American Academic Scientific Research Journal for Engineering, Technology, and Sciences
ISSN (Print) 2313-4410, ISSN (Online) 2313-4402
https://asrjetsjournal.org/index.php/American_Scientific_Journal/index

Simulation Prediction of Background Radiation Using Machine Learning

Peter Oluwasayo Adigun[a]*, Tobi Titus Oyekanmi[b], Ayodeji Adedotun Adeniyi[c]

[a]*Department of Computer Science, New Mexico Highlands University, 1005 Diamond St, Las Vegas, New Mexico, USA*

[b.c]*Department of Media, Art, and Technology, New Mexico Highlands University, 1005 Diamond St, Las Vegas, New Mexico, USA*

[a]*Email: poadigun@nmhu.edu*

[b]*Email: toyekanmi@live.nmhu.edu*

[c]*Email: aadeniyi1@live.nmhu.edu*

Abstract

The simulation of the natural background radiation dataset is research that implemented the application of machine learning in radiation physics. This is achieved by training natural background radiation datasets using different machine learning algorithms. The background radiation dataset is acquired from a field study carried out in the Gwagwalada Area, Abuja, Federal Capital Territory, Nigeria. The different machine learning algorithms applied are Random Forest, Naïve-Bayes, Support Vector Machine, and Kernel Support Vector Machine. Random Forest algorithms have the best test accuracy of 94.0%, a trained score of 98%, a K-fold cross validation score of 96.9%, and efficiently classify the effect of background radiation as harmful or harmless. This result established the integrated application of artificial intelligence and therefore indicates that machine learning has the ability to classify and categorize the effect of background radiation datasets.

Keywords: Machine Learning (ML); Random Forest (RF); Naïve-Bayes (NB); Support Vector Machine (SVM); Kernel Support Vector Machine (KSVM).

Received: 11/23/2024
Accepted: 1/23/2025
Published: 2/3/2025

----

* Corresponding author.

1. Introduction

In the field of science and engineering, simulation, modeling, and prediction are pertinent approaches to describing and understanding the dynamism of the real world. Some of the systems employed in making such decisions and predictions are artificial intelligence (AI) and machine learning (ML). Machine learning (ML) is a field of inquiring, learning, and interpreting, and it is the sole aim of the system (machine) being trained for. “It is a field of inquiry devoted to understanding and building methods, methods that model data to enhance performance on some set of tasks”[1].

In some parts of the world, little attention or zero attention is given to background radiation, which is hazardous to life and our environment. Background radiation is the amount of ionizing radiation present in the environment at a particular location which is not due to the deliberate introduction of radiation sources. Background radiation is also the intensity of ionizing radiation in a particular area per unit of time (hour). Background radiation originates from a variety of sources, both natural and artificial. These include cosmic radiation and environmental radioactivity such as naturally occurring radioactive materials (NORMs) including radon and radium, and man-made fallout from nuclear weapons testing and nuclear accidents.

Radionuclides or radioisotopes are the main elements that cause ionizing radiation. These radioisotopes are heavy nuclei and unstable atoms that have high amounts of energy. There are classes of nuclei which are numerous and unstable elements. They break up or disintegrate spontaneously by emitting some corpuscular or electromagnetic radiation of very high energy. In the first case, the atomic number Z or the mass number A or both, of the nucleus change, thereby producing an altogether new nucleus. In the second case, the nucleus makes a transition from a quantum state of higher energy to one of lower energy. This spontaneous transformation of a nucleus is known as radioactivity which was the first nuclear phenomenon to be discovered[2].

Radionuclides or radioisotopes are the main elements that cause ionizing radiation. These radioisotopes are heavy nucleus unstable atoms having high amounts of energy. There are classes of nuclei which are numerous and unstable elements. They break up or disintegrate spontaneously with the emission of some corpuscular or electromagnetic radiation of very high energy. In the first case either, the atomic number Z or the mass number A or both, of the nucleus change, thereby producing an altogether new nucleus. In the second case, the nucleus makes a transition from a quantum state of higher energy to one of lower energy. This spontaneous transformation of a nucleus is known as radioactivity which was the first nuclear phenomenon to be discovered[2].

The content of radioactive material is found throughout nature and detectable amounts occur naturally in soil, rocks, water, air, and vegetation, from which it is inhaled and ingested into the body. Unknowingly, this is one of the major causes of medical complications and even radiation exposure to background radiation is another worse unknown case. In addition to this internal exposure which occurs due to ingestion and inhalation, humans also receive external exposure from radioactive materials that remain outside the body and from cosmic radiation from space. The worldwide average natural dose to humans is about 2.4 mSv (240 mrem) per year[3]. This is four times the worldwide average artificial radiation exposure, which in 2008 amounted to about 0.6

millisieverts (60 mrem) per year. In some rich countries, like the US and Japan, artificial exposure is, on average, greater than natural exposure, due to greater access to medical imaging. In Europe, average natural background exposure by country ranges from under 2 mSv (200 mrem) annually in the United Kingdom to more than 7 mSv (700 mrem) annually for some groups of people in Finland [4].

The simulation and training of a system that can selectively identify various background radiation have not been commonly studied. Therefore, the use of ML to correctly predict various background radiation, including alpha, beta, and gamma background radiation, is necessarily essential. This research aims to employ ML algorithms for the simulation and prediction of background radiation (alpha, beta, and gamma). This research focuses on exploratory data analysis of background radiation datasets through supervised learning.

