

# Mitigating the effect of fake missing energy using Machine learning technique in the ATLAS experiment

K G Tomiwa<sup>1\*</sup> Xifeng Ruan<sup>1</sup> Bruce Mellado<sup>1</sup>

<sup>1</sup> School of Physics and Institute for Collider Particle Physics, University of the Witwatersrand, Johannesburg, Wits 2050, South Africa.

E-mail: [kehinde.gbenga.tomiwa@cern.ch](mailto:kehinde.gbenga.tomiwa@cern.ch)

## Abstract.

The missing transverse momentum in the ATLAS experiment is the momentum imbalance in the plane transverse to the beam axis. That is the resultant of the negative vectorial sum of the momenta of all particles that are involved in the proton-proton collision. A precise measurement of the missing transverse energy is essential for many physics studies at the LHC, such as Higgs boson measurements and dark matter search. The result presented in this study are from the implementation of Boosted Decision Tree (BDT) based on vertex variables and fake/real missing samples. The preliminary results show the BDT classifiers can improve signal purity to about 50% as compared to the nominal selection.

## 1. Introduction

The Standard Model (SM) of particle physics has been very successful in the predictability of new particles that have been confirmed by many experiments. Despite its success, the SM is not perfect because the theory holds many unexplained phenomena like its inability to explain gravity, the reason for the mass of neutrinos, dark matter, and dark energy. A number of theoretical efforts to explain some of these phenomena have been published in terms of physics beyond the standard model. One of these theories is the heavy scalar model, developed by the high energy physics group of the University of the Witwatersrand [1].

The heavy scalar model resulted from the explanation of the discrepancies seen in the ATLAS and CMS observed data and the standard model prediction, Ref [2] gives a detailed list of excess used for the postulation. The model postulated a heavy scalar particle with the mass range within twice the mass of Higgs boson and top quark ( $2m_h < m_H < 2m_t$ , where  $H$  is the heavy scalar particle and  $h$  is the Higgs boson). Ref [1] gives a comprehensive list of decay modes of the heavy scalar particle, one of which is the associated production of Higgs boson and a scalar mediator ( $H \rightarrow hS$ ), the mediator scalar mostly decays to dark matter particle. This gives the possibilities of searching for the heavy scalar particle in a various Higgs final state with an additional requirement of missing transverse energy ( $E_T^{miss}$ ).

The heavy scalar particle was first searched in the Higgs to di-photon and missing transverse energy [3]. Since the analysis is heavily based on the reconstruction of missing transverse energy, a considerable effort was put into studying real and fake missing transverse energy with the ATLAS experiment. In Run 2, it was discovered that the jet vertex tagger algorithm



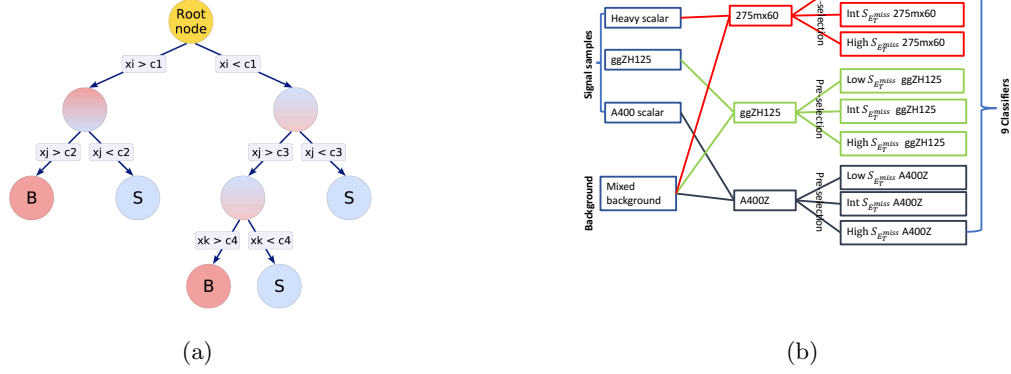


Figure 2: (a) Schematic view of a decision tree. (b) BDT classifiers by categories

Table 1: Signal and Background dataset used for MVA training

Acronym	Description
275mx60	Heavy scalar decaying to one Higgs boson in association with dark matter (Real $E_T^{miss}$ from dark matter candidate).
ggZH125	$Z$ boson produced from gluon-gluon fusion (Real $E_T^{miss}$ from neutrinos produced by $Z$ decay).
A400Z	heavy pseudo-scalar particle decaying to $Z$ boson and heavy scalar (Real $E_T^{miss}$ from neutrinos produced by $Z$ decay and dark matter candidate produced heavy scalar decay).
Mixed background	$\gamma\gamma$ , $\gamma - Jet$ and W/Z background sample

### 3. BDT samples and Categorization

Table 1 gives the acronyms and description of signal and background samples used for this analysis. All the signal samples are normalized to their corresponding cross-section and each samples of the combined background sample is scaled so that they can describe the Higgs side-band.

