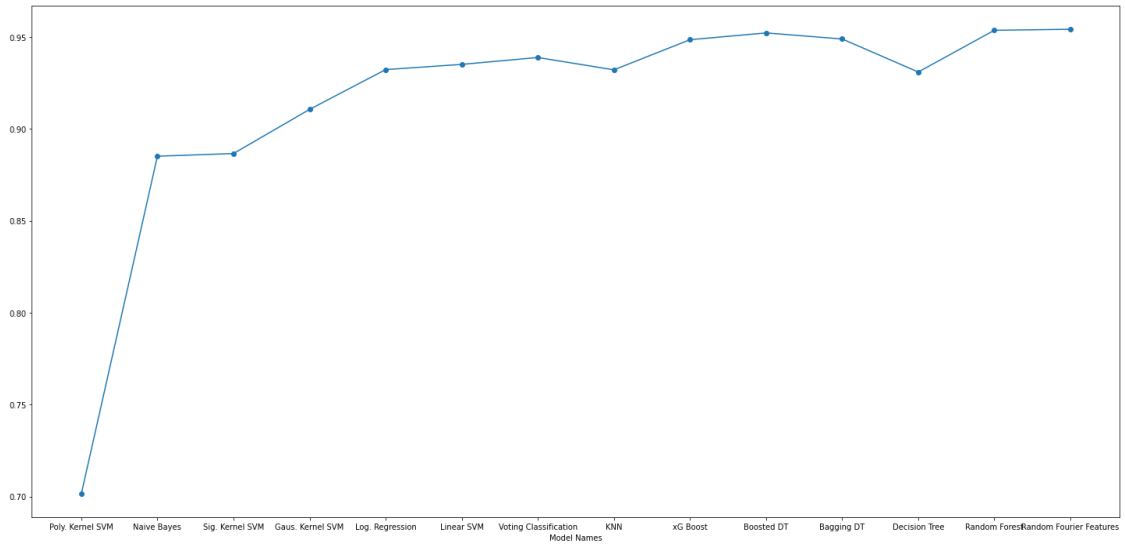


Figure 14. Test scores of the ML models for binary classification of the detector data.

In fig. 15, test and train scores of each ML model is plotted. From these results, it shows that there can be overfitting in some of the models trained such as decision tree, random forest, and Random Fourier Features. Because of the overfitting, the scores of these models are not reliable. However, xG boost and boosted decision tree models have provided high accuracy without overfitting. Therefore, xG boost and booster decision tree models can be considered to use in the later stages of the project.

