Probing the Supersymmetric Standard Model at the LHC through Vector Boson Fusion and Machine Learning

Umar Sohail Qureshi[1,†], Alfredo Gurrola[1]

[1]Department of Physics and Astronomy, Vanderbilt University, Nashville, TN, USA

Background and Motivation

- Supersymmetry (SUSY) restores symmetry between fermions and bosons.
- SUSY introduces new *superpartners* of SM particles.
- Spins of partner SUSY and SM particles differ by 1/2.

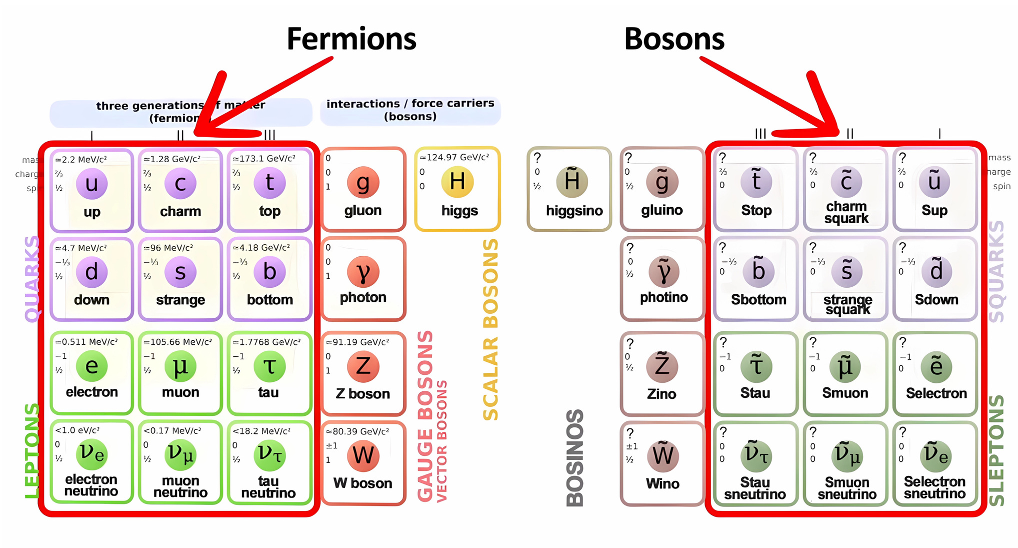

Figure 1. A Supersymmetric Standard Model (SSM).

- SUSY models can simultaneously address:
  - Dark matter (DM) relic density.
  - The gauge hierarchy problem.
  - Higgs boson mass.

Experimental Constraints

- ATLAS and CMS have conducted various SUSY searches.
- Bounds on the colored SUSY sector exclude at 95% CL:
  - Gluinos $\tilde{g}$ up to 2.31 TeV.
  - Stops $\tilde{t}$ up to 1.25 TeV
  - Sbottoms $\tilde{b}$ up to 1.24 TeV.
- Bounds on the electroweak sector are more relaxed:
  - Charginos $\tilde{\chi}_1^\pm$ and neutralinos $\tilde{\chi}_{1,2}^0$ up to 950 GeV.
- Since these bounds are model-dependent, we consider charginos and neutralinos down to 200 GeV.

Vector Boson Fusion (VBF)

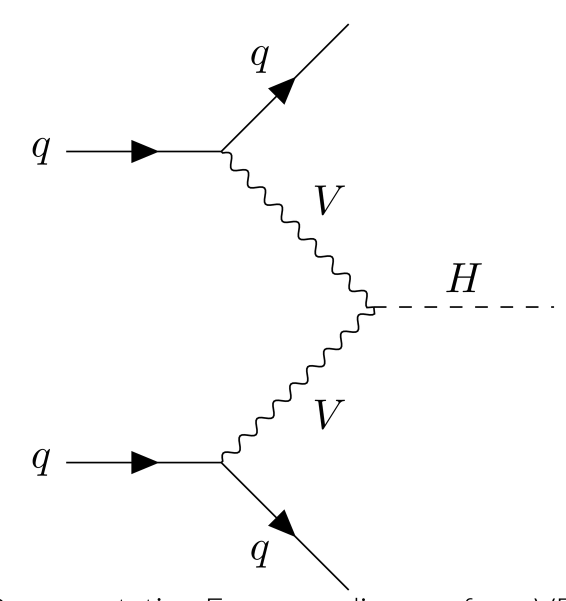

Figure 2. Representative Feynman diagram for a VBF process.

