Implementation of AI frameworks for S/B Analysis of tZ FCNC process

Saqlain, A.*[1] and Kartal, S.[1]
[1] İstanbul University, Türkiye
* a.saqlain@ogr.iu.edu.tr

INTRODUCTION

This study focuses on the training and implementation of a supervised AI framework in order to Identify, classify and distinguish between processes simulated at the generator level using Madgraph5_aMC@nlo generator. We aim to train the model to identify the signal and background by going through their respective lhe files and be able to suggest cuts, if required, to separate signals from background events. We compared the results of the AI to human results [1] on tZ FCNC searches. Our results conclude, though over a limited population, that the S/B ratio provided by the AI model was quite accurate, even comparable to the cuts implemented by classical methods to extract signal (neglecting the detector deficiencies).

METHODOLOGY

> Event files for the process was generated from Madgraph5.
> AI was trained to recognize and label these files one by one based on the pdg id of particles.
> A mixed file of all events was provided to AI and it was instructed to separate all the events of a specific process which it did.
> After getting familiar enough with the processes, AI was asked to provide cuts in order to separate tZ process in the best way from the other processes in the combined file.

ANALYSIS

Do keep in mind that this study is still ongoing and this is simply an initial step regarding the separation of tZ signal from the backgrounds.

Cuts proposed by Dr.Yao-Bei Liu (as control) in his study [1]:

1. Select 3 leptons with pT > 30 GeV and one b jet with pT > 40 GeV
2. Select $|M(l_1\ l_2) - m_z| < 15$ GeV
3. Select 100 GeV < $M_T(t)$ < 200 GeV

Cuts proposed by the AI model for the tZ process:

1. Select number of leptons as 3
2. Select 85 < M(l+ l-) < 97
3. Select events with 1 b-quark
4. Select 30 < MET <150
5. Select DeltaR(l+,l-) < 2.5

Table 1: Events of Signal and Background kept after every Cut proposed by Researcher

Cuts	Signals	Backgrounds			
	tZ	$t\bar{t}$	$t\bar{t}W$	$t\bar{t}Z$	WZ
No Cuts	2210	32284	39	6	3960
Cut 1	705	0	15	0	0
Cut 2	572	0	2	0	0
Cut 3	294	0	1	0	0

Table 2: Events of Signal and Background kept after every Cut proposed by AI

Cuts	Signals	Backgrounds			
	tZ	$t\bar{t}$	$t\bar{t}W$	$t\bar{t}Z$	WZ
No Cuts	2210	32284	39	6	3960
Cut 1	2210	0	39	0	3960
Cut 2	2106	0	17	0	3746
Cut 3	1721	0	17	0	0
Cut 4	1220	0	14	0	0
Cut 5	1085	0	11	0	0

WHY IS IT SIGNIFICANT?

With more advancement in ML and the enormous amount of data produced during LHC runs, it calls for more effecient ways to perform data analysis. The main motivation of this study was simply this, being the first step in developing an AI framework which would provide a better Analysis and can be easily trained on simulation data to assess the real data generated by LHC runs.

Morever, the main aim is the development of an AI framework that is able to apply cuts effectively on a number of processes, not just one and also has the capability to get significantly better in its analysis capabilities if trained on only one set of data.

Even with a very limited set of data to train the model, the AI produced significant results when it came to the S/B effeciency, only falling short of complete removal of background by a factor of 10 events.

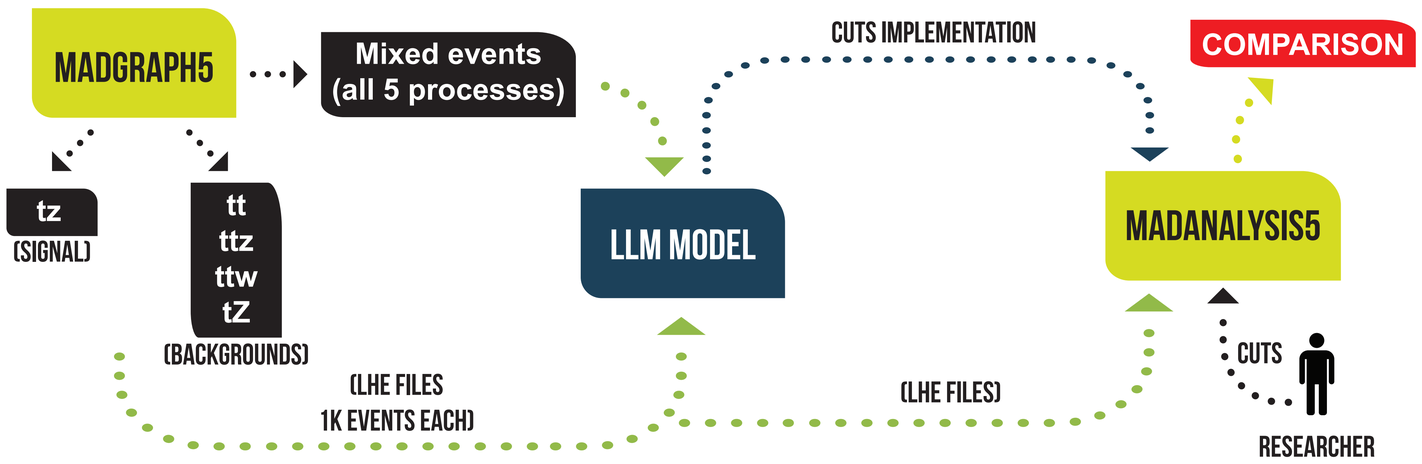
Table 3: Signal vs background efficiency after every cut proposed by Researcher

Cuts	Signal	Background	S/B
No Cuts	2210	36288	11.263
Cut 1	705	15	26.273
Cut 2	572	2	23.874
Cut 3	294	1	17.128

Table 4: Signal vs background efficiency after every cut proposed by AI

Cuts	Signal	Background	S/B
No Cuts	2210	36288	11.263
Cut 1	2210	3999	28.046
Cut 2	2106	3763	27.49
Cut 3	1721	17.61	41.274
Cut 4	1220	14	34.729
Cut 5	1085	11	34.773

WORK FLOW



FUTURE PROSPECTS

> Training the AI model on a large set of data, preferably starting with SM LHC processes.

> Application of this model on open data from LHC runs to improve its effeciency.

> Production of a base trainable model for researchers.

REFERENCES

[1] Yao-Bei Liu and Stefano Moretti 2021 Chinese Phys. C 45 043110