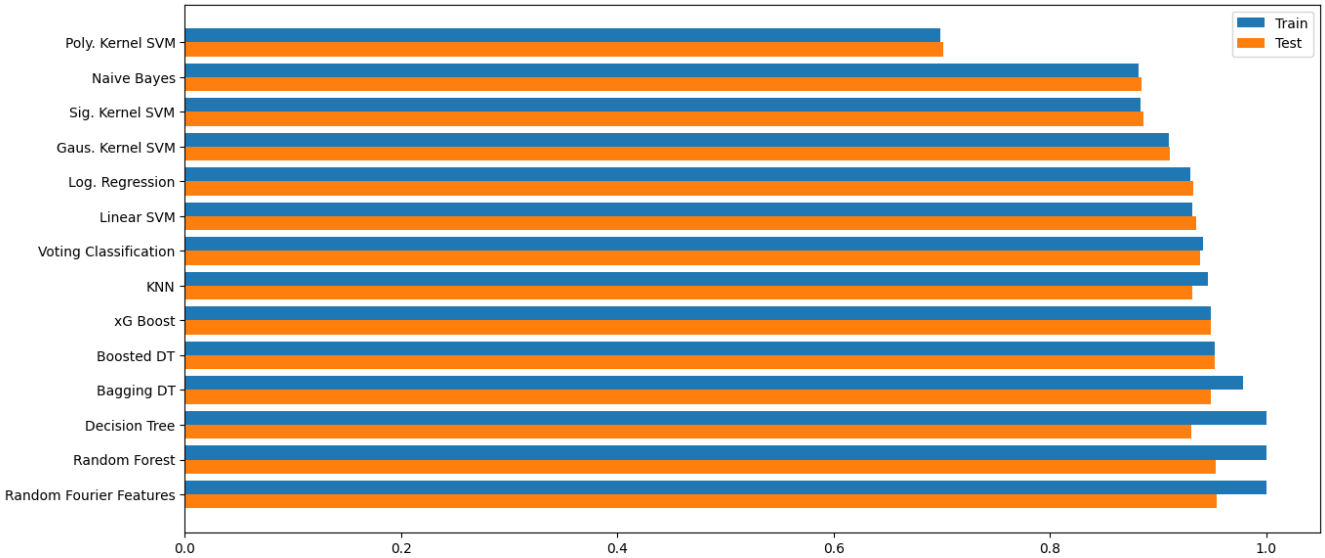


Figure 15. Train and test score comparisons for every ML model trained with the simulation data.

f. Comparison of the ML model accuracies for different types of data:

As it is explained in the pre-processing of the data part, 6 different datasets are formed with the GEANT4 simulation results. In the fig. 16, a dataset with original data and 3 datasets with different precisions are used to train the ML models. Blue ones show the results for the original GEANT4 simulation data. Orange ones show the data which is rounded to 5 decimal places. Green ones show the data which is rounded to 3 decimal places. Red ones show the data which is rounded to integer. As it is expected, decreasing the precision of the data decreases the accuracy of the models for binary classification. However, there are no significant accuracy changes between the 3 datasets with rounded x, y and z positions.

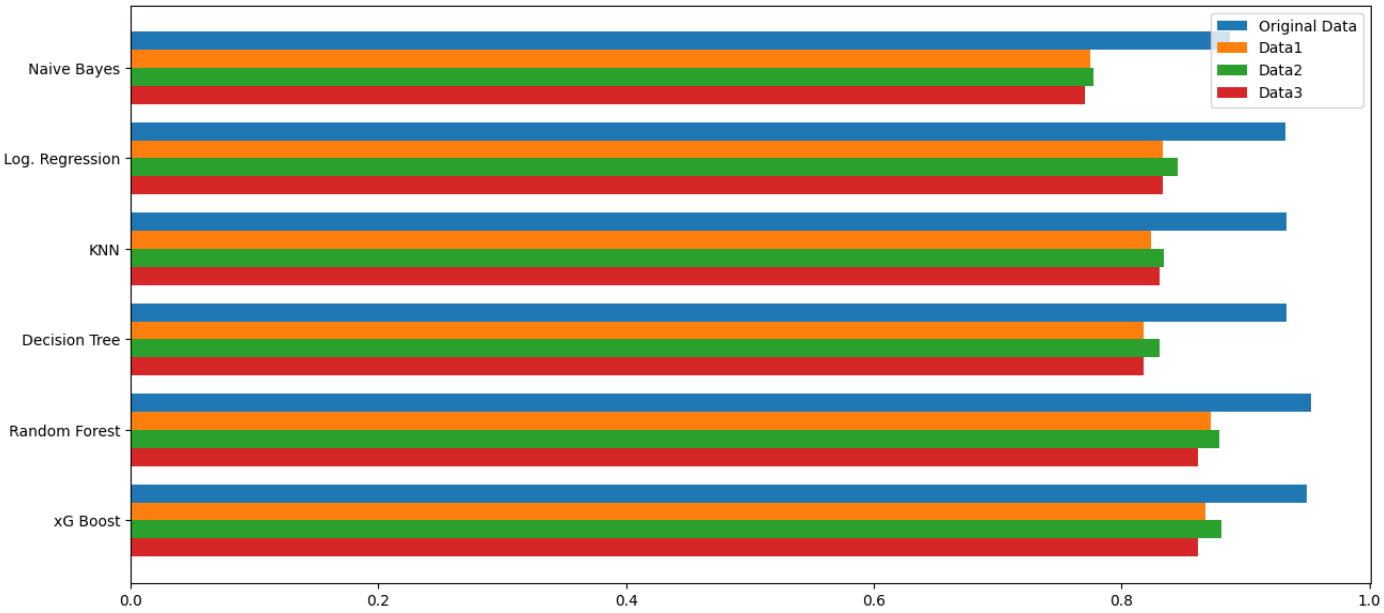


Figure 16. Test score comparisons of each ML model trained with 4 different datasets (with different position precisions).