The objectives of the research were:

- To survey and acquire sufficient background ionizing radiation data from the area of study.
- To analyze and optimize the collected background ionizing radiation data.
- To make a preprocessing algorithm for the simulation to be implemented.
- To train the system that can recognize, and interpret the dataset of alpha, beta, or gamma radiation using machine learning.
- To test and verify the system for simulation of different ionizing radiation.
- To confirm and establish the standard operation of the system.

1.1 Background Radiation

All the background radiations are ionizing radiations, that is; they have non-zero rest energy, and the ability to ionize (electron transition) material or substance in their part of propagation. Mainly, background radiation is nuclear radiation (particle radiation: subatomic particle emission), while the minor ones are cosmic radiation (muons, positrons) [5]. The ionizing radiation consists of subatomic particles or electromagnetic waves that have quantifiable energy to ionize atoms or molecules by removing electrons from them. Typical ionizing subatomic radiation includes alpha-ray, beta-ray, and gamma-ray. These are naturally created by radioactive decay, and almost all are energetic enough to ionize.

1.1.1 Alpha Background Radiation

Alpha radiation is characterized by particles that are the same as helium-4 nuclei (two protons and two neutrons). They undergo interaction with matter strongly, due to their charge properties and mass number [5]. Their usual velocities only penetrate a few centimeters of air or a few millimeters of low degree of thickness of the material. This indicates that “alpha particles from ordinary alpha decay do not penetrate the outer layers of dead skin cells and cause no damage to the live tissues below”. Also, there are rare alpha particles that possess very high energy alpha particles that are composed of about “10% of cosmic rays” and these variants of alpha radiation are capable of damaging the body cells and tissues, especially to astronauts, and even penetrating thin metal plates [5]. However, they are deflected by the Earth's magnetic field and then shielded away by Earth’s atmosphere.

Moreover, alpha radiation is dangerous when alpha-emitting radioisotopes are ingested or inhaled into the body system. This brings the radioisotope close enough to sensitive live tissue for the alpha radiation to damage cells. Also, per unit of energy, alpha particles are at least 20 times more effective at cell damage as gamma rays and X-rays. Examples of highly poisonous alpha-emitters are all isotopes of radium, radon, and polonium.

1.1.2 Beta Background Radiation

Beta radiation has two forms namely beta-minus (electron) and beta-plus (positron). Beta-minus (β^-) radiation is made up of an energetic electron [5]. It has more penetrating power than alpha radiation but a lesser one than gamma. Beta radiation from radioactive decay can be absorbed with a few centimeters of plastic or a few millimeters of metal. Beta radiation is emitted when a "neutron decays into a proton in a nucleus, releasing the beta particle and an antineutrino". Also, there is a variant form of beta radiation from linac accelerators which is far more energetic and penetrating than natural beta radiation. This beta radiation is used therapeutically in radiotherapy to treat superficial tumors.

Beta-plus (β^+) radiation is the emission of positrons, which are distinctly the antimatter form of electrons. When a positron slows to speeds similar to those of electrons in the material, the positron will annihilate an electron, releasing two gamma photons of 511 keV in the process [6].

1.1.3 Gamma Background Radiation

Gamma (γ) radiation is made up of photons with a wavelength less than $3x10^{-11}$ meters, a frequency greater than 10^{19} Hz, and an energy value greater than 41.4 keV). Gamma radiation emits radiation which is a nuclear process that occurs to eliminate an unstable nucleus of excess energy which involves nuclear reactions (production of a new nucleus, nuclear particles, and energy liberation[7]. Unlike alpha and beta particles that possess an electric charge and mass, gamma radiation is mainly composed of photons, which have neither mass nor electric charge and have a penetrating power that can pass through matter than either alpha or beta radiation.

Gamma radiation is absorbed mainly by thick or dense layers of material, "where the stopping power of the material per given area depends mostly on the total mass along the path of the radiation, regardless of whether the material is of high or low density"[6]. All gamma Radiation that approaches the Earth through space is absorbed by the atmosphere. Air also can absorb gamma Radiation thereby reducing the energy passing through it by half, an average 500ft (150m).

2. Methods

2.1 Description of the Study

The location of the study was Abuja (9.0765° N, 7.3986° E), Federal Capital Territory, Nigeria. The city is the capital and eighth most populous city of Nigeria.[8] . The field study was carried out in the Gwagwalada area, Abuja, Nigeria. Gwagwalada (8.9508° N, 7.0767° E) is a local government area and the main town in the Federal Capital Territory in Nigeria. Gwagwalada has an area of 1,043 km^2 and a population of 157,770 at 2006.[9] Although, Gwagwalada was more than a quarter of a million as total population in 2023. It is bordered by other local governments which are Abaji, Kuje and Kwali.

This study covered a field study in Gwagwalada residential area. This area is highly populated with residents, landfills and dumpsites which are located in this vicinity quite a lot. This area is well polluted and this area is chosen due to the following activities:

- Consumer items: air travel, cigarettes, building materials, hydrocarbon products, food materials etc.
- Occupational exposure: people and residents that work and harbor around.
- Inhalation of air: mainly from radon, depends on indoor accumulation from the dumpsite and landfill.
- Terrestrial radiation from the ground: depends on soil and building materials
- Cosmic radiation from space: depends on altitude

The residential area was chosen to check and measure the background radiation dose rate that the resident could probably be exposed to due to the activities mentioned above.

