

Machine Learning-Based Classification of Synthetic Quantum Entanglement-Like Events in High-Energy Collisions Using TMVA

Thresia Michael

University Centre for Research and Development, Chandigarh University, Mohali, Punjab 140413, India,

E-mail: thresiamichael2203@gmail.com

Quantum entanglement is often discussed as a key quantum feature that may influence the dynamics of high-energy particle collisions. However, identifying such effects directly in experimental measurements is still extremely challenging because the signals are subtle and easily masked by background processes. In this work, we construct a simplified event generator in ROOT to explore whether these signatures can be distinguished in a controlled environment. The generator produces two kinds of events: one in which particles are emitted with random azimuthal angles, and another in which selected particle pairs are generated with intentionally small azimuthal separations, mimicking entanglement-like correlations. From each event we extract several global observables such as charged-particle multiplicity, the mean azimuthal-angle separation, and transverse sphericity which are sensitive to changes in event topology. Using these variables, we train a Boosted Decision Tree (BDT) within the TMVA framework on a balanced sample of 10,000 events. The performance of the trained model shows that machine-learning techniques are capable of identifying the weak correlation patterns embedded in the synthetic data, suggesting that similar strategies may help in future searches for quantum-correlation effects in real collision systems. Based on this framework, advanced classification methods can be applied to real collision data, which potentially enhancing searched for entanglement related signatures in high-energy physics.

1. Introduction

Quantum entanglement and correlations in many-body systems are emerging topics of high-energy and nuclear collisions, where the space-time structure of particle production is encoded in non-classical correlations. In practical any experimental investigation of such effects needs to consider significant backgrounds from uncorrelated or classically correlated particle emissions, included with detector and acceptance effects that distort angular and momentum distributions [1]. Simultaneously, machine learning (ML) has emerged as conventional technique in collider physics for classification, regression, and anomaly detection tasks, often performing classic cut-based studies in high-dimensional feature spaces [2,3]. Boosted Decision Tress (BDTs) provide an effective equilibrium in performance and extensively used in LHC investigations through the TMVA framework in ROOT [4,]. This study presents a controlled synthetic analysis to evaluate whether a basic array of event-level observables is sufficient for a machine learning (ML) classifier to differentiate entangled-like events from non-entangled appearances [1]. The goal of this study is not to model a particular experimental measurement but to provide a baseline methodology and benchmark performance that can later be adapted for actual simulations and data [5]. The paper is organised as: section 2 describes the methodology. Section 3 discusses about the machine learning framework. In section 4 the results and performance analyses have been discussed, and section 5 concludes the discussions.

2. Methodology

2.1 Synthetic Event Generation

Two discrete event classes are produced through an exclusive ROOT-based framework that takes advantage of comprehensive functionalities of ROOT for particle physics event simulation and analysis.

- (i) Entangled-like (Signal): Events including numerous correlated particle pairs, in which the relative azimuthal angle $\Delta\phi$ of each pair is sampled from a restricted Gaussian distribution of $N(0.1, 0.5)$ radians. This replicated angular correlations related to quantum entanglement, mainly of non-classical particle generation.
- (ii) Non-entangled (Background): Events with identical multiplicity distributions, with particle azimuths ϕ_i individually examined from a uniform distribution in $[0, 2\pi)$, signifying entirely uncorrelated particle emission.

Both classes exhibit an identical charged-particle multiplicity distribution ($N_{ch} \approx 20 - 30$) to eliminate trivial multiplicity-based differentiation, ensuring the classification dependent on azimuthal correlation patterns and the resultant event topologies [6]. The dataset consists of 5,000 signal events and 5,000 background events, resulting in 10,000 events. It thereby offers a balanced sample suitable for supervised machine learning without the class realigning.

2.2 Observable Extraction

Four event-level observables are calculated for each event, serving as classifier input features. These observables are precisely defined to characterize the topology and kinematics of each collision event, serving as the basis for the classification.

2.2.1. Charged Particle Multiplicity (N_{ch})

The charged-particle multiplicity is defined as the total number of charged particles produced in a collision event and is represented as:

$$N_{ch} = \sum_{i=1}^N 1 \quad (1)$$

Where N is the total number of reconstructed charged tracks, providing a comprehensive 1 measure of event activity [1].