- A quark from each LHC proton radiates a vector boson.
- These bosons fuse to produce a particle e.g. a Higgs.
- The quarks are minimally deflected from their initial directions.
- Leading to energetic jets in the forward direction.
- These jets have large mass and are in opposite hemispheres.

Proposed Analysis Strategy

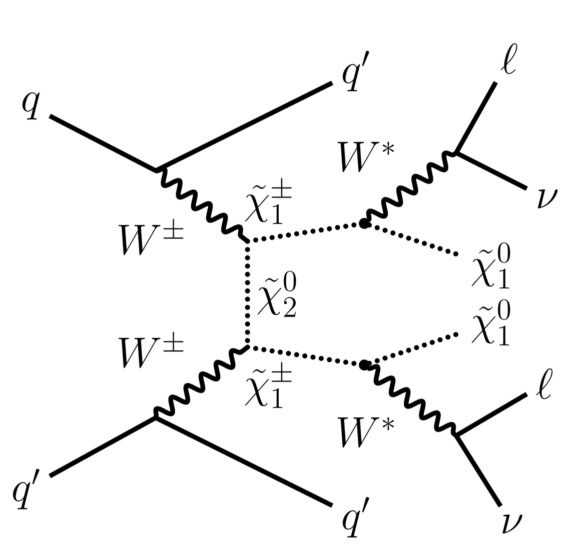

Figure 3. Feymnann diagram for our signal. All chargino and neutralino combinations are considered, thus all lepton multiplicities.

- The R-parity conserving MSSM is probed for this study.
- We define ewkino $= \{\tilde{\chi}_1^+, \tilde{\chi}_1^-, \tilde{\chi}_1^0, \tilde{\chi}_2^0\}$.
- The following α_S exclusive process is considered as signal:

$$pp \to \text{ewkino ewkino } jj \quad (1)$$

- The LSP $\tilde{\chi}_1^0$ is purely bino; the NLSPs $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ are purely wino and mass-degenerate.
- The sleptons $(\tilde{e}, \tilde{\mu}, \tilde{\tau})$ are left-handed, mass-degenerate, and heavier than the $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_{1,2}^0$.
- As such, the branching fraction ratios are:

$$\mathcal{B}\left(\tilde{\chi}_1^\pm \to \tilde{\chi}_1^0 W^{\pm,*}\right) = 1 \quad \text{and} \quad \mathcal{B}\left(\tilde{\chi}_2^0 \to \tilde{\chi}_1^0 Z^*\right) = 1 \quad (2)$$

- The only relevant SM backgrounds are $pp \to V$ + jets, $pp \to VVjj$, and $pp \to t\bar{t}$ with semi-leptonic decay.
- The VBF jet topology massively mitigates SM backgrounds.