2.2 Design and Technique of the Study

The study design was an experimental-based and field data collection design. The study involved the measurement of the amount of radiation per unit of time from naturally occurring radioactive materials (NORM) which are present in the surroundings. The radiation per unit of time is called a dose or dose rate which is based on radioactivity, a term used to describe the disintegration of atoms.[10] The unit of absorbed radiation dose is the sievert (Sv). Since one sievert is a large quantity, radiation doses normally encountered are expressed in millisievert (mSv) or microsievert (µSv).[10] The standard and average radiation exposure due to all-natural sources is 2.4 mSv (0.27µSv) a year as published by IAEA, reported by UNSCEAR and established by ICRP[11]. This established value was used to indicate and distinguish the measured to be either harmful or harmless. Any value less than or equal to 0.27µSv/hr was interpreted as harmless and any value greater than the standard is interpreted as harmful. Also, these radiations were categorically the three radiations, namely alpha, beta, and gamma. The quantification of these radiations and expression with time is the technique used to record the amount of radiation exposure or natural background radiation in an area

2.3 Instrumentation of the Study

1. Gamma Scout

Gamma Scout (model GS2 with serial number A20) is a part of the Gamma Scout series which is a radiation measuring device that measures alpha, beta, gamma radiation and also x-rays.[12] The gamma senses ionizing radiation using a G-M (Geiger Muller) tube within a thin mica window. The window monitor is optimized to detect low and high levels of ionization. Gamma Scout has the following accessories: Cumulative dose function, two bytes data memory, audio warning alert, visual indicator of annual dose, audible pulse mode (total count rate per second), dosage exposure chart, multi-functional keypad, digital LCD screen display, a reliable 9v battery, measures in mSv/hr or mRem/hr units, wide range temperature (-40 -700C), USB 2.0 data transfer port to connect to a personal computer (pc). The Gamma Scout sets a new standard in portable Geiger counter performance and functionality. This radioactivity meter has a wide measuring range and is used for diverse, long-lasting measurement types.[12] The radioactivity meter provides a certified measurement of the environmental radiation as well as of the radiation which is artificially increased up to 500 times over the limit value. This radioactivity meter has diverse applications.

2. **Timer and Stopwatch**

A timer and stopwatch were used to record the time when the measurement and recording were done. It is necessary to maintain steady and regular timing for the measurement and fixation of the device (Gamma Scout) at the point of reading the background radiation.

3. Pegs

Pegs are used as point indicators, used to mark out where point where measurement and reading were taken

2.4 Software Toolkit of the Study

1. Microsoft Excel

Microsoft Excel 2016 is the spreadsheet application used for data entry, data sorting and data cleansing. This application is also used for formatting the dataset file in *csv* format.

2. Python Programming Language

Python is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation[13]. This programming language is used to code, model, train and simulate the background radiation.

2.4 Data Modelling

Microsoft Excel application was used to optimize and format the natural background dataset for Python programming language Firstly, the library functions of the Python programming software were initialized and the dataset was imported. Secondly, the information and description of the dataset were computed and also, and the visualization of the dataset was performed to view the trend between harmfulness and harmlessness of the effect of the background radiation. Thirdly, the dataset was prepared for modelling, training and simulation. Recoding, categorization and other data manipulation include, the column for the effect of radiation is coded as a binary value (0 and 1), the radiation value categorized as the independent variable, and the summation and effect of radiation categorized as the dependent variable. Then, the dataset was trained using the library function of the science kit tool. Machine learning algorithms, namely random forest, Naive-Bayes, SVM, and kernel

SVM, were used in training and testing the model. A confusion matrix was also initialized for evaluating the model and assessing the performance metrics of each machine learning algorithm. Lastly, each algorithm was used to simulate the background radiation data in which the machine was trained using different algorithms, and each model was tested and evaluated and performance metrics were also computed.

3. Results

The dataset of the background radiation collected from the field consists of 1554 data each for alpha, beta and radiation. The dataset was used for modeling and training machine learning algorithms namely random forest, Naives-Bayes, SVM, and

3.1 Data Visualization and Optimization

The visualization of the background radiation dataset is essential because its easily reveals the trend and illustrates the variation between the harmfulness and harmlessness of the effect of the natural background radiation. "Input 1 is the code used to generate the count plot, which is a bar chart.

