// ACADEMIC / 2026 / NOV

Physically constrained tabular variational autoencoders for small sample blasting fragmentation data

Paulo Lopes
https://linktr.ee/pftlopes

Abstract

Small blasting datasets challenge generative models because physical bounds, heterogeneous fragmentation regimes and algebraically related variables must be represented simultaneously. This study examines an existing 62-record fragmentation benchmark in which separate tabular variational autoencoders were fitted within fragmentation-index groups, with a Gaussian copula fallback for subsets smaller than 20 records. Eight engineering variables were generated, and fragmentation index was reconstructed by adding an empirical residual to the logarithm of the in-situ-block-size to fragment-size ratio. The reported workflow attained an overall SDV fidelity score of 0.8568, compared with 0.8261 for a pure TVAE and 0.8023 for Gaussian simulation. Train-on-synthetic-test-on-real regression produced $R^2 = 0.7030$ and mean absolute error = 0.0364 m, while the reverse transfer yielded $R^2 = 0.1302$. Reported support precision and recall were 0.9247 and 0.9194. These results favour the grouped workflow descriptively, especially for marginal fidelity, but do not isolate the contributions of grouping, reconstruction and rejection constraints. They also do not establish independent predictive validity because split chronology and task-specific predictor lists are incompletely documented. The classification label based on in-situ block size is a threshold proxy, rather than a measured post-blast oversize outcome, and can become circular if its defining variable is retained as an input. The analysis develops the assumptions behind residual reconstruction, interprets transfer asymmetry and specifies the additional validation needed to establish engineering utility without overstating the evidence.

Keywords: rock fragmentation; synthetic data; TVAE; physical constraints; fragmentation index; residual reconstruction; small samples; evaluation

1 Introduction

Blasting fragmentation depends on explosive loading, geometric confinement, rock stiffness and the structure of the unbroken rock mass. Predictive models are useful only if these relationships are represented in the data available for calibration. Existing multivariate and neural-network studies establish a technical basis for modelling fragment size with blast-design and rock descriptors [1,2]. More recent work combines clustering, optimisation and augmentation [3]. Across these approaches, the limited number of independently measured blasts remains a fundamental constraint.

A synthetic-data workflow can encode relationships that an unconstrained statistical simulator would ignore. Positivity, consistent units and feasible combinations of variables help exclude implausible records. However, physical plausibility is a property of the joint sample, not a consequence of a

generator's name. A table can obey simple bounds while violating response relationships, misrepresenting rare conditions or reproducing the training records almost exactly. Reviews of tabular generation therefore emphasise that the generation mechanism and its evaluation must be considered together [4].

Variational autoencoders learn a latent representation through a reconstruction objective regularised toward a prior [5]. The tabular VAE implementation included in the CTGAN benchmark extends this idea to heterogeneous columns [6]. Copulas provide a simpler alternative that separates marginals from dependence [7]. Neither model class automatically resolves an ambiguous definition of fragmentation index or a mismatch between the target and the time at which predictors become available.

The benchmark examined here combines group-wise generation with an explicit reconstruction of FI around the ratio between in-situ block size and the fragment-size field labelled x50. The focus is whether the reported fidelity, utility and support diagnostics establish a useful joint-data representation, and what further evidence would be needed for prospective fragmentation prediction.

2 Data and problem formulation

2.1 Dataset and engineering variables

The underlying study selected the table identified as Hudaverdi_2012, containing 62 records with blast-design ratios, powder factor, rock stiffness, in-situ block size, fragment size and FI [2,8]. Table 1 lists the retained variables. The dataset identifier should be distinguished from a claim of 62 independent, newly measured field trials. Source publications compiled blasts from different mines and rock formations [2]; an operational validation should recover blast identifiers, location and campaign relationships before deciding which rows can be evaluated independently.

Table 1. Variables in the generation and utility workflow. Design ratios are dimensionless, PF is in kg/m³, xB and x50 are in metres, E is in GPa and FI is dimensionless. Descriptions identify their roles in the benchmark rather than asserting how each value was measured in the field.

