Simulation Based on Garfield++ & Different Configurations to Train Long Short Term Memory Model by Using HPC Resources

Muhammad Numan Anwar
Department of Physics
Polytechnic University of Bari
INFN, Italy

Outline

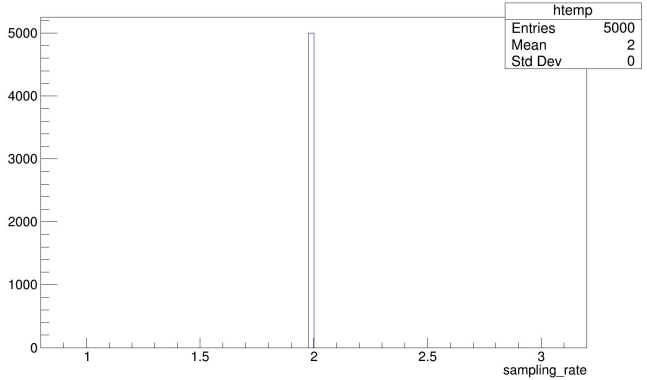
1. **Simulation based on Garfield++**
2. **Different Configurations for Training LSTM Model**
3. **Accepted Abstract at Bolgna Conference**
4. **Future Planning**

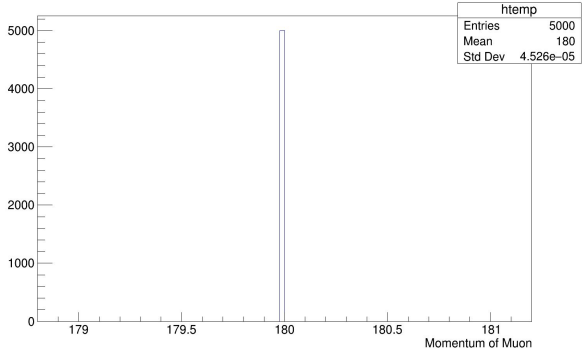
Simulation Based on Garfield ++

- **Muon beam is passed through mixture of gas having 90% He and 10% Isobutane C_4H_{10} by using a geometry of drift tubes mimicking what was used for the beam test at CERN in 2022**

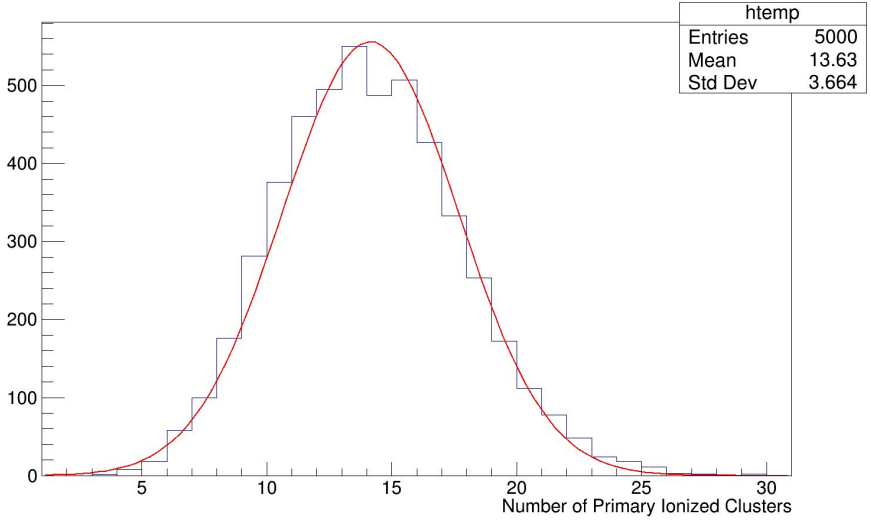
- **The simulation parameters are:**
  Cell Size of 0.8 cm
  Sampling rate of 2.0 GHz
  Time window 2000 ns
  Momentum of muon particles 180 GeV/c

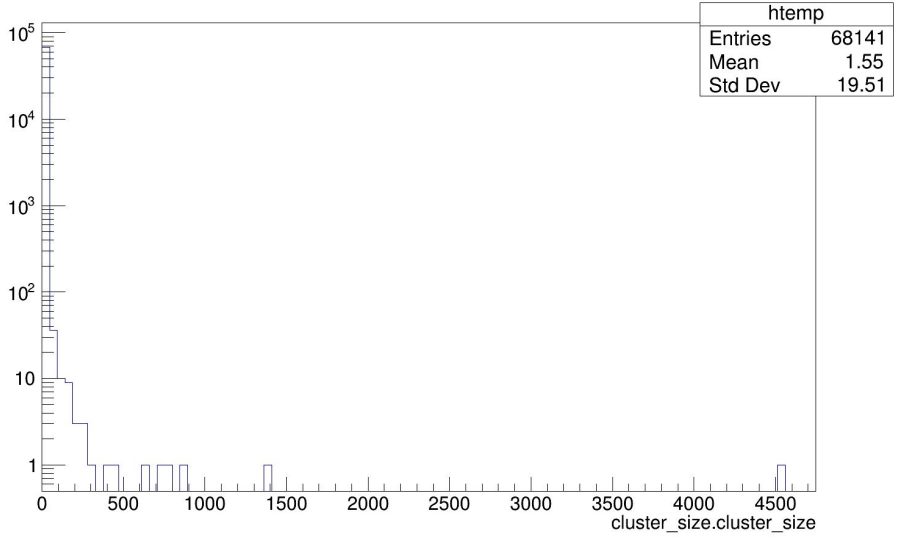
- **The simulation was conducted using Garfield++**





Simulation Based on Garfield ++





The above distribution shows the number of primary ionization clusters with mean value 13.63

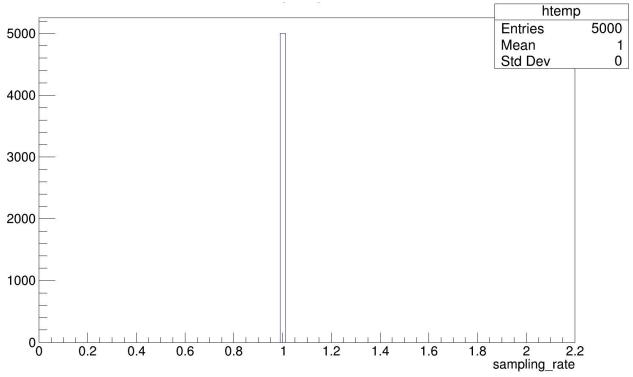
The above distribution shows the number of electrons per cluster with mean value 1.55