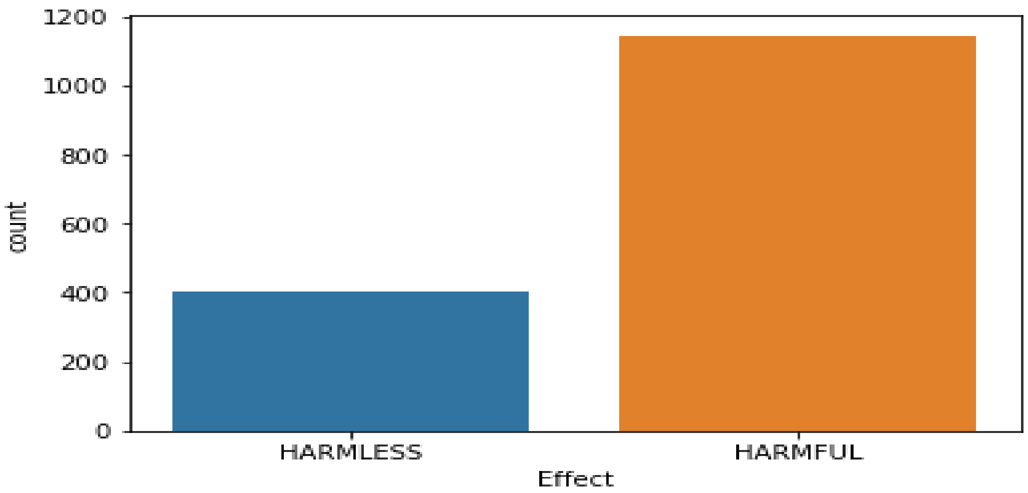

Figure 1: Illustration of the effect of background radiation

Input 2 was the code used to categorize the effect of background radiation namely harmful and harmless as a binary value, that is, "harmless = 0", "harmful = 1". Input 1: Code for categorization of the effect of background radiation

In [2]:

```
dataset = dataset.replace(['HARMLESS','HARMFUL'],[0, 1])

dataset
```

Figure 6

Out [1]:

	Alpha	Beta	Gamma	Summation	Effect
0	0.03	0.005	0.189	0.224	0
1	0.02	0.001	0.178	0.199	0
2	0.05	0.078	0.177	0.305	1
3	0.06	0.12	0.151	0.331	1
4	0.04	0.011	0.158	0.209	0
…	…	…	…	…	…
1539	0.063	0.007	0.199	0.269	0
1540	0.085	0.015	0.219	0.319	1
1541	0.08	0.01	0.226	0.316	1
1542	0.084	0.01	0.23	0.324	1
1543	0.086	0.012	0.234	0.332	1

Figure 7

3.2 Data Training

The "*sklearn library*" in the Python program is a science kit tool, which is used to train and test datasets. The "*sklearn library*" also includes sub-libraries. These libraries were initialized to use machine learning algorithms for training the machine so that classification prediction was performed. Also, the confusion matrix is part of the library initialized to evaluate the performance of the classification model through the calculation of performance metrics. Inputs 3– 7 are Python codes and "*sklearn library function*" used for training and testing the modelling machine.

In input 3, the "*sklearn.model_selection*" was used to split arrays or case matrices into random subsets for train and test data respectively.

Input 3: Initialization of the library functions for training and testing

In [3]:

```
from sklearn.model_selection import train_test_split
```

Thirty percent (30%) of the dataset was used for training the machine and the remaining 70% was reserved for testing and prediction purposes.

Input 4: Selection and splitting of the background radiation dataset for training and testing

In [4]:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)
```

Input 5: Training of the data of the independent variables

In [5]:

```
X_train
```

Output 2: Trained data of the independent variables

Out [2]:

	Alpha	**Beta**	**Gamma**
87	0.081	0.011	0.219
845	0.076	0.016	0.188
1063	0.055	0.039	0.18
434	0.081	0.018	0.22
1281	0.071	0.002	0.188
...	...	...	...
1206	0.081	0.018	0.22
893	0.01	0.02	0.22
278	0.086	0.012	0.234
515	0.005	0.015	0.188
91	0.072	0.025	0.179

Figure 8

Input 6: Training of the data of the dependent variable.

In [6]:

```
y_train
```

Output 3: Trained data of the dependent variable

Out [3]:

```
87      1

845     1

1063    1
```

```
434     1

1281    0

      ..

1206    1

893     0

278     1

515     0

91      1

Name: Effect, Length: 1080, dtype: int64
```

Figure 9

Input 7: Testing of the data of the independent variables

In [7]:

```
X_test
```

Output 4: Tested data of the independent variables

Out [4]:

	Alpha	Beta	Gamma
673	0.102	0.012	0.198
1147	0.096	0.008	0.211
527	0.01	0.001	0.179
867	0.074	0.012	0.211
1191	0.078	0.037	0.173
...	...	...	...
1149	0.089	0.02	0.22
306	0.067	0.024	0.221
1057	0.076	0.034	0.177
842	0.075	0.027	0.178
1189	0.081	0.003	0.231

Figure 10

Input 8: Testing of the data of the dependent variable

In [8]:

```
y_test
```

Output 5: Tested data of the dependent variable

Out [5]:

```
673     1
1147    1
527     0
867     1
1191    1
      ..
1149    1
306     1
1057    1
842     1
1189    1
Name: Effect, Length: 464, dtype: int64
```

Figure 11

3.3 Random Forest Data Simulation

Part of the sub-library functions of "*sklearn*", which is "*sklearn.ensemble*", is a module that includes two averaging algorithms based on randomized decision trees. The random forest classifier (RFC) was used to create a set of decision trees from a randomly selected subset of the training set.

Input 9: Initialization of random forest algorithm for training the data

In [9}:

```
from sklearn.ensemble import RandomForestClassifier



RFC = RandomForestClassifier(n_estimators=100, criterion= 'entropy')

RFC.fit(X_train, y_train)
```

Figure 12

Output 6: RFA classifier initialized

Out [6]:

```
RandomForestClassifier(criterion='entropy')
```

Input 10: Code for prediction of the trained data

In [10]:

```
y_predict_train = RFC.predict(X_train)

y_predict_train
```

Figure 13

Output 7: Prediction of the trained data, Out [7]:

```
array([1, 1, 1, ..., 1, 0, 1], dtype=int64)
```

Input 11: Code for computation of the confusion matrix of the trained data

In [11]:

```
from sklearn.metrics import classification_report ,confusion_matrix

cm = confusion_matrix(y_train, y_predict_train)

sns.heatmap(cm,annot = True)
```

Figure 14

Output 8: Computation of the confusion matrix of the trained data

Out [8]:

```
<AxesSubplot:>
```

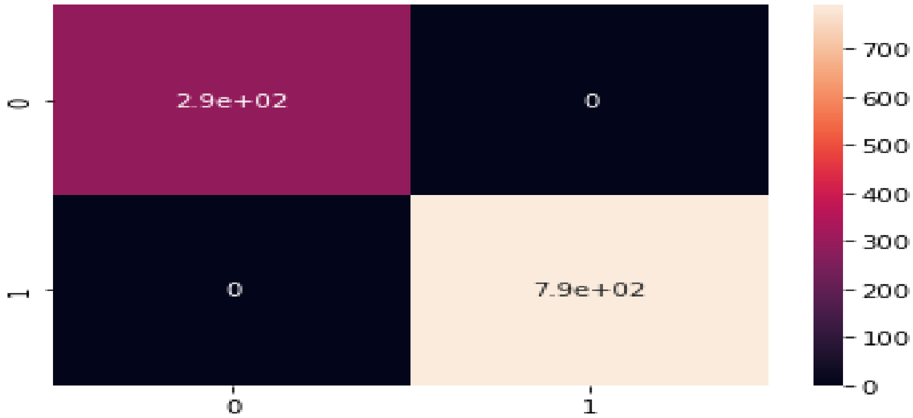

Figure 2: Confusion matrix chart of trained data by RFA

Input 12: Code to generate the performance metrics of the trained data

In [12]:

```
print(classification_report(y_train, y_predict_train))
```

Output 9: Performance metrics of the trained data

Out [9]:

```
          precision    recall  f1-score   support

       0       1.00      1.00      1.00       289

       1       1.00      1.00      1.00       791

accuracy                           1.00      1080

macro avg      1.00      1.00      1.00      1080

weighted avg   1.00      1.00      1.00      1080
```

Figure 15

Input 13: Initialization of random forest algorithm for testing the data

In [13]:

```
from sklearn.metrics import classification_report ,confusion_matrix

y_predict_test = RFC.predict(X_test)

cm = confusion_matrix(y_test, y_predict_test)

sns.heatmap(cm,annot = True)
```

Figure 16

Output 10: Computation of the confusion matrix of the tested data

Out [10]:

```
<AxesSubplot:>
```

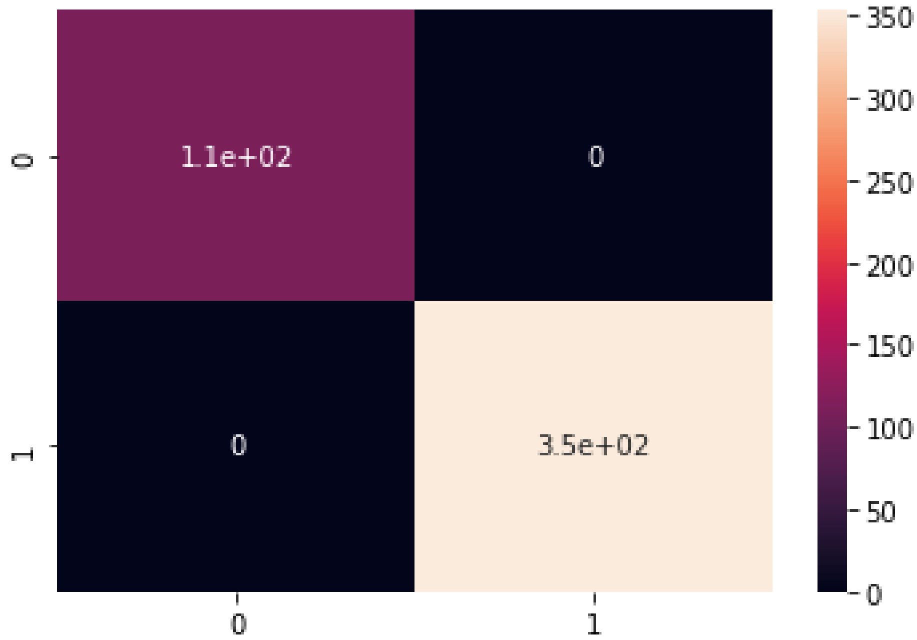

Figure 3: Confusion matrix chart of tested data by RFA

Input 14: Code to generate the performance metrics of the tested data

In [14]:

```
print(classification_report(y_test, y_predict_test))
```

Output 11: Performance metrics of the tested data

Out [11]:

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       111
           1       1.00      1.00      1.00       353

    accuracy                           1.00       464
   macro avg       1.00      1.00      1.00       464
weighted avg       1.00      1.00      1.00       464
```

Figure 17

3.4 Naive-Bayes Data Simulation

The Naïve-Bayes models are a group of extremely fast and simple classification algorithms that are often suitable for very high dimensional datasets.