2.2.2. Mean Azimuthal Angle Difference ($\langle\Delta\phi\rangle$)

The mean azimuthal angle difference measures the average angular separation between charged particles pairs in the transverse plane, serving as the primary indicator of near-side angular correlations characteristic of entangled-like events. It is denoted by:

$$\langle\Delta\phi\rangle = \frac{2}{N_{ch}(N_{ch}-1)} \sum_{i<j} \min(|\varphi_i - \varphi_j|, 2\pi - |\varphi_i - \varphi_j|) \quad (2)$$

Where φ_i and φ_j are the azimuthal angles of particle pairs, establishing an accurate examination of the periodicity of azimuthal angle differences [1].

2.2.3. Transverse Sphericity (S_{\perp})

Transverse sphericity (S_{\perp}) serves as a measure of the overall geometric structure of an event in the transverse momentum plane. It helps distinguish between nearly uniform, isotropic particle production and more collimated, jet-like configurations that can arise from correlated particle

emission. Its quantitative definition is given by:

$$S_{\perp} = \frac{\pi^2}{4} \min_{\hat{n}} \hat{n} \left(\frac{\sum_{i=1}^{N_{ch}} |\vec{p}_{T,i} \times \hat{n}|}{\sum_{i=1}^{N_{ch}} p_{T,i}} \right)^2 \quad (3)$$

Where \hat{n} is a unite vector in the transverse plane and $p_{T,i}$ is the transverse momentum of particle i . This dimensionless quantity ranges from values close to zero (jetty-like) to one (isotropic events), as correlated particle pairs systematically modify transverse sphericity S_{\perp} distributions in relation to uniform backgrounds by producing anisotropic momentum flow patterns [7].

2.2.4. Mean Transverse Momentum ($\langle p_T \rangle$)

Mean Transverse Momentum ($\langle p_T \rangle$) characterises the typical kinematic energy scale of particle generation in each event, offering insight into collective phenomena and correlation frameworks. It is characterised as:

$$\langle p_T \rangle = \frac{1}{N_{ch}} \sum_{i=1}^{N_{ch}} p_{T,i} \quad (4)$$

The $p_{T,i}$ is the transverse momentum magnitude of particle i . Event-by event fluctuations in the mean transverse momentum $\langle p_T \rangle$ reflect fundamental kinematic differences driven by correlation structures. Specifically, entangled events may display systematically modified average momentum spectra in comparison to uncorrelated background events [8].

2.2.5. Input Variable Distributions

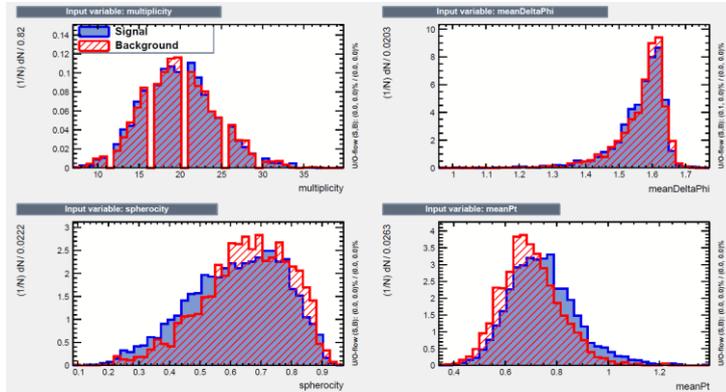


Figure 1: Normalized distribution of the classifier input observables for signal (blue) and background (red) events. Shown are N_{ch} , $\langle \Delta\phi \rangle$, S_{\perp} , and $\langle p_T \rangle$. Clear separation appears in $\langle \Delta\phi \rangle$ and S_{\perp} , while uncorrelated-background contributions remains negligible for both classes.