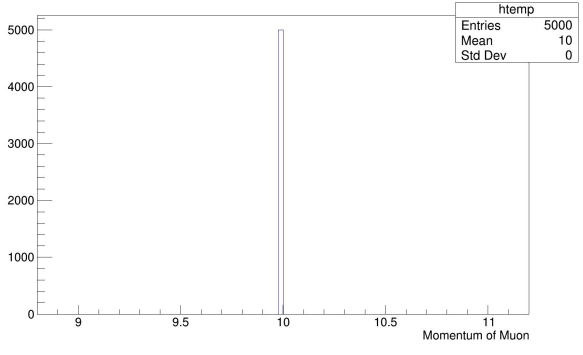
Simulation Based on Garfield ++

- **Muon beam is passed through mixture of gas having 90% He and 10% Isobutane C_4H_{10} by using a geometry of drift tubes mimicking what was used for the beam test at CERN in 2023**

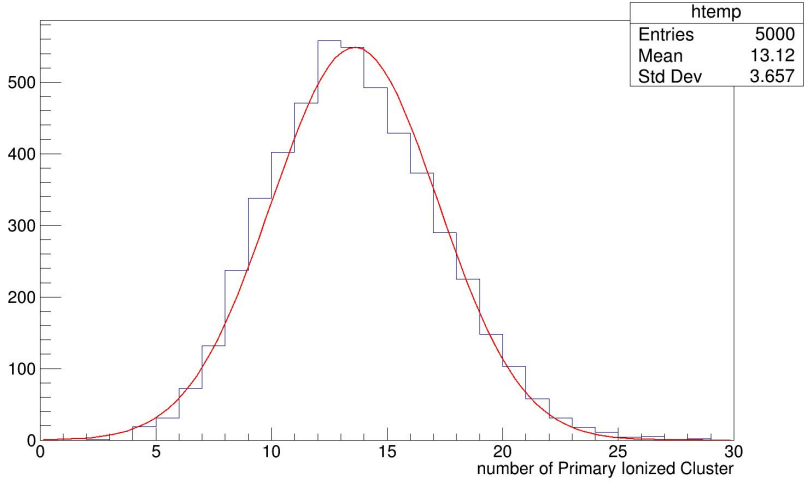
- **The simulation parameters are:**
  Cell Size of 0.8 cm
  Sampling rate of 1.0 GHz
  Time window 2000 ns
  Momentum of muon particles 10 GeV/c

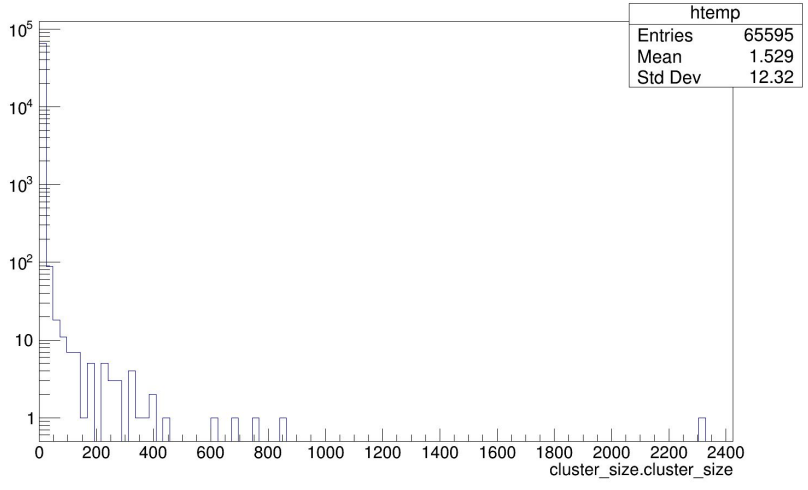
- **The simulation was conducted using Garfield++**





Simulation Based on Garfield ++





The above distribution shows the number of primary ionization clusters with mean value 13.12

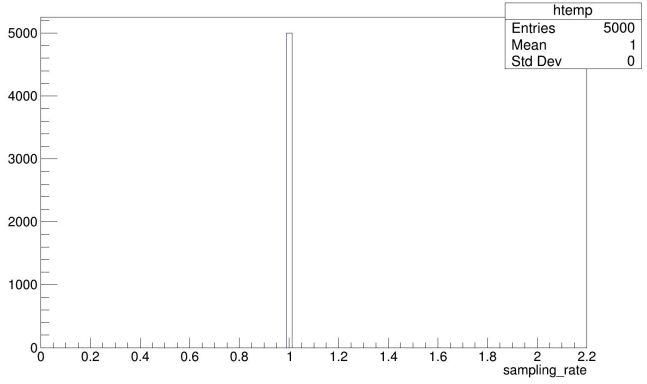
The above distribution shows the number of electrons per cluster with mean value 1.529

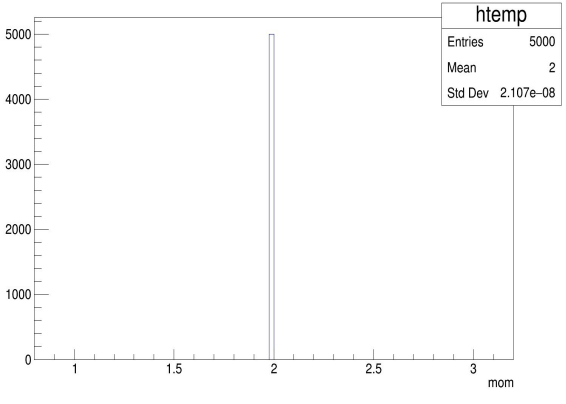
Simulation Based on Garfield ++

- **Muon beam is passed through mixture of gas having 90% He and 10% Isobutane C_4H_{10} by using a geometry of drift tubes mimicking what was used for the beam test at CERN in 2023**