Samples and Simulation

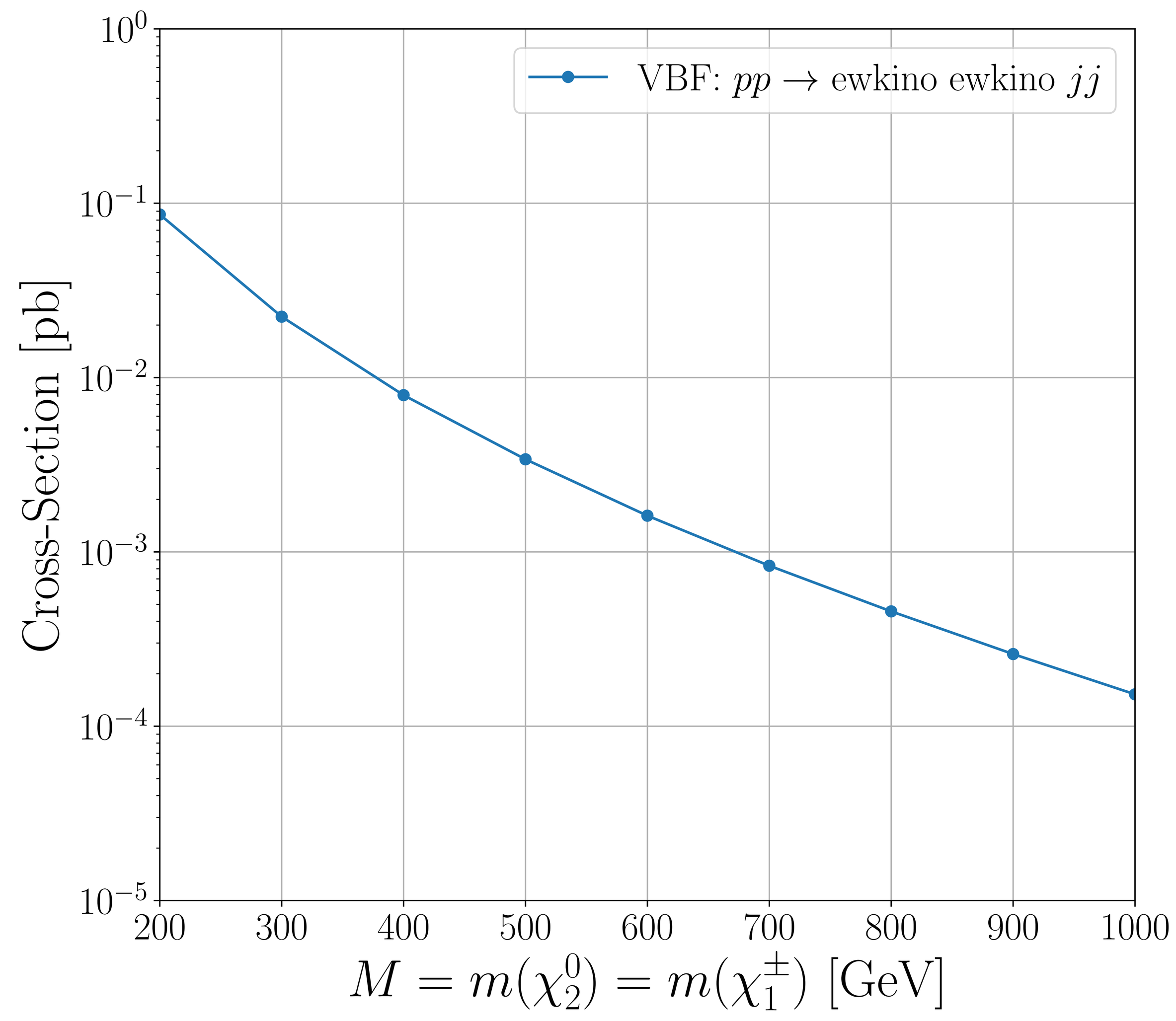

Figure 4. Signal production cross-section as a function of the $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^\pm$ masses.

- Samples are produced with the event generator `MadGraph5`.
- `Pythia8` is used for parton showers and hadronization.
- To ensure VBF production, the SUSY colored sector is decoupled.
- Further, generator-level cuts of $|\Delta\eta(jj)| > 3.5$ and $m(jj) > 200$ GeV are applied to suppress non-VBF contributions.
- We define $\Delta m = m(\tilde{\chi}_2^0) - m(\tilde{\chi}_1^0)$ and fix $\Delta m = 50$.
- With $M = m(\tilde{\chi}_2^0) = m(\tilde{\chi}_1^\pm)$, we let $M \in \{200, 300, \ldots, 1000\}$ GeV.
- To ensure sufficient statistics, 10M events are simulated.
- The event selection criteria is: $\geq 1\ell$ with $p_T \geq 5$ GeV and $|\eta| < 2.5$.
- All leptons must pass ID and isolation criteria.