In fig. 17, a dataset with original data and 3 datasets with different energy thresholds are used to train the ML models. Blue ones show the results for the original GEANT4 simulation data. Orange, green and red bands show data in 45-800 keV, 45-500 keV, and 45-200 keV, respectively. As it is expected, filtering the events by the energy range decreases the accuracy of the models for binary classification. However, the decrease of the accuracy in the models is more significant in the energy filtering. Therefore, it can be said that the energy range of the events affect the model accuracy more than the precision of the x, y and z positions of the events. So, it is important to have the possible widest range of energy in order to have high accuracy on the ML models for binary classification.

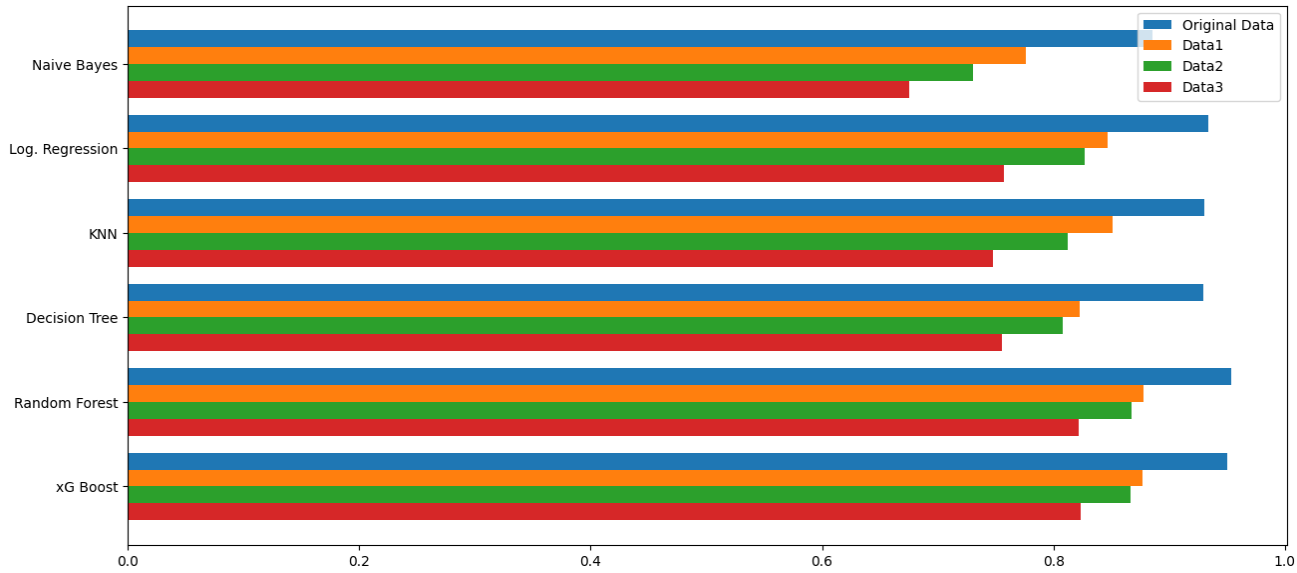
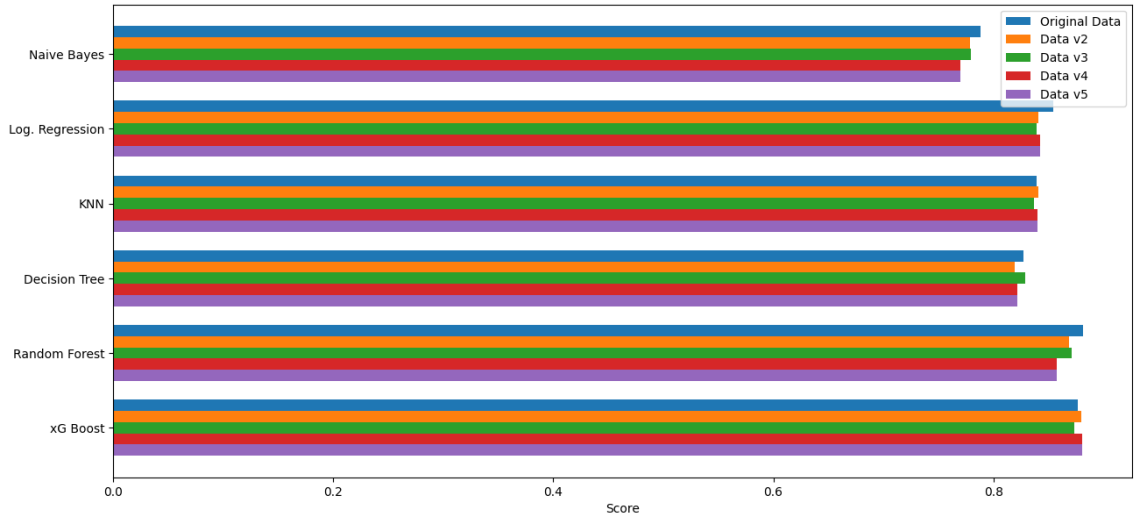