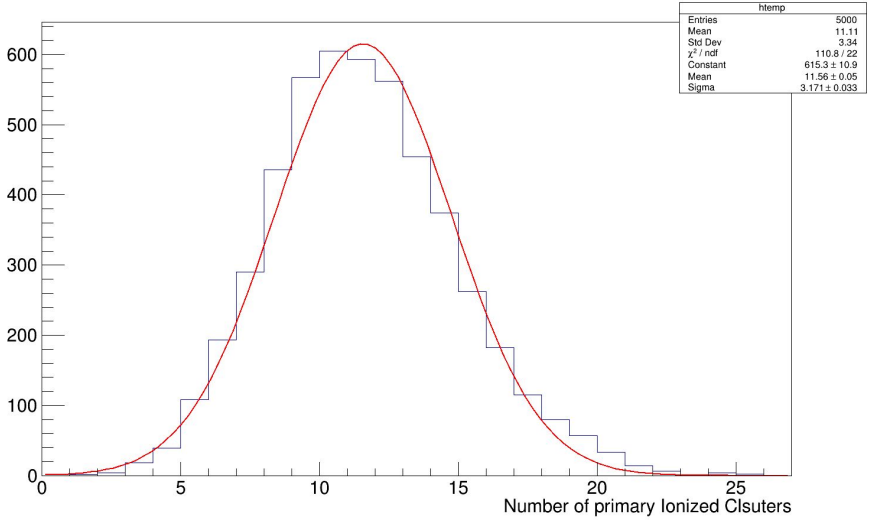
- **The simulation parameters are:**
  Cell Size of 0.8 cm
  Sampling rate of 1.0 GHz
  Time window 2000 ns
  Momentum of muon particles 2.0 GeV/c

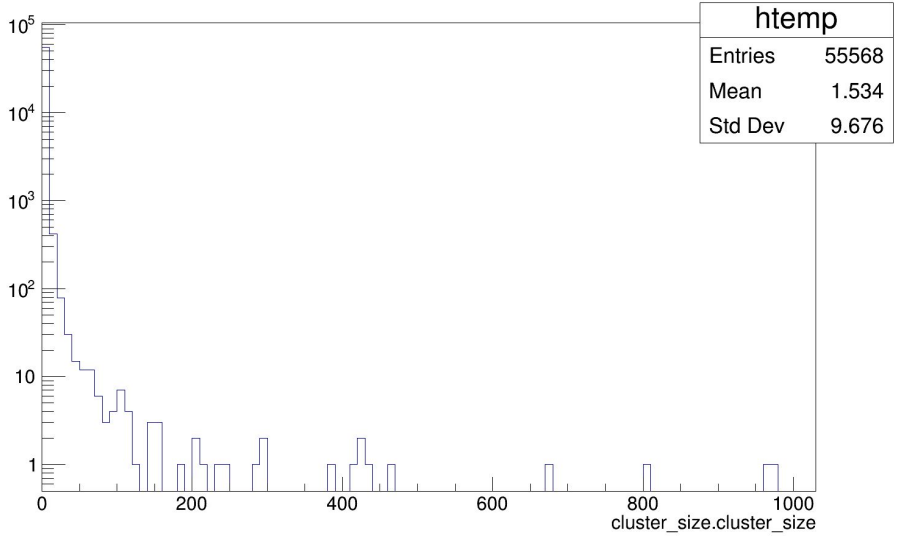
- **The simulation was conducted using Garfield++**





Simulation Based on Garfield ++





The above distribution shows the number of primary ionization clusters with mean value 11.11

The above distribution shows the number of electrons per cluster with mean value 1.534

Different Configuration to train LSTM Model by Using Different hyperparameters

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 96)	37632
flatten (Flatten)	(None, 96)	0
dense (Dense)	(None, 128)	12416
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 1)	129
dropout_1 (Dropout)	(None, 1)	0
dense_2 (Dense)	(None, 1)	2

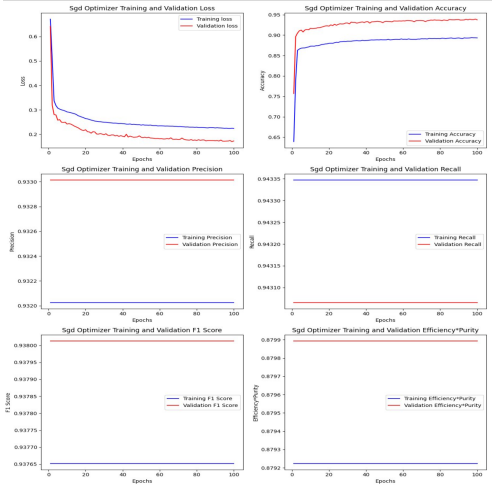
Total params: 50,179
Trainable params: 50,179
Non-trainable params: 0

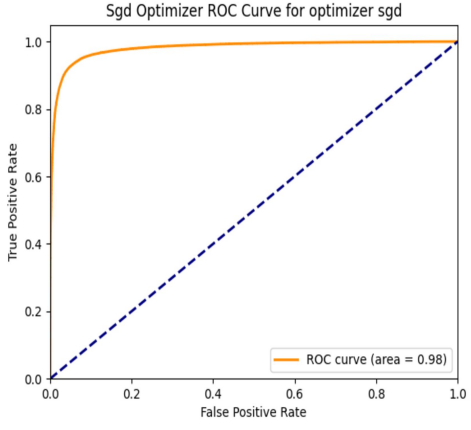
Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 32)	4352
flatten (Flatten)	(None, 32)	0
dense (Dense)	(None, 64)	2112
dropout (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 1)	65
dropout_1 (Dropout)	(None, 1)	0
dense_2 (Dense)	(None, 1)	2

Total params: 6,531
Trainable params: 6,531
Non-trainable params: 0

Random LSTM MODEL SUMMARY

Different Configuration to train LSTM Model by Using Different hyperparameters






```
Optimizer: sgd
Topology: [96, 128, 1]
Activation Functions: ['relu', 'selu']
Patience: [50]
Batch Size: [150]
Dropout Rates: [0.1, 0.1]
Train/Validation Split: 0.7
Number of Epochs: 100
Final Training Loss: 0.22435244917869568
Final Validation Loss: 0.1728692650794983
Final Training Accuracy: 0.8931867480278015
Final Validation Accuracy: 0.9374743700027466
Validation Accuracy: 0.9374743621297892
Precision: 0.9330137734360779
Recall: 0.9430656635865565
F1 Score: 0.9380127898819631
AUC Score: 0.98
Efficiency*Purity: 0.879893253380892
----------
```

- Different plots related to the Evaluation metrics over the epochs are shown above. This Evaluation metrics tell us about the best model by using different hyperparameters

- Area Under the Curve (AUC) value from Receiver Operating Characteristic (ROC) Curve is 0.98. Higher the value of AUC , best model would be consider fro the classification task