Input 15: Initialization of the Naïve-Bayes algorithm for training the data

In [15]:

```
from sklearn.naive_bayes import MultinomialNB

NB_classifier = MultinomialNB()

NB_classifier.fit(X_train, y_train)
```

Figure 18

Output 12: Initialized NBA classifier

Out [12]:

```
MultinomialNB()
```

Output 13: Code for prediction of the trained data

In [16]:

```
from sklearn.metrics import classification_report, confusion_matrix



y_predict_train = NB_classifier.predict(X_train)

y_predict_train
```

Figure 19

Output 14: Prediction of the trained data

Out [14]:

```
array([1, 1, 1, ..., 1, 1, 1], dtype=int64)
```

Input 17: Code for computation of the confusion matrix of the trained data

In [17]:

```
cm = confusion_matrix(y_train,y_predict_train)

sns.heatmap(cm,annot=True)
```

Figure 20

Output 15: Computation of the confusion matrix of the trained data

Out [15]:

```
<AxesSubplot:>
```

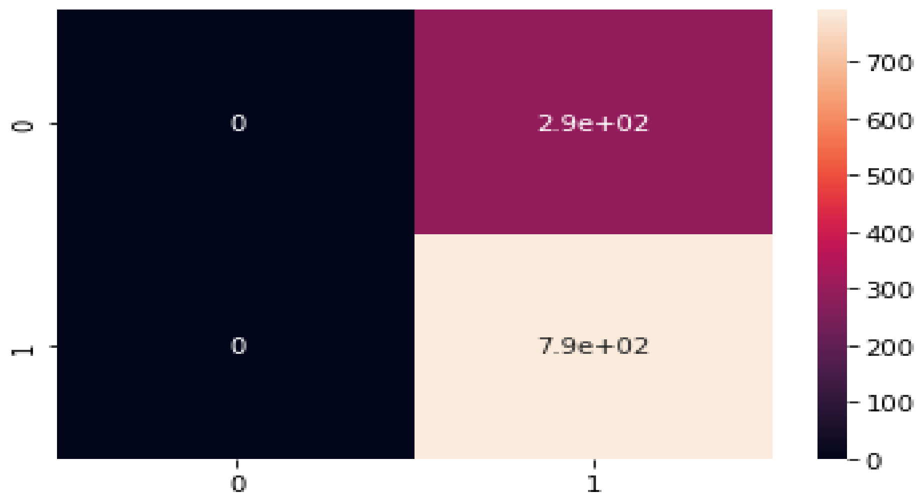

Figure 4: Confusion matrix chart of trained data by NBA

Input 18: Code to generate the performance metrics of the trained data

In [18]:

```
print(classification_report(y_train, y_predict_train))
```

Output 16: Performance metrics of the trained data

Out [16]:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	111
1	1.00	1.00	1.00	353
accuracy			1.00	464
macro avg	1.00	1.00	1.00	464
weighted avg	1.00	1.00	1.00	464

Figure 21

Input 19: Code for prediction of the tested data

In [19]:

```
from sklearn.naive_bayes import MultinomialNB

NB_classifier = MultinomialNB()

NB_classifier.fit(X_test, y_test)
```

Figure 22

Output 17: NBA classifier initialized

Out [17]:

```
MultinomialNB()
```

Input 19: Code for computation of the confusion matrix of the tested data

In [19]:

```
from sklearn.metrics import classification_report, confusion_matrix

y_predict_test = NB_classifier.predict(X_test)

cm = confusion_matrix(y_test,y_predict_test)

sns.heatmap(cm,annot=True)
```

Figure 23

Output 18: Computation of the confusion matrix of the tested data

Out [18]:

```
<AxesSubplot:>
```

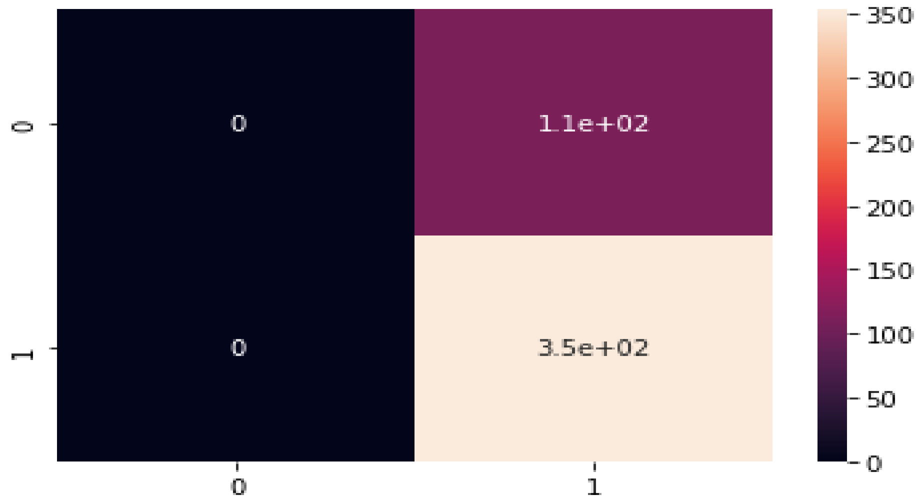

Figure 5: Confusion matrix chart of tested data by NBA

Input 20: Code to generate the performance metrics of the tested data

In [20]:

```
print(classification_report(y_test, y_predict_test))
```

Output 19: Performance metrics of the tested data

Out [19]:

	precision	recall	f1-score	support
0	0.00	0.00	0.00	111
1	0.76	1.00	0.86	353
accuracy			0.76	464
macro avg	0.38	0.50	0.43	464
weighted avg	0.58	0.76	0.66	464

Figure 24

3.5 Support Vector Machine Data Simulation

The SVM algorithm is a supervised learning algorithm used for outlier detection, regression and classification that is both powerful and adaptable.