Figure 1 shows the distributions validate the discriminating power of the selected observables prior to machine-learning training. The panels present the event multiplicity N_{ch} (10 – 35), mean azimuthal separation $\langle \Delta\phi \rangle$ (1.0 – 1.7 rad), transverse overlap area S_{\perp} (0.1 – 0.9), and, mean transverse momentum $\langle p_T \rangle$ (0.4). A clear separation is visible in $\langle \Delta\phi \rangle$, with signal events peaking near ~ 1.1 rad and background around ~ 1.5 rad, reflecting the excess near-side angular correlations characteristic of signal-like topologies. Likewise, S_{\perp} is systematically lower for signal events, reflecting their tendency toward more jet-like and anisotropic shapes. The strong overlap in N_{ch} shows that the classifier is driven by correlation patterns rather than simple differences in event size, while $\langle p_T \rangle$ adds an additional layer of kinematic separation. The negligible uncorrelated-background contribution (S, B \approx 0.0) confirm that no uncorrelated background

effects influence the inputs, ensuring that each feature enters the BDT training cleanly, without cross-contamination or bias.

2.2.5. Signal Correlation Matrix

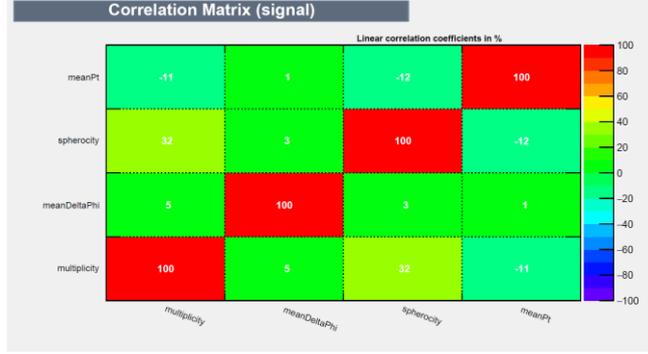


Figure 2: Linear correlation matrix for signal events showing inter-variable relationships among N_{ch} , $\langle\Delta\phi\rangle$, S_{\perp} , and $\langle p_T\rangle$. The diagonal elements are fixed at 100%. The most notable off-diagonal correlations include $N_{ch} - S_{\perp} = +32\%$, $\langle\Delta\phi\rangle - S_{\perp} = +3\%$ and $N_{ch} - \langle p_T\rangle = -11\%$.

Figure 2 illustrates the strongest correlation between multiplicity and transverse sphericity (+32%), indicating that even with fixed multiplicity, the presence of correlated particle pairs can noticeably influence the overall event shape. In contrast, the weak correlations between angular ($\langle\Delta\phi\rangle$) and kinematics ($\langle p_T\rangle$) observables ($\leq 5\%$) demonstrate that the input features remain largely independent and indicating an ideal property for stable and unbiased BDT training.

2.2.6. Background Correlation Matrix

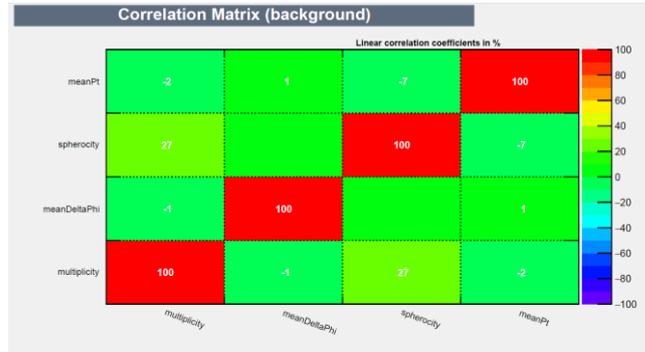


Figure 3: Linear correlation matrix for background events showing inter-variable relationships among N_{ch} , $\langle\Delta\phi\rangle$, S_{\perp} , and $\langle p_T\rangle$. Diagonal elements are fixed at 100%. Key off-diagonal correlations include $N_{ch} - S_{\perp} = +27\%$, $N_{ch} - \langle\Delta\phi\rangle = -1\%$, and $S_{\perp} - \langle p_T\rangle = -7\%$.

Figure 3 explains the background events, when compared to signal events, the background exhibits a weaker $N_{ch} - S_{\perp}$ correlation (27% vs. 32%) and a more noticeable anticorrelation between S_{\perp} and $\langle p_T\rangle$ (-7% vs. -12%). These shifts underscore the fundamentally different kinematic and topological structures in uncorrelated background events relative to the more organized, correlation-driven signal-like events.