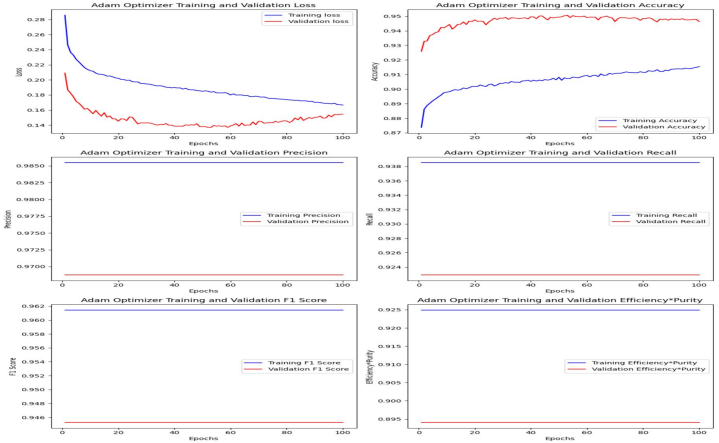
- Single Value of Mentioned Evaluation Metrics and Hyperparameters in .txt file

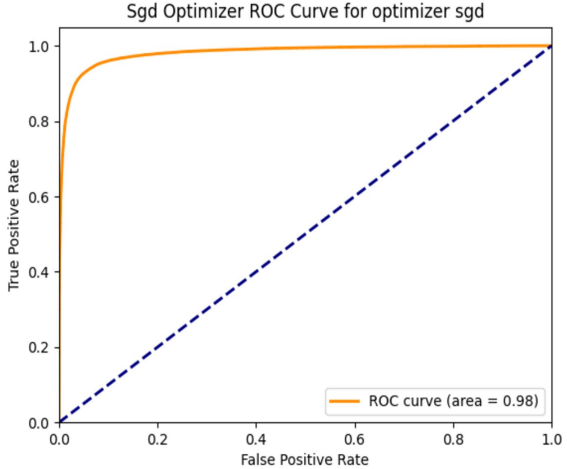
Different Hyperparameters, Evaluation Metrics and HPC Resoursources in Table

Optimizer	**sgd**	**Precision**	**0.93301**
Topology	**[96, 128, 1]**	**Recall**	**0.94306**
Bach size	**[150]**	**F1 Score**	**0.9380**
Train/Validation Split	**0.7**	**AUC Score**	**0.98**
Patience	**[50]**	**Efficiency*Purity**	**0.87**
Dropout Rate	**[0.1, 0.1]**	**Final Validation Accuracy**	**0.937474**
Number of Epochs	**100**	**Final Training Accuracy**	**0.89318**
Final Training loss	**0.22435**	**Activation function**	**Relu, Selu**
Final Validation loss	**0.17286**		

Partionable Resources	**Usage**	**Request**	**Allocated**
CPUS	**1.92**	**4**	**4**
Disk (KB)	**1**	**1**	**857735**
Memory (MB)	**925**	**5000**	**5120**
Run Remote Usage	**12 min 59 sec**	**2hr/job**	**//**

Different Configuration to train LSTM Model by Using Different hyperparameters






```
Optimizer: adam
Topology: [96, 128, 1]
Activation Functions: ['relu', 'selu']
Patience: [50]
Batch Size: [150]
Dropout Rates: [0.1, 0.1]
Train/Validation Split: 0.7
Number of Epochs: 100
Final Training Loss: 0.16662417352199554
Final Validation Loss: 0.15447402000427246
Final Training Accuracy: 0.9154653549194336
Final Validation Accuracy: 0.9463963508605957
Validation Accuracy: 0.9463963409631635
Precision: 0.968756035278213
Recall: 0.9229086597432333
F1 Score: 0.9452767570170756
AUC Score: 0.98
Efficiency*Purity: 0.894073334136784
----------
```

- Different plots related to the Evaluation metrics over the epochs are shown above. This Evaluation metrics tell us about the best model by using different hyperparameters

- Area Under the Curve (AUC) value from Receiver Operating Characteristic (ROC) Curve is 0.98. Higher the value of AUC , best model would be consider fro the classification task

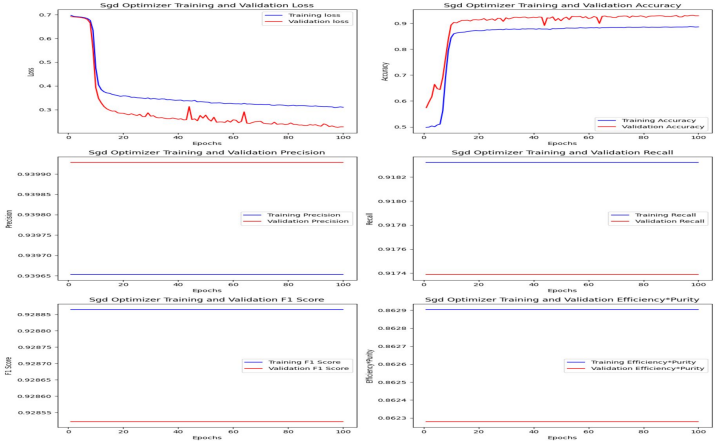
- Single Value of Mentioned Evaluation Metrics and Hyperparameters in .txt file

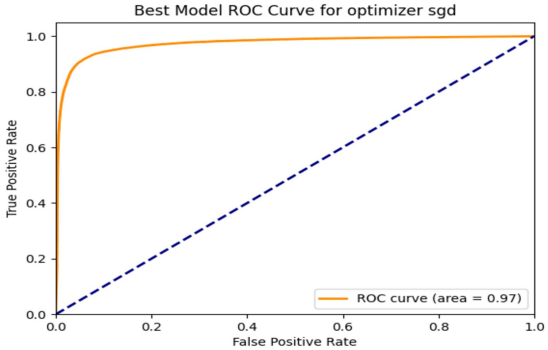
Different Hyperparameters, Evaluation Metrics and HPC Resoursources in Table

Optimizer	**adam**	**Precision**	**0.9687**
Topology	**[96, 128, 1]**	**Recall**	**0.9229**
Bach size	**[150]**	**F1 Score**	**0.9380**
Train/Validation Split	**0.7**	**AUC Score**	**0.98**
Patience	**[50]**	**Efficiency*Purity**	**0.8940**
Dropout Rate	**[0.1, 0.1]**	**Final Validation Accuracy**	**0.9463**
Number of Epochs	**100**	**Final Training Accuracy**	**0.9154**
Final Training loss	**0.1666**	**Activation function**	**Relu, Selu**
Final Validation loss	**0.15447**		

Partionable Resources	**Usage**	**Request**	**Allocated**
CPUS	**1.94**	**4**	**4**
Disk (KB)	**1**	**1**	**857737**
Memory (MB)	**759**	**5000**	**5120**
Run Remote Usage	**9 min 16sec**	**2hr/job**	**//**