Analysis using Machine Learning

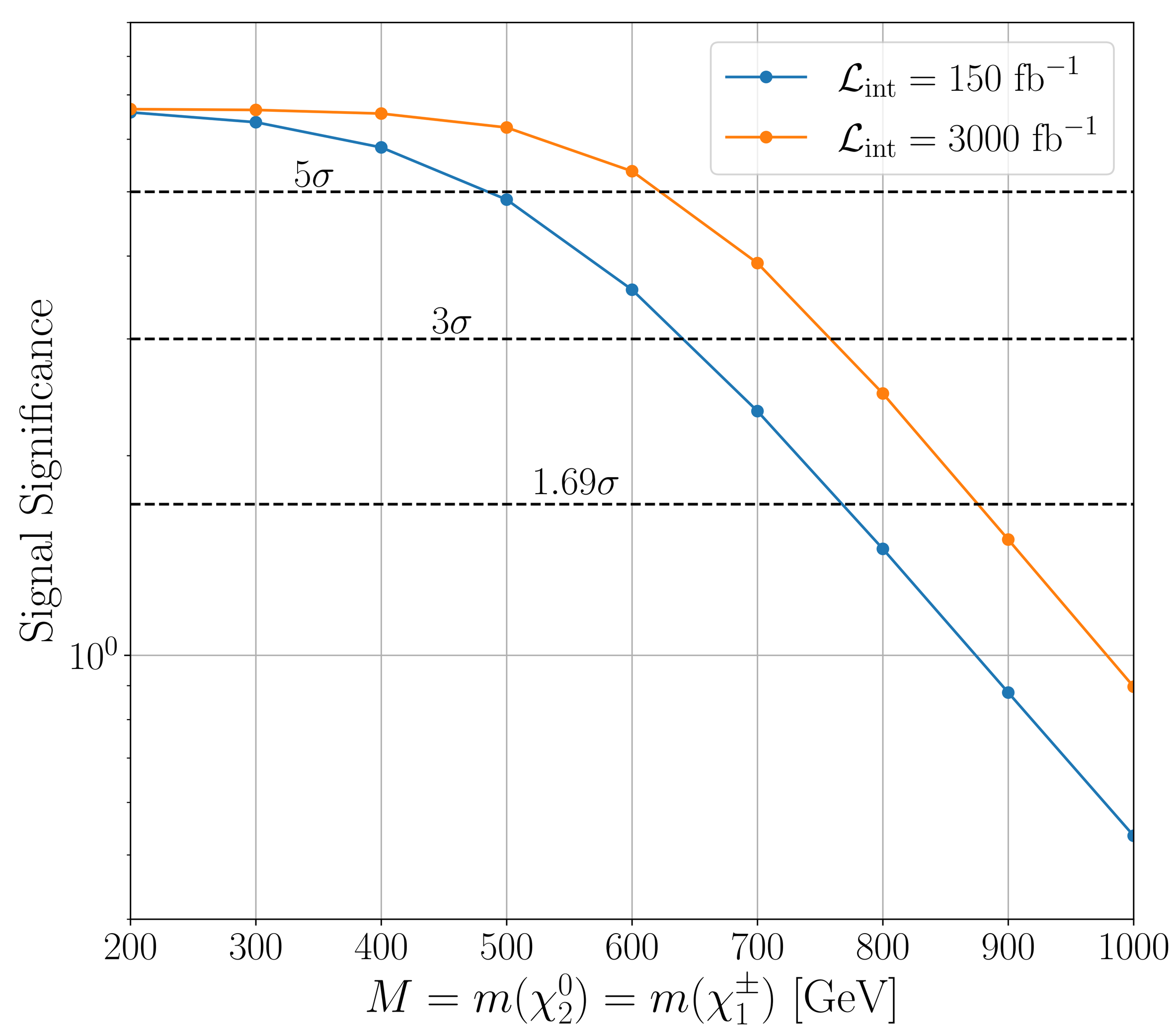

Figure 5. Signal significance as a function of the $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^\pm$ masses.

- Signal and background events are combined for training.
- A boosted decision tree (BDT) is trained for signal-background discrimination (binary classification).
- We proceed with the BDT's predictions for yet-unseen data.
- The BDT output distributions are normalized to:

$$N = \mathcal{L}_{\text{int}} \cdot \sigma \cdot \varepsilon \quad (3)$$

Where $\mathcal{L}_{\text{int}}$ is the integrated luminosity, σ is the cross-section, and ε is the efficiency factor.

- A bin-by-bin calculation is used to compute signal significance:

$$\mathcal{S} = \frac{\sum s_i w_i}{\sqrt{\sum (s_i + b_i)\, w_i^2 + \beta^2 \sum w_i^2 \left(s_i^2 + b_i^2\right)}} \quad (4)$$

Where s_i and b_i are the number of signal and background events in the i^{th} bin, β is the systematic uncertainty, and w_i is the weight of the i^{th} bin defined as:

$$w_i = \ln\left(1 + \frac{s_i}{b_i}\right) \quad (5)$$

Compressed Mass Spectrum Scenario

- The mass gap between the LSP and the NSLPs is small.
- This is the so-called compressed spectrum scenario.
- These scenarios have been challenging experimentally.
- Soft decay products that are challenging to detect.
- VBF and machine learning significantly alleviate this problem.
- The choice of $\Delta M = 50$ is motivated by the fact that it is:
  - Large enough such that $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^\pm$ are not long-lived.
  - Small enough to obtain on-mass shell W^*/Z^* decay.