Figure 17. Test score comparisons of each ML model trained with 4 different datasets (with different energy thresholds).

In fig. 18, a dataset with original data and 4 datasets with pixel and channel mappings are used to train models. Blue ones show the results for the original GEANT4 simulation data which is between 45-800 keV energy range. Orange ones show the data with 256 pixels and z position with 0.5 precision (also between 45-800 keV energy range). Green ones show the data with 256 pixels and integer z position values (also between 45-800 keV energy range). Red ones show the data with 35 channels and z position with 0.5 precision (also between 45-800 keV energy range). Purple ones show the data with 35 channels and integer z position values (also between 45-800 keV energy range). As it is expected, decreasing the precision of z position, and mapping the x and y positions decreases the accuracy of the models. However, these decreases are not significant. Therefore, for the binary classification model accuracies, energy range and energy resolution are more crucial than the x, y, z position precisions.



g. Classification of the data by Neural Network:

In addition to the ML models explained above, the classification was made by building a Neural Network. Sequential API was used in the creation of this model. Also, binary_crossentropy was used as the loss function. The number of epochs was set to 25. The validation loss and the validation accuracy by epoch are shown in figs. 19 and 20.

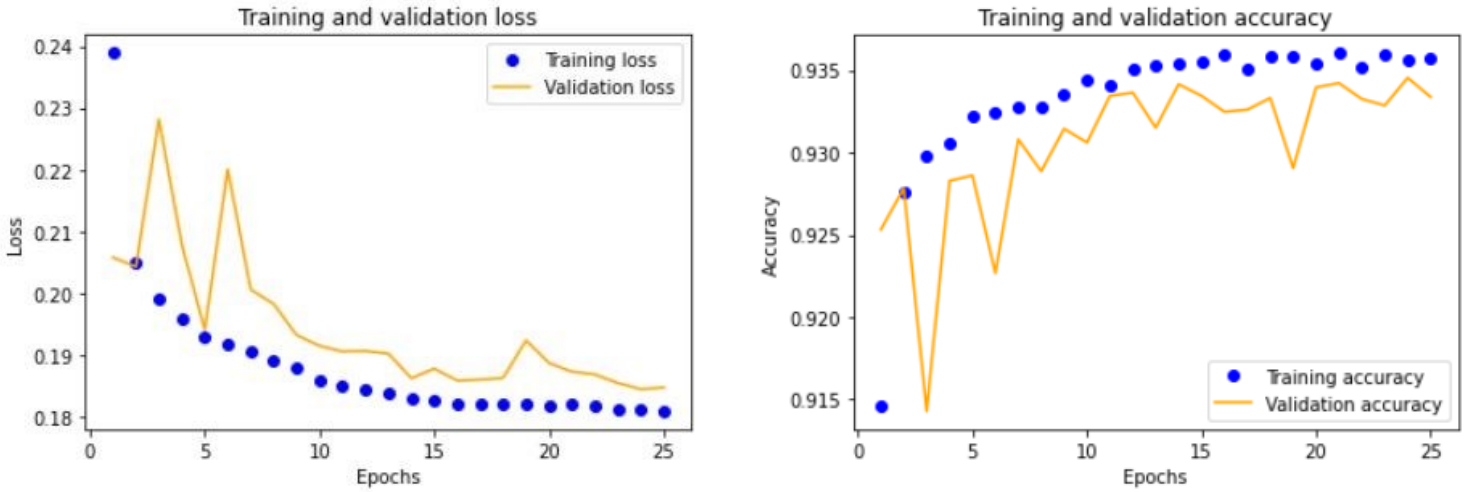


Figure 19 and 20. Validation loss (right), validation accuracy (left)

As it can be observed from the figure 21, the accuracy rate of the neural network model is 94%.

	precision	recall	f1-score	support
0	0.94	0.92	0.93	36790
1	0.93	0.95	0.94	40798
accuracy			0.94	77588
macro avg	0.94	0.93	0.94	77588
weighted avg	0.94	0.94	0.94	77588

Figure 21. Classification report of Neural Network

h. Calculation of the response matrix of the detector:

As another objective of the project, the response matrix of the detector is created. Detector response shows how internal characteristics of the detector affect the output signal of the detector. The detector's response is caused by certain electronic characteristics of the instrument so that a set of parameters causing the effect can be taken and the real spectrum that is detected by the instrument can be estimated. To be able to understand and characterize the response of the detector, the response matrix of the detector is calculated. Response matrix of the iXRD is firstly calculated with the initial GEANT4 simulations of the response of the detector. In the figure 22, the calculated response matrix of the iXRD is given in logarithmic scale for both axes.