Different Configuration to train LSTM Model by Using Different hyperparameters






```
Optimizer: sgd
Topology: [96, 128, 1]
Activation Functions: ['relu', 'sigmoid']
Patience: [50]
Batch Size: [150]
Dropout Rates: [0.1, 0.1]
Train/Validation Split: 0.7
Number of Epochs: 100
Final Training Loss: 0.31019169092178345
Final Validation Loss: 0.22750797867774963
Final Training Accuracy: 0.8861634731292725
Final Validation Accuracy: 0.9291471838951111
Validation Accuracy: 0.9291471818853064
Precision: 0.9399283664620992
Recall: 0.9173889933764003
F1 Score: 0.9285219172555066
AUC Score: 0.97
Efficiency*Purity: 0.8622799379545896
----------
```

- Different plots related to the Evaluation metrics over the epochs are shown above. This Evaluation metrics tell us about the best model by using different hyperparameters

- Area Under the Curve (AUC) value from Receiver Operating Characteristic (ROC) Curve is 0.97. Higher the value of AUC , best model would be consider fro the classification task

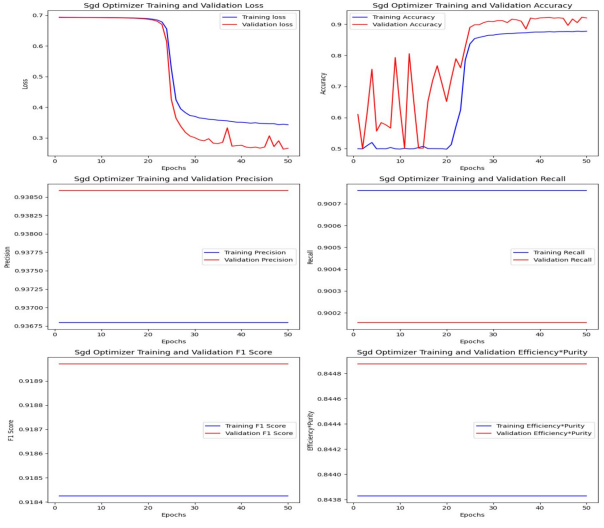
- Single Value of Mentioned Evaluation Metrics and Hyperparameters in .txt file

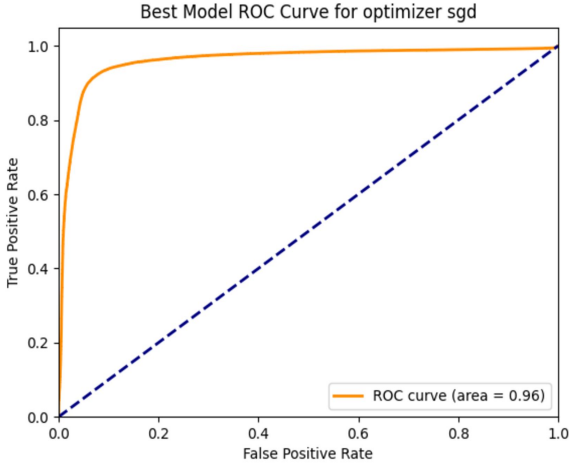
Different Hyperparameters, Evaluation Metrics and HPC Resoursources in Table

Optimizer	sgd	Precision	0.9399
Topology	[96, 128, 1]	Recall	0.9173
Bach size	[150]	F1 Score	0.9285
Train/Validation Split	0.7	AUC Score	0.97
Patience	[50]	Efficiency*Purity	0.86
Dropout Rate	[0.1, 0.1]	Final Validation Accuracy	0.8861
Number of Epochs	100	Final Training Accuracy	0.9291
Final Training loss	0.310	Activation function	Relu, Sigmoid
Final Validation loss	0.2275		

Partionable Resources	Usage	Request	Allocated
CPUS	1.81	4	4
Disk (KB)	1	1	857736
Memory (MB)	773	5000	5120
Run Remote Usage	8 min 14sec	2hr/job	//

Different Configuration to train LSTM Model by Using Different hyperparameters





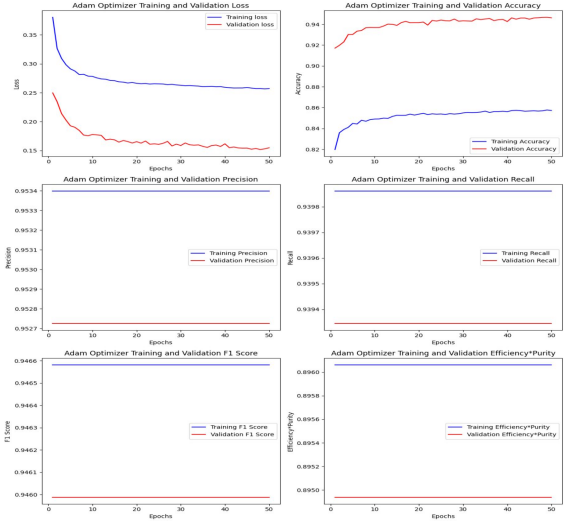

```
Optimizer: sgd
Topology: [32, 64, 1]
Activation Functions: ['relu', 'sigmoid']
Patience: [50]
Batch Size: [150]
Dropout Rates: [0.1, 0.1]
Train/Validation Split: 0.7
Number of Epochs: 50
Final Training Loss: 0.34286627173423767
Final Validation Loss: 0.266116738319397
Final Training Accuracy: 0.8774964213371277
Final Validation Accuracy: 0.9203687906265259
Validation Accuracy: 0.9203687751251128
Precision: 0.9385884509624198
Recall: 0.9001553683866219
F1 Score: 0.9189702490895241
AUC Score: 0.96
Efficiency*Purity: 0.8448754328395058
----------
~
~
```

- Different plots related to the Evaluation metrics over the epochs are shown above. This Evaluation metrics tell us about the best model by using different hyperparameters

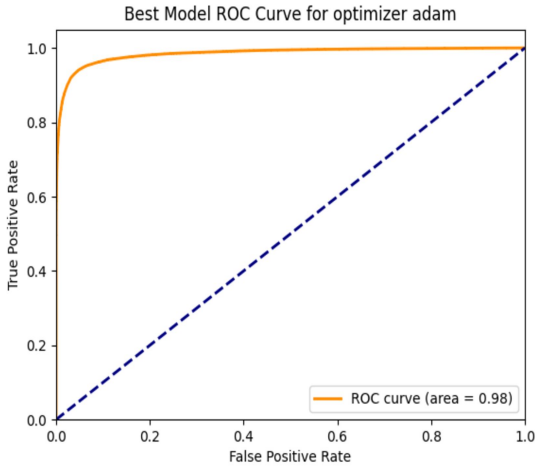
- Area Under the Curve (AUC) value from Receiver Operating Characteristic (ROC) Curve is 0.96. Higher the value of AUC , best model would be consider fro the classification task