Conclusion and Future Work

Our analysis strategy can extend LHC constraints to $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^\pm$ masses at a:

- $\geq 5\sigma$ **signal significance** for masses up to **660 (520)** GeV.
- $\geq 3\sigma$ **signal significance** for masses up to **770 (620)** GeV.
- $\geq 95\%$ **confidence level** for masses up to **880 (750)** GeV.

With an integrated luminosity of 3000 (150) fb^{-1}.

- As such, we advocate for an experimental search using our methodology at ATLAS and CMS.

Appendix: Efficiencies and Uncertainties

- MadGraph and Pythia do not consider detector response.
- For example, light-jet identification efficiency is around 80%.
- We consider these effects as an ε factor.
- So, for a final state with dijets and 1 muon, we have $\varepsilon \approx 0.61$.

Systematic uncertainties are taken into account as follows:

- 3% on the CMS measurement of $\mathcal{L}_{\text{int}}$.
- 2-5% in jet energy scale uncertainties.
- 1-2% shape-related in the BDT distribution.
- 5-10% in the signal and background predictions.

Appendix: Kinematic Distributions

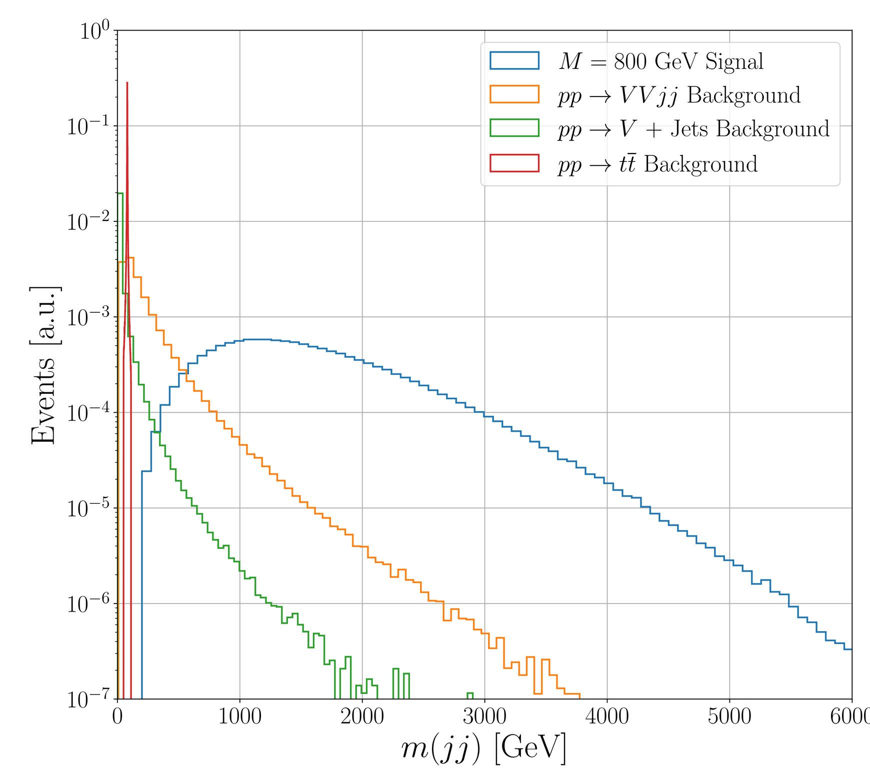

Figure 6. Dijet invariant mass distributions for a $M = 800$ GeV signal event and relevant SM backgrounds.