Input 21: Initialization of the SVM algorithm for training the data

In [21]:

```
from sklearn.svm import SVC

classifier = SVC(kernel = 'linear', random_state = 0)

classifier.fit(X_train, y_train)
```

Figure 25

Output 20: SVMA classifier initialized

Out [20]:

```
SVC(kernel='linear', random_state=0)
```

Input 22: Code for prediction of the tested data, In [22]:

```
print(classifier.predict([[0.95,0.560,0.960]]))
```

Output 21: Prediction of the tested data

Out [21]:

```
[1]
```

Input 23: Code to generate the accuracy score of the tested data

In [23]:

```
from sklearn.metrics import accuracy_score

print(accuracy_score(y_test, y_predict_test)*100)
```

Figure 26

Output 22: Accuracy score of the tested data

Out [22]:

```
76.07758620689656
```

Input 24: Code to generate the accuracy score of the trained data

In [24]:

```
from sklearn.metrics import confusion_matrix, accuracy_score

cm = confusion_matrix(y_train,y_predict_train)

print(cm)

accuracy_score(y_train, y_predict_train)*100
```

Figure 27

Output 23: Accuracy score of the trained data

Out [23]:

```
[[  0 289]

 [  0 791]]

73.24074074074073
```

Figure 28

3.6 Kernel SVM Data Simulation

The kernel SVM algorithm is a machine learning which transforms the data into the required form using the kernel trick.

Input 25: Initialization of Kernel SVM algorithm for training the data

In [25]:

```
from sklearn.svm import SVC
```

```
classifier = SVC(kernel = 'rbf', random_state = 0)

classifier.fit(X_train, y_train)
```

Figure 29

Output 24: Initialized KSVM algorithm

Out [24]:

```
SVC(random_state=0)
```

Input 26: Code for prediction of the tested data

In [26]:

```
print(classifier.predict([[0.95,0.560,0.960]]))
```

Output 25: Prediction of the tested data

Out [25]:

```
[1]
```

Input 27: Code to generate the accuracy score of the tested data

In [27]:

```
from sklearn.metrics import accuracy_score

print(accuracy_score(y_test, y_predict_test)*100)
```

Figure 30

Output 26: Accuracy test of the tested data

Out [26]:

```
76.07758620689656
```

Input 28: Code to generate the accuracy score of the tested data

In [28]:

```
from sklearn.metrics import confusion_matrix, accuracy_score

cm = confusion_matrix(y_train,y_predict_train)

print(cm)

accuracy_score(y_train, y_predict_train)*100
```

Figure 31

Output 27: Accuracy test of the trained data

Out [27]:

```
[[  0 289]

 [  0 791]]

73.24074074074073
```

Figure 32

3.7 Retest of Random Forest Modeled Machine

The RFA is once again used for retesting the trained data. Below is the successful operation of the random forest algorithm model in classifying the effect of natural background radiation.

Input 29: Code to predict the effect of the natural background radiation

In [29]:

```
y_newpredict = RFC.predict([[0.95,0.560,0.960]])

y_newpredict
```

Figure 33

Output 28: Prediction of the effect of the natural background radiation

Out [28]:

```
array([1], dtype=int64)
```

Input 30: Code to predict the effect of the natural background radiation

In [30]:

```
y_newpredict = RFC.predict([[0.02,0.001,0.178]])

y_newpredict
```

Figure 34

Output 29: Prediction of the effect of the natural background radiation

Out [29]:

```
array([0], dtype=int64)
```

4. Discussion

The findings from this study using different machine learning algorithms for modelling and simulating background radiation datasets had made contributions as part of the widespread application of artificial intelligence and machine learning in the contemporary world. Using different machine algorithms which were random forest (RF), Navier-Bayes (NB), support vector machine (SVM) and kernel SVM (KSVM), the classification of the natural background radiation dataset was achieved and established at various accuracies. Each machine could identify and classify whether alpha, beta or gamma radiation was harmful or not. RF had the best accuracy so far. The training accuracy of the random forest was 100%, while NBA was just 67%, SVM was 68%, and KSVM was also 68%. The testing accuracies for the machine learning algorithms were; RF 96%, NB 68%, SVM 67% and KSVM 67% also. The accuracies were based on the algorithm's other factors namely “recall”, “precision” and “F1 score”. The random forest machine algorithm had the best and most efficient result. According to Rigatti, the random forest algorithm can identify, classify and predict [14]. The machine can visualize and measure the trend of harmfulness and harmlessness of the background radiation dataset. The findings from this research revealed there was radiation exposure in the Gwagwalada area, Abuja, Nigeria which may be harmful. The study was able to establish the applicable use of machine learning in radiological studies, and medical physics. The applicability of machine learning is related to Guillaume and colleagues’ study in which a neural network was used in building a physics-constrained stable machine learning-based radiation emulator [15]. Sarrut and colleagues also used deep learning, another type of machine learning to model Monte Carlo simulation for particle transport [16]. Their study was able to achieve and theoretically establish a Monte Carlo Simulation called the Technique Variance Reduction Technique, used to accelerate the processes of dose estimation [16]. Additionally, Hafermann and his colleagues. (2022), made use of a random forest machine to create a simulation called plasmode simulation [17]. The plasmode simulation was a simulation based on subsampling a dataset of about 200,000 individuals from a pharmacoepidemiologic study, used to specify patient outcomes [17]. Ivan used ANN to label ionizing radiations [18]. AI and ML were used to optimized glioma systems in medical science by Zhang and his colleagues [19] and Farouk and his colleagues also used machine learning for the enhancement of the accuracy of X-ray radiation [20]. Gachancipa used machine learning to build a computational model to detect radiation [21].