Variable	Meaning	Type	Role
S/B	Spacing-to-burden ratio	Blast-design ratio	Generated input variable
H/B	Bench-height-to-burden ratio	Blast-design ratio	Generated input variable
B/D	Burden-to-hole-diameter ratio	Blast-design ratio	Generated input variable
T/B	Stemming-to-burden ratio	Blast-design ratio	Generated input variable
PF [kg/m³]	Powder factor	Explosive loading	Generated input variable
xB [m]	In-situ block size	Rock-mass descriptor	Generated rock descriptor and FI reconstruction term
E [GPa]	Modulus of elasticity	Rock-mechanical descriptor	Generated input variable

Variable	Meaning	Type	Role
x50 [m]	Fragment-size field x50	Fragmentation response	Regression target and FI reconstruction term
FI	Fragmentation index	Fragmentation descriptor	Grouping variable and reconstructed response

The label x50 is retained for continuity with the experimental record. It does not resolve whether all source values represent a mass median, an arithmetic mean or another characteristic particle size. The originating literature uses the expression mean fragment size [1,2]. These estimands should not be treated as interchangeable without the corresponding size-distribution information. Likewise, E should retain its source definition: laboratory intact-rock modulus and rock-mass modulus are different properties.

2.2 Fragmentation index and measurement semantics

The benchmark describes a ratio-based interpretation of FI in related source tables:

$$\mathrm{FI} = x_B / x_{50} \qquad (1)$$

For Hudaverdi_2012, the experimental record treats reported FI as varying around that ratio [8]. A discrepancy from Eq. (1) does not, by itself, establish an independent measurement. Rounding, digitisation, a different particle-size convention, unit conversion or a different definition of the block-size descriptor could all produce departures from equality. The residual should therefore be interpreted as an empirical discrepancy until its origin is resolved, rather than assumed to quantify physical stochasticity alone.

This distinction changes the modelling objective. If FI is an exact derived descriptor under a source's definition, it should be calculated after generating its defining variables. If it is a separately estimated index with demonstrable measurement error, a calibrated residual model can represent that uncertainty. Preserving inconsistency without investigating its provenance can make a synthetic table faithful to transcription noise while weakening its engineering meaning.

2.3 Group definition and joint synthesis

The retained FI groups are FI ≤ 4, 4 < FI < 8 and FI ≥ 8. They are data-analysis strata, not established universal fragmentation classes. The boundaries assign exactly 4 to the first group and exactly 8 to the third. These endpoint rules differ from some related benchmarks and are preserved explicitly. A pooled generator represents group-conditioned distributions combined with group weights; changing those weights changes the intended synthetic population even if every within-group generator is unchanged.

Using FI to construct groups is legitimate for retrospective joint-data synthesis when groups and all fitted quantities are derived only from real training records. It does not make FI available before a blast. For predictive evaluation, the generator may use the training target, but the final predictor must use only information available for the stated prediction task. In addition, evaluating fidelity after balancing rare groups requires a declared target mixture or appropriate weighting; it cannot be compared uncritically with the original unbalanced sample.

3 Methodology

3.1 Grouped generators and baseline comparisons

Separate models were fitted within the three FI regimes, and their sampled outputs were recombined. The project reports poor preliminary global-model representation of the high-FI regime, which motivated grouping [8]. These preliminary observations explain the design choice but are not a controlled ablation. A pure TVAE and a Gaussian simulation baseline were retained for the main comparison.

The generation variables were S/B, H/B, B/D, T/B, PF, xB, E and x50. Derived specific-surface fields were excluded to avoid fitting multiple deterministic representations of the same quantity. This is sensible for joint generation, although it does not resolve predictor leakage in a downstream task that includes a derived FI as a feature. The generation schema and each prediction schema must be documented separately.

Within each group, TVAE was trained for 300 epochs with source-derived minimum and maximum enforcement enabled. A Gaussian copula was substituted when a group contained fewer than 20 records [8]. The threshold is an implementation heuristic, not an empirically established point at which a neural model becomes unreliable. The number of real records per group, the model actually used for each group and the synthetic allocation per group are required to reproduce the comparison.