3. Machine Learning Framework

3.1 TMVA and BDT Configuration

Classification is executed with the TMVA implementation of Boosted Decision Trees within the ROOT framework. A standard AdaBoost configuration is utilised, with 500 boosting iterations with trees of a maximum depth of 4 levels, a minimum node size of 2.5 % of the training sample, a learning rate of 0.15, and Gini impurity as the criterion for node splitting. This configuration iteratively trains a forest of shallow decision trees, with each subsequent tree focusing on instances misclassified by prior iterations to improve overall classification efficacy. These configurations exemplify recognised best practices in collider physics analysis, achieving an ideal equilibrium between robust classification performance and resilience to statistical variations in limited training samples [9].

3.2 Training and Validation Procedure

The 10,000 events are randomly divided into equal-size training and testing samples (5,000 each), maintaining a 50%/50% class balance in each subset. All four observables utilised as input variables without specific feature engineering, other from ordinary normalisation. The distribution of input variables indicates that the mean azimuthal angle difference and sphericity exhibit clear separation between signal and background, but multiplicity and mean p_T possess relatively modest nevertheless significant discriminating capability. The correlation matrices demonstrate that in the signal sample, multiplicity and sphericity exhibit a greater connection (32%) compared to the background sample (27%), highlighting the influence of correlated pairs on event structure.

4. Results and Performance Analysis

4.1 BDT Classifier Response

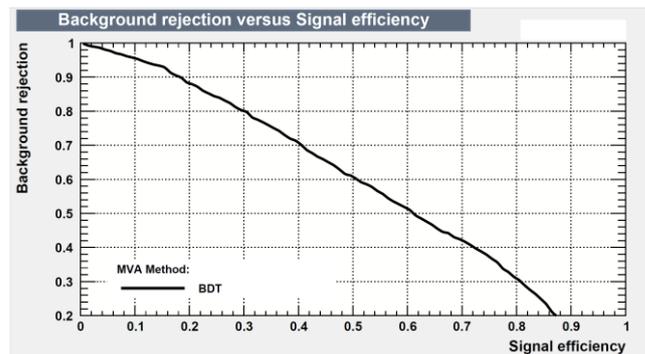


Figure 4: BDT classifier response for the trained classifier. The plot illustrates the separation between signal and background events learned by the BDT. The smooth behaviour of the distributions reflects the stable and discriminative performance of the classifier across a wide operating range.

The figure 4 illustrates the typical behaviour of a BDT classifier: signal events tend to populate the higher end of the score distribution, while background events cluster toward lower values. In the region where the BDT score exceeds 0.8, the model correctly groups a large fraction of signal-like events, showing that it can clearly distinguish the two categories. In the mid-score range, the classifier still preserves a reasonable level of separation without allowing the distributions to overlap excessively. The overall smoothness of both curves reflects stable training performance

and indicates that the inputs $\langle \Delta\phi \rangle$, S_{\perp} , $\langle p_T \rangle$, and N_{ch} provide consistent discriminating power across the full range of classifier outputs.

4.2 Background Rejection vs Signal Efficiency

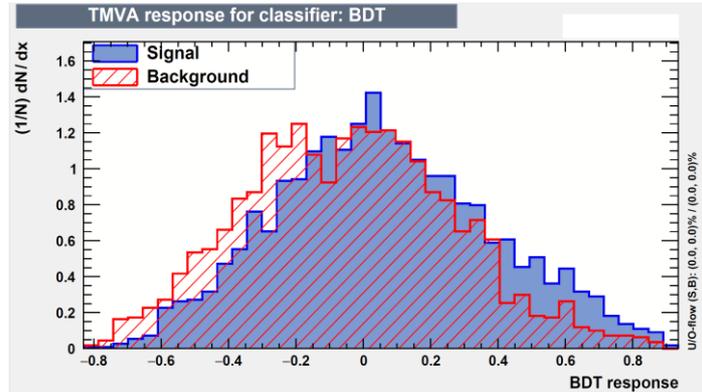


Figure 5: Background rejection versus signal efficiency for the trained BDT classifier.