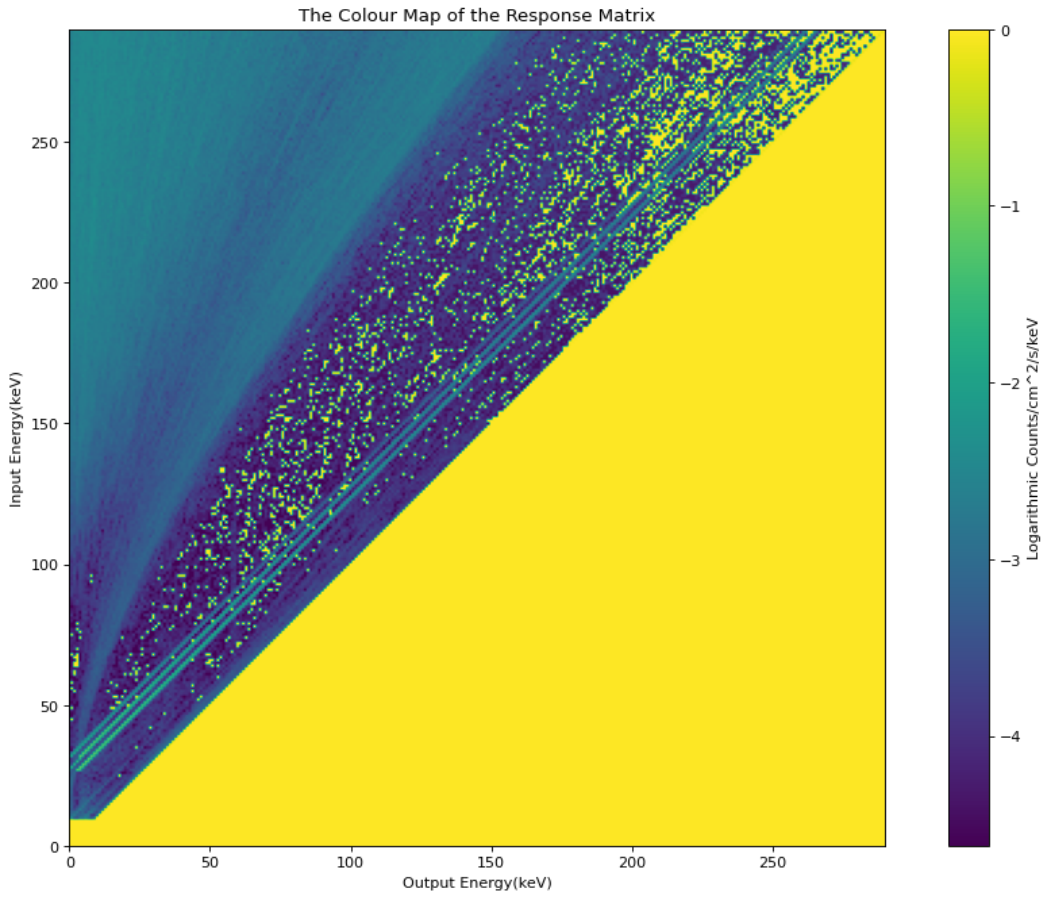


Figure 22. Color Map of the response matrix of iXRD.

Also, calculated response matrix data is transformed into FITS format in order to be able to create RMF (response matrix file) to be used the response files in standard fitting

programs such as ISIS and XSPEC. In this transformation process HEASP and HEASOFT tools by HEASARC is used. Also, AstroPy module of Python is used. In this response FITS file, calculated data is presented in EBOUNDS and MATRIX headers. Also, response file includes information about data format, type, unit and specific information about detector and mission.

4. RESULTS & DISCUSSION

As, it is planned, tools for pre-processing of the raw GEANT4 simulation data is finished. These data pre-processing tools include creating the weighted e-clouds for each event, forming datasets using the specific count rates of each event sources, pixel and channel mappings of the x, y and z positions of the events. After that visualization of the events with different plot techniques (such as histogram, 2-d and 3-d scatter plots) are done. With the visualized data and the calculated correlations of the features, data is analyzed more to start binary classification on data using machine learning models and statistical methods.

After the creation of the different datasets, different ML models (both linear, nonlinear and ensemble methods) are used to classify the data. These parameters of each model are tuned with using cross validation. After the tuning of the models is done, test and train scores are compared to check if the models are overfitted. The models with the best F1 scores are selected to use on the data. Then, these models are used on the different datasets (with different position precisions, different energy thresholds and pixel/channel mapping) to analyze the scores of the models with different data sensitivities. Also developed Neural Network is used for classification of the data.

Apart from them, the response matrix of iXRD was calculated and plotted with the simulation data and converted into FITS format to be used in standard fitting programs such as ISIS and XSPEC.