- Single Value of Mentioned Evaluation Metrics and Hyperparameters in .txt file

Different Configuration to train LSTM Model by Using Different hyperparameters



- Different plots related to the Evaluation metrics over the epochs are shown above. This Evaluation metrics tell us about the best model by using different hyperparameters



- Area Under the Curve (AUC) value from Receiver Operating Characteristic (ROC) Curve is 0.98. Higher the value of AUC , best model would be consider fro the classification task

```
Optimizer: adam
Topology: [32, 64, 1]
Activation Functions: ['relu', 'selu']
Patience: [50]
Batch Size: [150]
Dropout Rates: [0.2, 0.2]
Train/Validation Split: 0.7
Number of Epochs: 50
Final Training Loss: 0.2571702301502228
Final Validation Loss: 0.15482334792613983
Final Training Accuracy: 0.8572527170181274
Final Validation Accuracy: 0.9461912512779236
Validation Accuracy: 0.9461912380014768
Precision: 0.9527255385763752
Recall: 0.9393449995911358
F1 Score: 0.9459879561480262
AUC Score: 0.98
Efficiency*Purity: 0.8949379706444899
----------
```

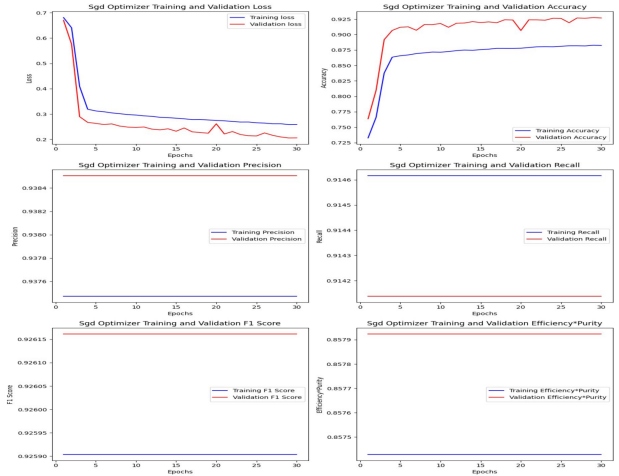
- Single Value of Mentioned Evaluation Metrics and Hyperparameters in .txt file

Different Hyperparameters, Evaluation Metrics and HPC Resoursources in Table

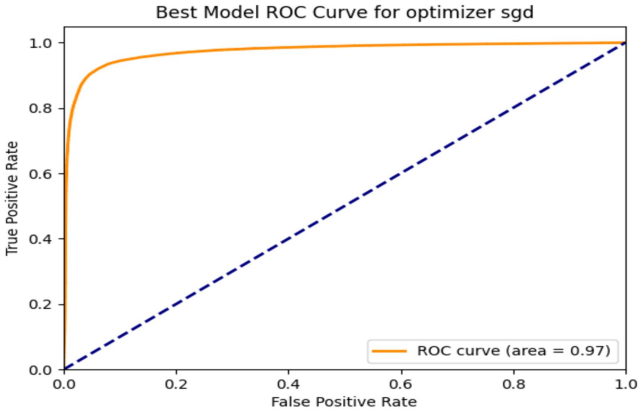
Optimizer	**adam**	**Precision**	**0.9527**
Topology	**[32, 64, 1]**	**Recall**	**0.93934**
Bach size	**[150]**	**F1 Score**	**0.9459**
Train/Validation Split	**0.7**	**AUC Score**	**0.98**
Patience	**[50]**	**Efficiency*Purity**	**0.8949**
Dropout Rate	**[0.2, 0.2]**	**Final Validation Accuracy**	**0.9461**
Number of Epochs	**50**	**Final Training Accuracy**	**0.8572**
Final Training loss	**0.2571**	**Activation function**	**Relu, Selu**
Final Validation loss	**0.1548**		

Partionable Resources	**Usage**	**Request**	**Allocated**
CPUS	**1.73**	**4**	**4**
Disk (KB)	**1**	**1**	**419696**
Memory (MB)	**7815**	**5000**	**5120**
Run Remote Usage	**2 min 53 sec**	**2hr/job**	**//**

Different Configuration to train LSTM Model by Using Different hyperparameters



- Different plots related to the Evaluation metrics over the epochs are shown above. This Evaluation metrics tell us about the best model by using different hyperparameters