Classifiers are formed from the combination of the mixed background sample with each of the signal sample. Each combination of samples are categorized by the preselection in Table 2, this results in nine different classifiers, Figure 2b is a schematic description of how the classifiers are formed. The events are binned by missing transverse energy significance ( $S_{E_T^{miss}}$ ) and the number of central jets,  $S_{E_T^{miss}}$  is defined as the ratio of  $E_T^{miss}$  and square sum of transverse energy of all particle detected.

#### 3.1. BDT Training and hyper-parameters

The entire dataset is divided into test-set and train-set in 50 to 50 ratio. The trainset is used to model development and training while the test-set is used for evaluation and BDT performance measurement. The following BDT hyper-parameters are used to configure decision trees:

**Number of trees (NTrees):** Number of trees in the forest ( $T$ ). It represents the total number of aggregation in the boosted decision tree. NTrees is set to 800 for this study.

Table 2: Table showing BDT categories.

BDT cat	Pre-Selection
Low $S_{E_T^{miss}}$	$S_{E_T^{miss}} > 2.5 \ \&\& \ S_{E_T^{miss}} < 3.5 \ \sqrt{GeV} \ \&\& \ N_{jet} \geq 1$
Int $S_{E_T^{miss}}$	$S_{E_T^{miss}} > 3.5 \ \&\& \ S_{E_T^{miss}} < 5.5 \ \sqrt{GeV} \ \&\& \ N_{jet} \geq 1$
High $S_{E_T^{miss}}$	$S_{E_T^{miss}} > 5.5 \ \sqrt{GeV} \ \&\& \ N_{jet} \geq 1$

**Minimum node size:** The minimum percentage of training event required in a leaf node. This is one of the stop criteria of a decision tree.

**Shrinkage:** The learning rate ( $g$ ), see above description for more details,  $g$  is set to 0.06.

**Number of cuts:** The number of grid points in a variable range used in finding the optimal cut in node splitting.

**Maximum depth:** The maximum tree depth allowed, Tree depth is defined as the length of the longest path from the tree root to a leaf. the root node is considered to have a depth of 0.

### 3.2. Variables

The variables below are used for this MVA study, they include photon and jet kinematics and the number of interaction per bunch crossing.

- **Photon pointing Vertex  $\sum p_T^2$ :** Photon pointing vertex sum  $p_T$  squared.
- **Pile-up Vertex  $\sum p_T^2$ :** Pile-up vertex sum  $p_T$  squared.
- $\mu$ : Number interaction per crossing.
- $\Delta\phi(\gamma\gamma, jet_1)$ : Angular distance between diphoton system and leading jet ( $jet_1$ ).
- $\Delta\phi(jet, E_T^{miss})$ : Angular distance between jets system and missing transverse energy.
- $\sum p_T^2 (\sum p_{T(php)}^2 - \sum p_{T(PU)}^2)$ : difference between photon pointing and pile-up vertex  $p_T$ .
- **Corrected jet vertex tagger ( $JVT_{corr}$ ):** Corrected jet vertex fraction is the ratio of track  $p_T$  and  $p_T$  of jets in the calorimeter.
- $R_{p_T}^{jet}$ : Scalar sum of jet primary vertex track and jet track  $p_T$  associated with the diphoton system.
- $\Delta\phi(\gamma\gamma, E_T^{miss})$ : Angular difference between diphoton system and missing transverse energy.
- $\Delta\phi(\gamma\gamma 1, \gamma 2)$ : Angular difference between leading and sub-leading photon.
- $\Delta\phi(softjet, E_T^{miss})$ : Angular difference between soft jets (jets with  $p_T$  less than 30 GeV) and missing energy.
- $\Delta\phi(Fjet, E_T^{miss})$ : Angular difference between forward jets (jets outside the central region of the detector,  $\eta \leq 2.4$ )
- **Difference of ref jets and  $p_T^{jj}$ :** Scalar difference of ref jets and the  $p_T$  of two leading jets.

### 3.3. Results

The performance of the BDT classifiers is measured in terms of background rejection efficiency against the signal acceptance efficiency, in the Receiver Operating Curve (ROC). By definition, the top right-most point on the ROC distribution gives the best trade-off point between signal acceptance and background rejection. Figure 3a, 3b and 3c show the performance of the classifiers, an additional benchmark distribution of no improvement spectrum is added. The no

improvement spectrum represents the minimum signal to background significance. The region below the distribution is the no improvement region and the region above the distribution is the region with maximum improvement.

The ROC curve of the intermediate  $S_{E_T^{miss}}$  classifiers shows that we can attain up to 64%, 80%, and 33% improvement insignificance if we cut on 0.9 for A400Z, ggZH125 and 275mx60 classifiers respectively. Another way of checking the performance of the BDT classifiers is to evaluate the trained information on a new dataset. Figures 4a, 4b and 4c show the evaluation performance on data side-band, mixed background, pseudoscalar signal, heavy scalar signal and standard model Higgs production samples. The performance of all the classifiers give a background like a shape on the data side-band and the shape of the data side-band distribution is similar to background shape.