3.2 Residual reconstruction

The log-ratio and residual are calculated as:

$$r = \log(x_B) - \log(x_{50}), \qquad \epsilon = \log(\mathrm{FI}) - r \qquad (2)$$

For generated block and fragment sizes, the synthetic log-ratio is:

$$r_{\mathrm{syn}} = \log\left(x_{B,\mathrm{syn}}\right) - \log\left(x_{50,\mathrm{syn}}\right) \qquad (3)$$

A residual sampled from the corresponding empirical group pool reconstructs FI:

$$\mathrm{FI}_{\mathrm{syn}} = \exp\left(r_{\mathrm{syn}} + \epsilon_{\mathrm{syn}}\right) \qquad (4)$$

Equations (2) and (3) should be evaluated with consistently expressed lengths. Formally, the logarithms can be written for lengths divided by the same reference length, which cancels in the difference; the resulting ratio and FI are dimensionless. Positivity is required before taking logarithms. Equations (1)-(4) retain the mathematical content of the original formulation.

The reconstruction encodes an assumption that the residual distribution within a group can be reused across generated covariate combinations. Sampling residuals independently of those combinations discards any within-group association between the discrepancy and block size, fragment size, stiffness or blast design. Grouping can reduce this heterogeneity but does not eliminate it. Residual-versus-covariate plots and conditional residual summaries are needed to test whether the approximation is acceptable.

The empirical residual pool also has finite support. It cannot supply an unobserved discrepancy simply by being resampled many times. Because the exponential function is nonlinear, a zero mean log residual would not imply an unbiased reconstruction on the FI scale. Back-transformation should

therefore be assessed through conditional means, medians and quantiles instead of assumed unbiased from log-scale centring.

Reconstructed FI can leave the group used to generate the record. Without an explicit post-reconstruction rule, assigning a record to the original generator group and assigning it by its reconstructed FI need not give the same result. Rejecting migrated samples changes the residual distribution; relabelling changes the output group counts. The benchmark summary does not identify that rule, so neither procedure is retrospectively asserted here.

3.3 Constraints and their statistical consequences

The reported admissibility rules are xB > 0, x50 > 0, PF > 0 and xB ≥ x50 [8]. The first three exclude values incompatible with the adopted definitions. The ordering condition is a domain-informed restriction for this dataset, not a universal physical law for every representative block-size and fragment-size statistic. Its validity depends on compatible measurement definitions and whether the comparison concerns the same rock volume and size-distribution convention.

Minimum-maximum enforcement is a data-range constraint. It does not demonstrate physical validity beyond the calibration sample, and it prevents exploration of tails outside the observed bounds [9]. Rejection sampling has a different effect: it conditions the fitted distribution on acceptance. If acceptance depends on FI regime or covariate combination, the retained mixture may differ from the intended one. Rejection counts, acceptance rates and post-filter group weights should therefore accompany fidelity scores.

Identical support rules should be considered when comparing a constrained method with Gaussian simulation. An unconstrained Gaussian can generate negative values and infeasible combinations; a rejection-truncated or transformed Gaussian baseline may be a more informative comparator. Without holding these rules constant, the observed advantage combines the effects of generator family, grouping, reconstruction, range enforcement and filtering. The present comparison evaluates the complete workflow rather than isolating one component.

3.4 Fidelity evaluation

The reported SDMetrics assessment summarises Column Shapes and Column Pair Trends [10]. Numerical shapes are commonly assessed through a complement of the two-sample KS statistic, while discrete or categorical comparisons use an appropriate distributional similarity measure. Pair-trend metrics depend on column types and the implementation version. A score near one indicates agreement under those specific diagnostics, not a probability that the synthetic table is scientifically correct.