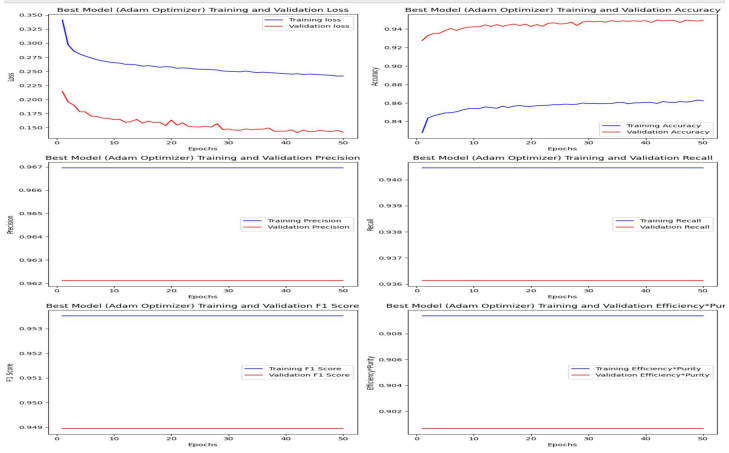


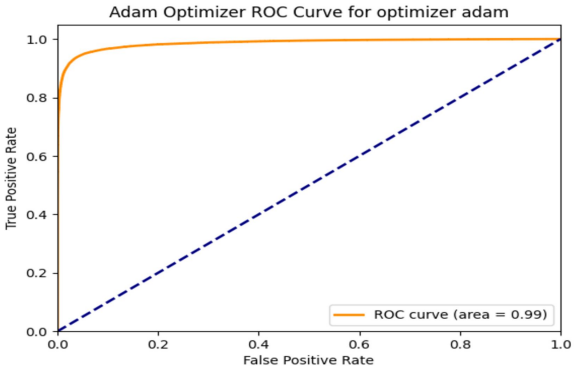
- Area Under the Curve (AUC) value from Receiver Operating Characteristic (ROC) Curve is 0.97. Higher the value of AUC , best model would be consider fro the classification task

```
Optimizer: sgd
Topology: [32, 64, 1]
Activation Functions: ['relu', 'selu']
Patience: [50]
Batch Size: [150]
Dropout Rates: [0.1, 0.1]
Train/Validation Split: 0.7
Number of Epochs: 30
Final Training Loss: 0.258718341588974
Final Validation Loss: 0.20604060590267181
Final Training Accuracy: 0.8823309540748596
Final Validation Accuracy: 0.9268807768821716
Validation Accuracy: 0.9268807941586676
Precision: 0.9385048062796457
Recall: 0.9141385231825987
F1 Score: 0.9261614299620968
AUC Score: 0.97
Efficiency*Purity: 0.8579233976122462
----------
```

- Single Value of Mentioned Evaluation Metrics and Hyperparameters in .txt file

Best Model Based on the Higher Efficiency* Purity, F1 Score and Recall Value






```
Optimizer: adam
Topology: [96, 128, 1]
Activation Functions: ['relu', 'selu']
Patience: [50]
Batch Size: [150]
Dropout Rates: [0.2, 0.2]
Train/Validation Split: 0.7
Number of Epochs: 50
Final Training Loss: 0.24181483685970306
Final Validation Loss: 0.1418175995349884
Final Training Accuracy: 0.862377405166626
Final Validation Accuracy: 0.9494729042053223
Validation Accuracy: 0.949472885388465
Precision: 0.9621178695241097
Recall: 0.9361354158148663
F1 Score: 0.948948824486328
AUC Score: 0.99
Efficiency*Purity: 0.9006726118498657
```

- Different plots related to the Evaluation metrics over the epochs are shown above. This Evaluation metrics tell us about the best model by using different hyperparameters

- Area Under the Curve (AUC) value from Receiver Operating Characteristic (ROC) Curve is 0.99. Higher the value of AUC , best model would be consider fro the classification task

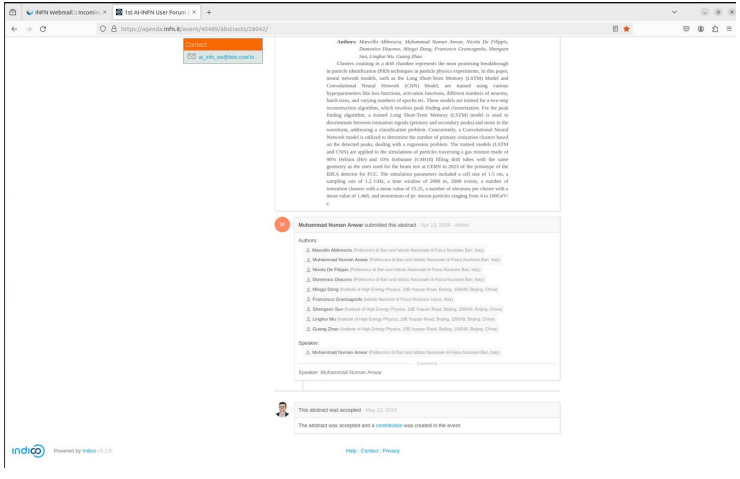
Single Value of Mentioned Evaluation Metrics and Hyperparameters in .txt

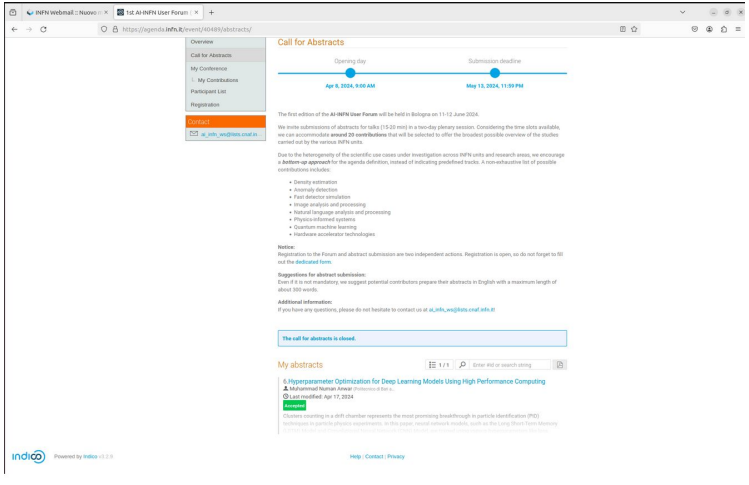
Different Hyperparameters, Evaluation Metrics and HPC Resoursources in Table

Optimizer	**adam**	**Precision**	**0.93301**
Topology	**[96, 128, 1]**	**Recall**	**0.94306**
Bach size	**[150]**	**F1 Score**	**0.9380**
Train/Validation Split	**0.7**	**AUC Score**	**0.98**
Patience	**[50]**	**Efficiency*Purity**	**0.90**
Dropout Rate	**[0.2, 0.2]**	**Final Validation Accuracy**	**0.94947**
Number of Epochs	**50**	**Final Training Accuracy**	**0.862377**
Final Training loss	**0.2418**	**Activation function**	**Relu, Selu**
Final Validation loss	**0.1418**		

Partionable Resources	**Usage**	**Request**	**Allocated**
CPUS	**1.76**	**4**	**4**
Disk (KB)	**1**	**1**	**419696**
Memory (MB)	**830**	**5000**	**5120**
Run Remote Usage	**8 min 39 sec**	**2hr/job**	**//**

Accepted Abstract at Bolognia Conference





The submitted abstract related to “Hyperparameter Optimization for Deep Learning Models Using High Performance Computing” has been accepted at AI INF Forum Bolognia Conference

Future Planning

- **I would repeat the above task for all the possible hyperparameters and then select the best model based on the highest accuracy, F1 score, Recall. Efficeincy*Purity and highest AUC value among all different configurations for training the LSTM model.**

- **Then, we will apply the trained models to the real beam test data to classify signals from noise in the waveform and determine the number of primary clusters based on the detected peaks**