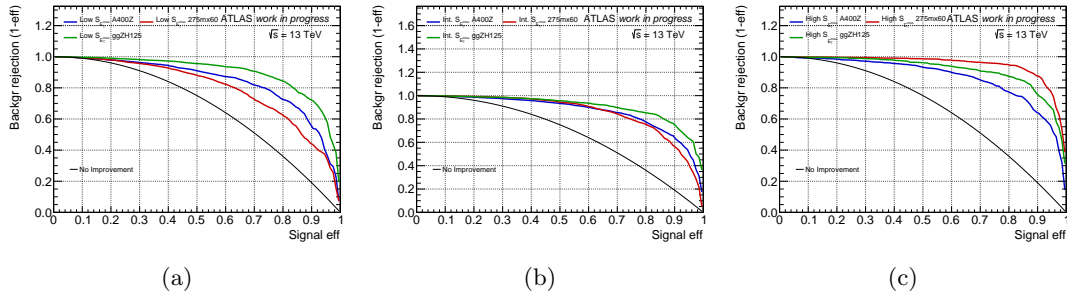


Figure 3: ROC for (a) Low  $S_{E_T^{miss}}$  region (b) intermediate  $S_{E_T^{miss}}$  region (c) High  $S_{E_T^{miss}}$  region

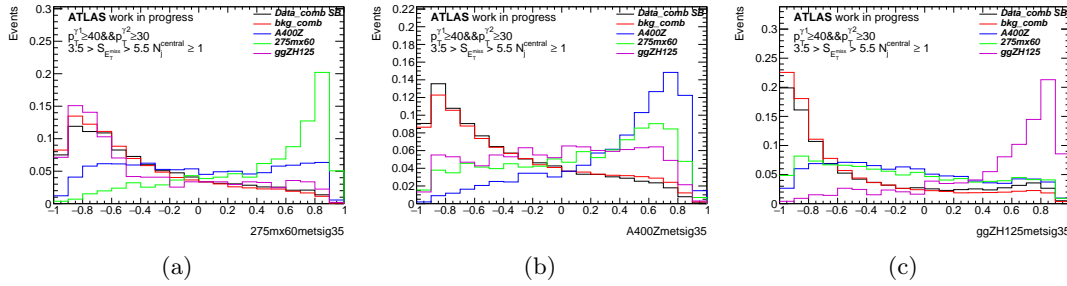


Figure 4: Evaluation of classifier in the intermediate  $S_{E_T^{miss}}$  category for (a) 275mx60 (b) A400Z (c) ggZH125 samples

The performance results suggest we can define a new set of BDT categories, by choosing BDT 0.4 as a working point in the intermediate category. Table 3 gives a summary of the new categories. The lower region represents the background region and the high region represent the BDT signal region.

#### 4. Conclusion

The search for dark matter particle with missing transfer energy signature is very important to most beyond the standard model and in particular the heavy scalar model. The clean signatures of missing transverse energy are mostly contaminated by fake missing energy arising from pileup interactions. The purity of this search region depends heavily on the correct reconstruction primary vertex of interaction. The photon pointing method is developed for correctly identifying

Table 3: Table showing BDT categories.

BDT classifier	BDT High	BDT Low
A400Met35	A400BDT > 0.4	A400BDT < 0.4
ggZh125Met35	ggZh125BDT > 0.4	ggZh125BDT < 0.4
275mx60Met35	275mx60BDT > 0.4	275mx60BDT < 0.4

the vertex of interaction. The diphoton analysis has developed a number of techniques to reject pileup objects. The jet vertex tagger and the forward jet tagger to suppress pile up jets in the forward region and outside the forward region of the detector respectively.

It was observed that the JVT suppression method sometimes causes fake missing energy by removing non-identified physics objects. The same vertex method was developed to reduce fake missing transverse energy caused by JVT. The same vertex variable ensures that the diphoton system is from the actual vertex of interaction and not the pileup vertex. The same vertex method attends over 50% performance of suppressing fake missing transverse energy. The ATLAS detector has seen an increase in average interaction per bunch and increase in pileup interaction. A multivariate analysis technique is developed to further suppress the fake missing energy due to an increase in pileup interaction. The MVA method uses photon and jet kinematics and  $\langle \mu \rangle$  to develop classifiers. The classifiers have about 80% performance for fake missing transverse energy rejection.

## References

- [1] von Buddenbrock S, Chakrabarty N, Cornell A S, Kar D, Kumar M, Mandal T, Mellado B, Mukhopadhyaya B, Reed R G and Ruan X 2016 *Eur. Phys. J.* **C76** 580 (*Preprint 1606.01674*)
- [2] von Buddenbrock S, Chakrabarty N, Cornell A S, Kar D, Kumar M, Mandal T, Mellado B, Mukhopadhyaya B and Reed R G 2015 (*Preprint 1506.00612*)
- [3] Aad G, Abbott B and Abdallah J (ATLAS Collaboration) 2015 *Phys. Rev. Lett.* **115** 131801. 19 p URL <https://cds.cern.ch/record/2021099>
- [4] Aaboud M *et al.* (ATLAS) 2017 *Phys. Rev.* **D96** 112004 (*Preprint 1706.03948*)