These diagnostics are useful for detecting marginal and pairwise departures. They do not establish higher-order consistency, causal relationships or accurate conditional tail probabilities. The overall score in the benchmark is the arithmetic mean of the two reported property scores, up to rounding. Thus, an increase can be driven primarily by marginals even when pairwise agreement changes little. The property scores should be interpreted alongside the aggregate.

3.5 Utility tasks and predictor availability

Random Forest was used for TSTR and TRTS utility assessment [8,11]. Regression predicted x50 and was summarised by R^2 and MAE, the latter in metres under the reported schema. Classification used

the binary label xB > 0.15 m and reported accuracy, F1 and AUROC. This is a threshold on the pre-blast in-situ block-size descriptor. It is not a direct observation of the oversize fraction or maximum particle size in the post-blast product.

The classification task becomes deterministic if xB is included among its inputs. Removing xB alone may also be insufficient when FI and x50 reconstruct it closely. Similarly, if FI and xB are supplied when predicting x50, Eq. (4) embeds the target within a predictor constructed from that target. Such a task can test preservation of algebraic relationships, but it cannot support a prospective forecast. The source summary does not give task-specific feature lists, so the utility scores are interpreted subject to this unresolved predictor audit.

3.6 Partition chronology and uncertainty

An independent TSTR assessment requires the real partition to precede every data-fitted operation: group proportions, scalers, residual pools, generator fitting, range estimates, support diagnostics used for tuning and downstream model selection [12,13]. The real evaluation rows must not contribute to a residual pool used to synthesize training data. TRTS provides a useful reverse-transfer diagnostic, but its test data are generated and do not replace a real hold-out set.

The benchmark summary does not document the real split counts, complete partition chronology, generator seeds, repetition counts or confidence intervals. The results below are therefore descriptive historical scores. They do not establish that the real evaluation set was contaminated; equally, the labels TSTR and TRTS alone cannot certify independence. With only 62 records, repeated evaluation should resample independent real blasts and refit the full pipeline, rather than attach uncertainty based on the much larger number of generated rows.

3.7 Support precision and recall

The project reports alpha-Precision and beta-Recall computed through nearest-neighbour support approximations in standardised feature space [8]. These are related to the fidelity and coverage concepts developed by Alaa et al. [14]. A complete implementation also requires the neighbourhood size, distance metric, support or quantile thresholds, scaling fit and reference-sample definitions. Without these, the scores should be described as the reported support diagnostics rather than assumed numerically equivalent to every implementation of the published metrics.

Precision reflects how much synthetic support is consistent with the real reference; recall reflects how much of that real reference is covered. Neither establishes that a record is novel. Originality or authenticity needs a separate comparison with training neighbours and held-out real neighbours [14]. The absence of such an audit is particularly relevant when group models are fitted to very few observations.

4 Results

4.1 Marginal and pairwise fidelity

The grouped workflow achieved an overall score of 0.8568, compared with 0.8261 for pure TVAE and 0.8023 for Gaussian simulation. Table 2 retains every original property score. The grouped workflow's Column Shapes score increased by 0.0640 relative to pure TVAE, while its Column Pair

Trends score decreased by 0.0024. The overall advantage of approximately 0.0307 is consequently driven by marginal similarity rather than superior performance on both diagnostic axes.

Table 2. Reported SDMetrics fidelity scores. Higher is better for these similarity summaries. Values reproduce the benchmark and have no accompanying confidence intervals; the overall value averages the two property scores up to rounding.

Data source	Property	Score
Proposed synthetic workflow	Overall	0.8568
Proposed synthetic workflow	Column Shapes	0.8027
Proposed synthetic workflow	Column Pair Trends	0.9110
Pure TVAE	Overall	0.8261
Pure TVAE	Column Shapes	0.7387
Pure TVAE	Column Pair Trends	0.9134
Gaussian baseline	Overall	0.8023
Gaussian baseline	Column Shapes	0.7849
Gaussian baseline	Column Pair Trends	0.8197

The Gaussian baseline's overall score of 0.8023 is not negligible despite its weaker support diagnostics. This illustrates why aggregate resemblance should not be treated as a physical-validity certificate. A simulator may retain broad column shapes or pairwise trends while allocating probability to implausible combinations, especially outside the observed support.