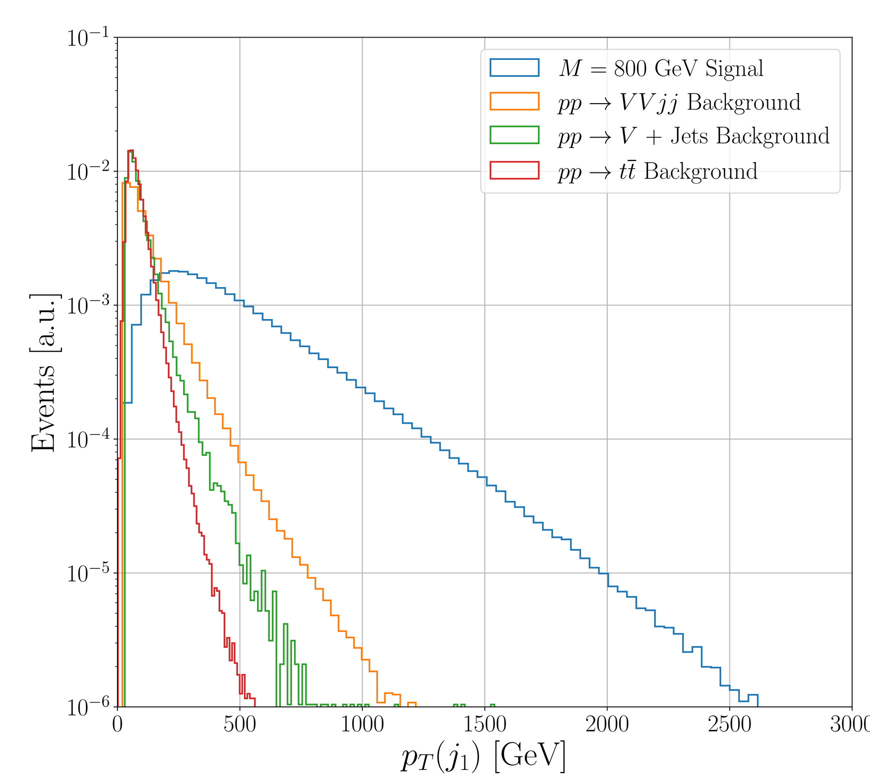

Figure 7. Leading jet transverse momentum distributions for a $M = 800$ GeV signal event and relevant SM backgrounds.

Appendix: ML Model Performance

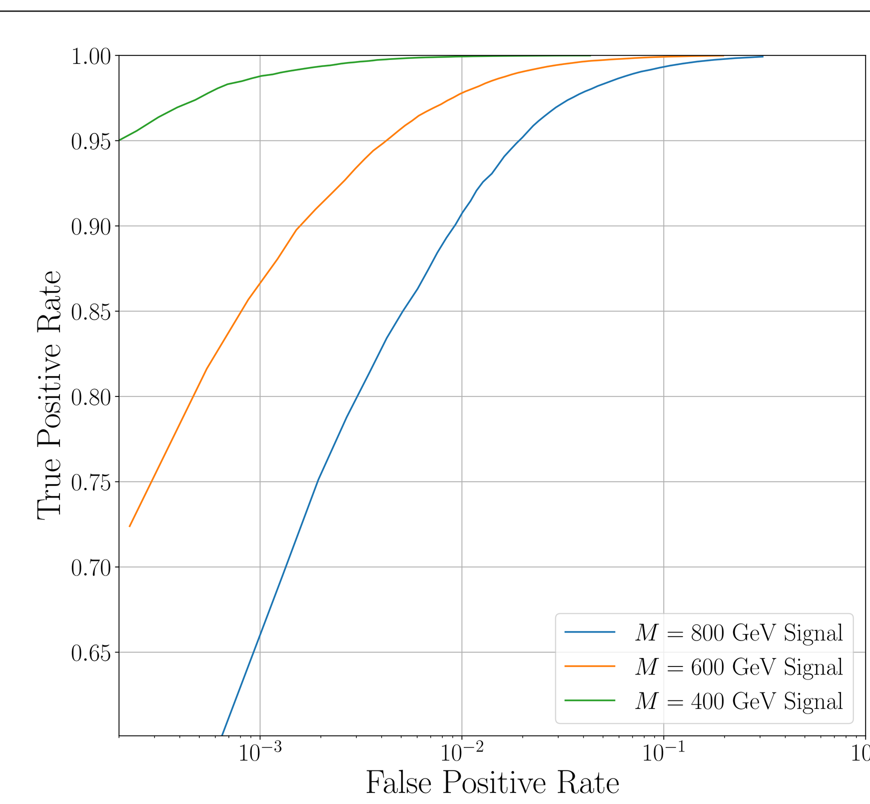

Figure 8. Receiver operating characteristic curve of the BDT for three signals.

Acknowledgements

- I thank Prof. Alfredo Gurrola for his invaluable support and guidance.
- This work is supported in part by NSF Award PHY-1945366 and a Vanderbilt Seeding Success Grant.

†uqureshi@cern.ch