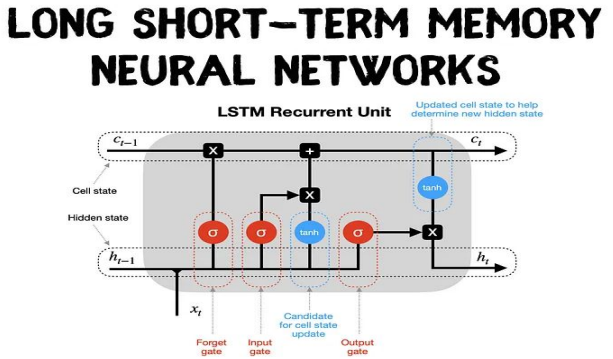
Thank you

ACCURACY and LSTM

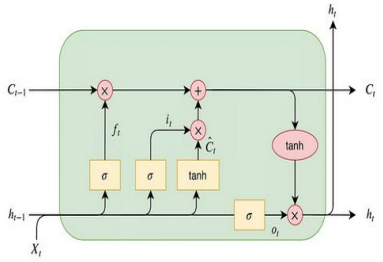
- The accuracy is defined as the ratio between the number of correct predictions to the total number of predictions
- Accuracy values range between 0 and 1. Obviously an accuracy values near to 1 means that our model fits well the datasets

$$Accuracy = \frac{True_{positive} + True_{negative}}{True_{positive} + True_{negative} + False_{positive} + False_{negative}}$$

- Forget Gate: This gate determines what information from the previous cell state should be forgotten or retained.
- Input Gate: It controls what new information should be stored in the cell state.
- Output Gate: This gate defines the output of the LSTM cell, considering the current input and the updated cell state

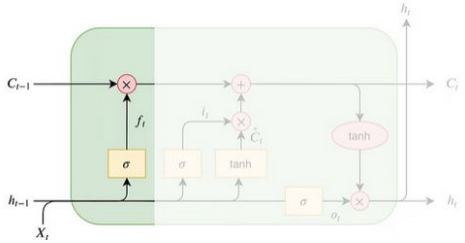


Long SHort Term Memory (LSTM)

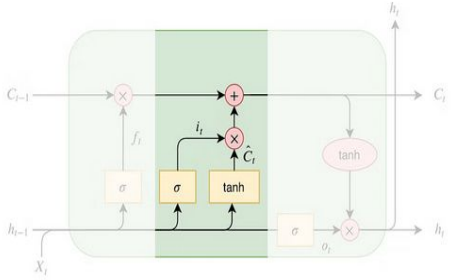


Forget Gate

state and the new input data.



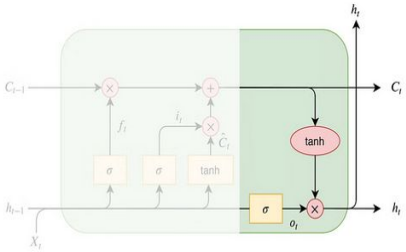
Input Gate



$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i)$$

Output Gate



$$o_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o)$$

$$\hat{C}_t = \tanh(W_C \cdot [h_{t-1}, X_t] + b_C)$$

$$C_t = i_t \cdot \hat{C}_t + f_t \cdot C_{t-1}$$

EXAMPLES of LOSS FUNCTIONS

- Mean Squared Error(MSE)/ Quadratic Loss/ L2:

$$MSE(y^{(i)}, y_{pred}^{(i)}) = \frac{\left(y^{(i)} - y_{pred}^{(i)}\right)^2}{n}$$

- Mean Absolute Error (MAE)/ L1 Loss:

$$MAE(y^{(i)}, y_{pred}^{(i)}) = \frac{\left|y^{(i)} - y_{pred}^{(i)}\right|}{n}$$

- Mean Bias Error (MBE):

$$MBE(y^{(i)}, y_{pred}^{(i)}) = \frac{\left(y^{(i)} - y_{pred}^{(i)}\right)}{n}$$

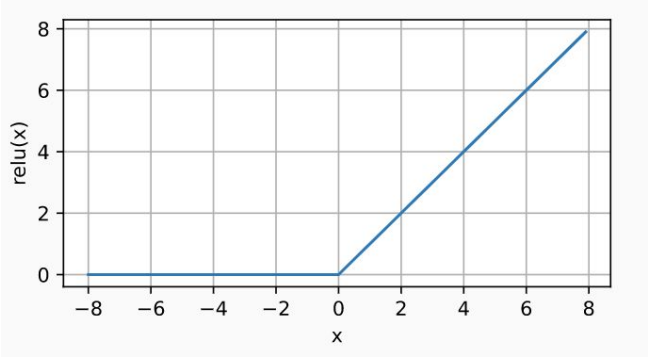
NUMBER OF EPOCHS

- **Epoch**: In terms of artificial neural networks, an epoch refers to one cycle through the full training dataset
- Number of epochs is a delicate choice:
  - ❏ A large number of epochs can induce our model to an overfitting problem
  - ❏ Too small number of epochs can lead to an under fitting problem
- To avoid a wrong choice we can use the ' EarlyStopping', also implemented by Keras:
  - ❏ It allows to stop the training when a monitor (set by us and tipically the loss function) has stopped improving.

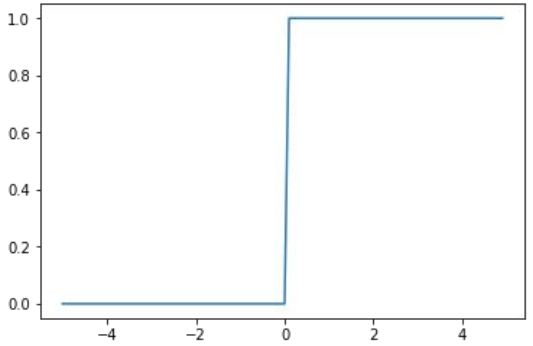
RECTIFIED LINEAR UNIT (RELU)

- One the most popular non-linear activation function is the REctified Linear Unit (ReLU)
- It provides a non-linear transformation and returns the max value between the input x (the argument) and 0
- The ReLU function is also differentiable in as given below:

$$\frac{dReLU(x)}{dx} = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$$



$$ReLU(x) = max(0, x)$$



SCALED EXPONENTIAL LINEAR UNIT (SELU)

• Another choice is the Scaled Exponential Linear Unit (SELU)

• The functions depends on two parameters and the equation is the following:

$$SELU(x) = \lambda \begin{cases} \alpha(e^x - 1) & x \leq 0 \\ x & x > 0 \end{cases}$$

- The function is not differentiable in zero

$$\frac{dSELU(x)}{dx} = \lambda \begin{cases} \alpha e^x & x \leq 0 \\ 1 & x > 0 \end{cases}$$