4.2 Regression transfer

In TSTR, the grouped workflow attained R^2 = 0.7030 and MAE = 0.0364 m. Pure TVAE attained 0.3350 and 0.0674 m, and Gaussian simulation attained -0.2804 and 0.0820 m. The grouped workflow's MAE is approximately 46.0% lower than pure TVAE's, calculated directly from the reported rounded scores. This is a descriptive contrast between synthetic workflows, not a demonstrated gain over a matched real-only predictor.

Table 3. Reported regression utility for the fragment-size field x50. MAE is in metres under the retained data schema. TSTR and TRTS evaluate different test distributions; their R^2 values do not share a common variance denominator.

Data source	Protocol	R^2	MAE
Proposed synthetic workflow	TSTR	0.7030	0.0364
Proposed synthetic workflow	TRTS	0.1302	0.0387
Pure TVAE	TSTR	0.3350	0.0674
Pure TVAE	TRTS	-0.0147	0.0519
Gaussian baseline	TSTR	-0.2804	0.0820
Gaussian baseline	TRTS	-0.3201	0.1070

Reverse transfer was weaker: the grouped workflow attained TRTS R^2 = 0.1302 with MAE = 0.0387 m. A low R^2 accompanied by a similar MAE is compatible with a narrower synthetic target distribution, since R^2 scales squared errors by target variance. Conditional-relationship mismatch, learner

differences and tail behaviour are also possible explanations. The table alone does not identify which mechanism dominates, and it does not prove that synthetic tails are uniformly underrepresented.

4.3 Block size threshold classification

The grouped workflow attained TSTR accuracy = 0.9194, F1 = 0.9580 and AUROC = 0.9930; its TRTS values were 0.9892, 0.9945 and 1.0000. Pure TVAE produced higher TSTR accuracy and F1 but weaker regression transfer. All reported classification values are preserved in Table 4, with the task explicitly identified as a block-size threshold proxy.

Table 4. Reported utility for the label xB > 0.15 m, described as oversize risk in the original experiment. The label measures an in-situ-block-size threshold and is not a validated post-blast oversize outcome. Scores require the feature-exclusion and class-balance checks discussed in Sections 3.5 and 4.3.

Data source	Protocol	Accuracy	F1	AUROC
Proposed synthetic workflow	TSTR	0.9194	0.9580	0.9930
Proposed synthetic workflow	TRTS	0.9892	0.9945	1.0000
Pure TVAE	TSTR	0.9839	0.9913	1.0000
Pure TVAE	TRTS	0.9839	0.9918	0.8376
Gaussian baseline	TSTR	0.9194	0.9580	0.9561
Gaussian baseline	TRTS	0.9516	0.9752	0.4505

Accuracy of 0.9194 and F1 of 0.9580 are also compatible, after rounding, with an always-positive classifier on 62 cases containing 57 positives. This arithmetic example does not recover the actual confusion matrix or class counts; it demonstrates why the two reported scores alone cannot exclude a majority-class explanation. AUROC measures ranking across thresholds and can remain high even when a chosen decision threshold performs poorly for the minority class. Confusion matrices, class-specific recall, precision-recall curves and prevalence are needed before interpreting the task as a useful engineering classifier.

4.4 Support coverage

The grouped workflow's overall alpha-Precision and beta-Recall were 0.9247 and 0.9194. Pure TVAE had similar precision, 0.9194, but lower recall, 0.7581. Gaussian simulation had precision = 0.2957 and recall = 0.8226. These values describe distinct trade-offs between producing records close to the reference support and covering that support.

Table 5. Reported nearest-neighbour support precision and recall, overall and by FI group. Every original score and endpoint convention is retained. Small group counts and unspecified support thresholds limit quantitative interpretation of between-group differences.