5. Conclusion

So far, each ML algorithm has demonstrated its capability and accuracy in classifying natural background radiation datasets as either harmful or harmless. In general, the training accuracy of random forest was 100% and had the most preferred result, while NBA was just 67%, SVM was also 68%, and KSVM was also 68%. Therefore, the testing accuracy for the machine learning algorithms are; RF was 96%, NB was 68%, SVM was 67% and KSVM was 67% also. The accuracies are based on the algorithm's other factors namely “recall”, “precision” and “F1 score”. In conclusion, the random forest machine learning algorithm has the highest percentage testing accuracy out of the four machine learning algorithms used to build the model. Therefore, random forest is highly recommended for classification analysis modelling for background radiation. This study is subjected to further studies in the area of prediction analysis, using the robust dataset, and future implementations.

References

[1] V. Kannade. "What Is Machine Learning? Definition, Types, Applications, and Trends for 2022." Spiceworks. https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-ml/ (accessed 18, March, 2024).

[2] J. R. Batley *et al.*, "A precision measurement of direct CP violation in the decay of neutral kaons into two pions.," *Physics Letters*, pp. 544(1-2):97-112, 2002.

[3] "Sources and effects of ionizing radiation," United Nations, New York, 2008.

[4] "(http://www.world-nuclear.org/info/Safety-and-Security/Radiation-and-Health/Naturally-Occurring-Radioactive-Materials-NORM," World Nuclear Association, 2019.

[5] D. Shahbazi-Gahrouei, M. Gholami, and S. Setayandeh, "A review on natural background radiation," *Advanced biomedical research*, vol. 2, no. 1, p. 65, 2013.

[6] Weisstein and W. Eric, "Radiation," *Eric Weisstein's World of Physics*, 2014.

[7] N. Kwan-Hoong, "Non-Ionizing Radiations – Sources, Biological Effects, Emissions and Exposures," in *Proceedings of the International Conference on Non-Ionizing Radiation at UNITEN (ICNIR2003)*, UNITEN, Malaysia, 2003: ICNIR, pp. 1-16.

[8] O. Bashir, A. Reigns, O. Unimke, and S. Mshelia, "Population dynamics to urban spaces needs in one of Africa’s largest cities: Abuja, the federal capital city of Nigeria," *I R E J*, vol. 4, no. 10, pp. 124-130, 2021.

[9] Y. Simon and A. Samuel, "Analysis of Land-Use Conflict between Farmers and Pastoralists in

Gwagwalada Area Council of Abuja, Nigeria," 07/29 2021.

[10] J. Chavaudra, "Radioactivity," in *Handbook of Radiotherapy Physics*: CRC Press, 2021, pp. Vol1: 15-Vol1: 24.

[11] "Radiation in Everyday Life," ed, 2023.

[12] N. Bassey, U. Kaigama, and A. Oluwasegun, "Radiometric Mapping of Song Area and Environs, Hawal Basement Complex, Northeast Nigeria," *International Journal of Science and Technology,* vol. 2, pp. 692-699, 10/01 2013.

[13] D. Kuhlman, "A Python Book: Beginning Python, Adanced." [Online]. Available: https://www.davekuhlman.org/python_book_01.pdf

[14] S. J. Rigatti, "Random forest," *Journal of Insurance Medicine,* vol. 47, no. 1, pp. 31-39, 2017.

[15] G. Bertoli, S. Schemm, F. Ozdemir, F. Perez Cruz, and E. Szekely, "Building a physics-constrained, fast and stable machine learning-based radiation emulator," in *EGU General Assembly Conference Abstracts*, 2023, pp. EGU-3418.

[16] D. Sarrut, A. Etxebeste, E. Muñoz, N. Krah, and J. M. Létang, "Artificial intelligence for Monte Carlo simulation in medical physics," *Frontiers in Physics,* vol. 9, p. 738112, 2021.

[17] L. Hafermann, N. Klein, G. Rauch, M. Kammer, and G. Heinze, "Using background knowledge from preceding studies for building a random forest prediction model: a plasmode simulation study," *Entropy,* vol. 24, no. 6, p. 847, 2022.

[18] L. Ivan, D. Hummel, and L. Lebel, "A MCREXS modelling approach for the simulation of a radiological dispersal device," *Journal of environmental radioactivity,* vol. 192, pp. 551-564, 2018.

[19] X. Zhang *et al.*, "Optimizing a machine learning based glioma grading system using multi-parametric MRI histogram and texture features," *Oncotarget,* vol. 8, no. 29, p. 47816, 2017.

[20] S. Farouk, A. M. Osman, S. M. Awadallah, and A. S. Abdelrahman, "The added value of using artificial intelligence in adult chest X-rays for nodules and masses detection in daily radiology practice," *Egyptian Journal of Radiology and Nuclear Medicine,* vol. 54, no. 1, p. 142, 2023.

[21] J. N. Gachancipa Parga, "Computational Models to Detect Radiation in Urban Environments," 2021.