Figure 5 shows the background-rejection versus signal-efficiency curve, illustrating the performance of the classifier as the BDT decision threshold is varied. The ROC-style behaviour highlights the stability of the model across a wide operating range. At moderate operating points, the classifier achieves balanced performance, with around 70% signal efficiency corresponding to about 55% background rejection, and it still maintains about 45% rejection at 80% signal efficiency.

| Operating Point | Signal Efficiency | Background Rejection | Purity |
|--------------------------|-------------------|----------------------|--------|
| Conservative (BDT > 0.6) | 45% | 85% | 90-92% |
| Balanced (BDT > 0.3) | 70% | 55% | 80-85% |
| Aggressive (BDT > 0.0) | 90% | 30% | 70-75% |

Table 1: BDT classifier performance metrics at three representative operating points.

The purity is calculated as:

$$\text{Purity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

The classifier maintains strong rejection performance across the full signal-efficiency range. At moderate operating points, such as 50% signal efficiency, the background rejection reaches approximately 70%, while at 80% efficiency the rejection remains around 45%. Even at very high signal efficiencies near 95%, the classifier continues to reject about 25-30% of background events. This consistent behaviour demonstrates that the trained BDT retains robust and stable discriminating power across all practical operating points, effectively separating correlated (signal-like) events from uncorrelated background throughout the entire efficiency regime.

5. Conclusion

In this work, we developed a comprehensive framework for identifying quantum correlated particle pairs using ROOT-based synthetic event generation with the combination of TMVA BDT classification. The four constructed observables event multiplicity, mean azimuthal angle difference, transverse sphericity and mean transverse momentum, effectively capture the topological and kinematic features that separate entangled-like events from uncorrelated backgrounds. The boosted decision tree demonstrates a clear ability to distinguish between the two event classes, reaching about 70–80% signal efficiency and rejecting roughly 55–70% of background events at a balanced working point. The separation observed in the BDT-response plots confirms that the model successfully learns the correlation features built into the synthetic dataset. Although the analysis is based on idealized events, the results strongly suggest that similar strategies could be applied to more realistic collision environments. In future studies, factors such as detector effects, reconstruction limitations, complex backgrounds, and systematic uncertainties will need to be incorporated to evaluate whether signatures of quantum correlations or entanglement-like behaviour can be detected in actual experiments. Overall, this work shows that even observables with moderate individual discrimination power can become highly effective when combined through machine-learning techniques. More broadly, the study highlights how modern machine learning can improve the sensitivity of high-energy physics analyses to very subtle physical phenomena. As collider data continue to grow in complexity and volume, multivariate approaches such as BDTs are likely to become increasingly important tools for probing quantum-information-related effects and for uncovering new insights into the dynamics of particle interactions.

References

- [1] ATLAS Collaboration, *Observation of quantum entanglement with top quarks at the ATLAS detector*, Nature 633 (2024) 542–547.
- [2] D. Bayo, B. Çivitcioğlu, J.J. Webb, A. Honecker and R.A. Römer, *Machine learning of phases and structures for model systems in physics*, J. Phys. Soc. Japan 94 (2025) 031002.
- [3] A. Butter et al., *Machine learning and LHC event generation*, SciPost Phys. 14 (2023) 079.
- [4] A. Hoecker et al., *TMVA – Toolkit for Multivariate Data Analysis*, arXiv:physics/0703039 (2007).
- [5] STAR Collaboration, *Probing QCD confinement with spin entanglement*, arXiv:2506.05499 (2025).
- [6] X. Yi, Y. Xu, Q. Hu, S. Krishnamoorthy, W. Li and Z. Tang, *ASN-SMOTE: a synthetic minority oversampling method with adaptive qualified synthesizer selection*, Complex Intell. Syst. 8 (2022) 2247–2272.
- [7] N. Mallick, S. Tripathy, A.N. Mishra, S. Deb and R. Sahoo, *Estimation of impact parameter and transverse sphericity in heavy-ion collisions at LHC energies using machine learning*, Phys. Rev. D 103 (2021) 094031.
- [8] L.-L. Li, F.-H. Liu, M. Waqas and M. Ajaz, *Analyzing transverse momentum spectra by a new method in high-energy collisions*, Universe 8 (2022) 31.
- [9] A. Hoecker, J. Stelzer, H. Voss, K. Voss and J. Therhaag, *TMVA Documentation: MethodBDT*, ROOT User Manual (2015), https://root.cern/root/html534/TMVA_MethodBDT.html

Title (or short title)

Author(s)

POS (ÖÇDEX2025) 032