Data source	FI group	alpha Precision	beta Recall
Proposed synthetic workflow	All	0.9247	0.9194
Proposed synthetic workflow	FI <= 4	0.9853	0.9200
Proposed synthetic workflow	4 < FI < 8	0.9726	0.9565
Proposed synthetic workflow	FI >= 8	0.9111	0.9286
Pure TVAE	All	0.9194	0.7581

Data source	FI group	alpha Precision	beta Recall
Pure TVAE	FI <= 4	0.9623	0.7200
Pure TVAE	4 < FI < 8	0.9697	0.6087
Pure TVAE	FI >= 8	1.0000	0.9286
Gaussian baseline	All	0.2957	0.8226
Gaussian baseline	FI <= 4	0.4561	0.8000
Gaussian baseline	4 < FI < 8	0.6023	0.9565
Gaussian baseline	FI >= 8	0.2683	0.7143

The high-FI precision of 1.0000 for pure TVAE is not evidence of complete representation or absence of memorisation. A concentrated generator can attain high precision while missing valid regions, and a small reference set can make scores sensitive to individual records. The grouped method's more balanced overall scores are favourable descriptively, but uncertainty and authenticity remain unmeasured.

5 Discussion

5.1 What the grouped formulation contributes

The grouped workflow offers a coherent way to impose local structure on a heterogeneous table. It allows each model to operate within a narrower response regime and then reconstructs FI with a transparent relationship. That structure can help retain marginal behaviour, consistent with the observed Column Shapes gain. It may also increase estimation variance because each generator sees fewer real records. A fallback copula reduces model complexity, but it does not resolve missing regimes or establish the correctness of the chosen group boundaries.

The comparison does not support attribution of the observed improvement specifically to TVAE. Grouping, residual reconstruction, range enforcement, rejection and the fallback generator all change together. A factorial or sequential ablation would compare the same base generator with and without each component, under matched real partitions, generated sample counts and filtering. Equal treatment of the baselines would clarify how much gain comes from domain restrictions rather than neural representation learning.

5.2 Residual modelling as a measurement problem

The residual construction is interpretable because it makes a familiar ratio explicit. Its reliability depends on why the residual exists. If departures arise primarily from rounding or digitisation, the appropriate response may be to reconcile the source data and propagate known precision. If they represent measurement uncertainty, its magnitude may depend on fragmentation method, rock type or size range. If FI has a genuinely different definition, residual reconstruction around Eq. (1) could impose the wrong relationship.

A useful next analysis would preserve the original measurement definitions and compare residuals by source, group and size range. Conditional residual sampling or a hierarchical residual model could then be evaluated against independent empirical sampling. These extensions require data; they cannot be justified solely because a more complicated model has greater flexibility. The present

formulation should be understood as a transparent empirical approximation with an explicitly testable exchangeability assumption.

5.3 The limits of the utility evidence

The positive TSTR result shows transfer within the recorded evaluation protocol, but a scientifically defensible prediction claim depends on the final inputs and their availability. FI grouping and FI reconstruction both use the response variable x50. This is acceptable within a training-only joint generator. It becomes circular when FI is later used to predict the same x50 and is constructed from evaluation information or supplied as a post-blast descriptor in a pre-blast task. These are different routes to optimistic interpretation and should be audited separately.

The block-size label poses a related but simpler problem. Predicting whether an observed xB exceeds 0.15 m is not an oversize-product forecast if xB is itself an input. Even without that variable, a proxy target should be validated against an actual engineering outcome such as the post-blast fraction above a declared screen size. The preserved classification results are therefore useful as reported transfer diagnostics but cannot carry the manuscript's operational claim.

5.4 Support restrictions and rare regimes

Minimum-maximum enforcement and the block-to-fragment-size ordering make the sample more conservative by construction. The same restrictions can suppress valid boundaries or rare combinations if the empirical bounds are narrow. A high-FI group with few observations is especially susceptible: the apparent regime can reflect one site, a measurement convention or a handful of extreme records. Sampling more synthetic examples from it reduces simulation noise without increasing the number of independent examples supporting its distribution.