5. IMPACT

It will shed light on future academic research and studies about classification of background events in X-ray imaging. This work can also be used by other cubesat missions that are launched.

6. ETHICAL ISSUES

There are no ethical issues in the project because the data will be simulated in the HEALAB, and real data will be received from the Sharjah Sat-1 which is developed in HEALAB.

7. PROJECT MANAGEMENT

The delay of the simulation data caused it to act differently from the plan. However, this setback did not cause a problem because the implementation of the ML models was faster than planned. The reason of this is that taking the Machine Learning course accelerated the process. Besides, we still do not have THEBES data. Due to the lack of this data, it was not possible to train the models with this data, which was planned.

The lack of resources regarding the goal of converting the calculated system response to the appropriate form and the arrival of simulation data caused the focus on the establishment of machine learning models and delayed the goal of converting the system response into the appropriate form. However, with the establishment of the models, this target is among the first things to be done from now on.

8. CONCLUSION AND FUTURE WORK

By applying these ML models to real data from the cube satellite which will be launched in the next 6 months, it will be able to classify the photons hitting the satellite's detector as crab or background effect. Besides, this work can be used by other cube satellites launched. Since it is a project that has not been commonly used before for cubesat missions, it can also contribute to future academic studies.

As the future work of the project, the studies below can be continued:

- Studies on the ML models and comparisons on the different datasets will be continued.
- Conversion of the calculated response matrix into FITS data format will be done.

9. 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