Rejection can also shift the relative frequency of regimes. Sampling a fixed number of accepted records per group may intentionally rebalance the table, while sampling once and dropping rejected rows can change it unintentionally. These choices affect both fidelity and downstream utility. Reporting pre-filter and post-filter distributions would make the intended target population clear and reveal whether good scores partly reflect the removal of difficult cases.

5.5 A stronger engineering evaluation

The next validation should define separate purposes for joint distribution replay and pre-blast prediction. For replay, all variables may be generated together and tested for constraints, marginals, dependencies, regime weights and authenticity. For prediction, the final learner should receive a documented pre-blast feature set and be compared with real-only, synthetic-only and hybrid alternatives on identical independent real rows. Generator selection must occur inside the training boundary [12,13].

A source- or campaign-blocked design is preferable when the intended use involves a new operation [15]. At minimum, repeated outer partitions should refit every fitted component and report paired real-test differences. Uncertainty should be based on independent blasts or operational groups, not synthetic rows. Publishing group counts, rejection rates, residual pools, software versions, random seeds and task-specific features would make the full workflow inspectable. Sample-level originality tests are also needed before describing it as a source of diverse new cases or assuming confidentiality from synthesis [14,16].

6 Conclusions

The reported grouped and reconstructed workflow provides the strongest overall fidelity and TSTR regression results among the three synthetic alternatives examined. Its fidelity advantage is predominantly marginal, while its support diagnostics are more balanced than those of pure TVAE or unconstrained Gaussian simulation. The reverse-transfer gap demonstrates that utility in one direction does not establish complete distributional equivalence.

The main contribution is an interpretable synthesis structure and a clearer account of its assumptions. Existing scores do not isolate each component, establish improvement over real-only prediction or validate post-blast oversize risk. Resolving FI semantics, documenting training-only residual construction, excluding circular predictors and testing independent real blasts are the priorities for establishing engineering validity. Physical constraints strengthen a generator only when their definitions, statistical effects and evaluation scope are explicit.

References

[1] Kulatilake, P.H.S.W.; Qiong, W.; Hudaverdi, T.; Kuzu, C. (2010). Mean particle size prediction in rock blast fragmentation using neural networks. Engineering Geology, 114(3-4), 298-311. https://doi.org/10.1016/j.enggeo.2010.05.008

[2] Hudaverdi, T.; Kuzu, C.; Fisne, A. (2012). Investigation of the blast fragmentation using the mean fragment size and fragmentation index. International Journal of Rock Mechanics and Mining Sciences, 56, 136-145. https://doi.org/10.1016/j.ijrmms.2012.07.028

[3] Krop, I.; Sasaoka, T.; Shimada, H.; Hamanaka, A. (2024). Optimizing mean fragment size prediction in rock blasting: A synergistic approach combining clustering, hyperparameter tuning, and data augmentation. Eng, 5(3), 1905-1936. https://doi.org/10.3390/eng5030102

[4] Fonseca, J.; Bacao, F. (2023). Tabular and latent space synthetic data generation: A literature review. Journal of Big Data, 10, 115. https://doi.org/10.1186/s40537-023-00792-7

[5] Kingma, D.P.; Welling, M. (2014). Auto-Encoding Variational Bayes. International Conference on Learning Representations. https://arxiv.org/abs/1312.6114

[6] Xu, L.; Skoularidou, M.; Cuesta-Infante, A.; Veeramachaneni, K. (2019). Modeling tabular data using conditional GAN. Advances in Neural Information Processing Systems, 32. https://proceedings.neurips.cc/paper/2019/hash/254ed7d2de3b23ab10936522dd547b78-Abstract.html

[7] Nelsen, R.B. (2006). An Introduction to Copulas. 2nd edition. Springer. https://doi.org/10.1007/0-387-28678-0

[8] Yu, K. (2025). MSc Independent Research Project report. Imperial College London. Project identifier irp-ky1024. Experimental source identified in the accompanying manuscript. Repository recorded as https://github.com/ese-ada-lovelace-2024/irp-ky1024

[9] DataCebo. TVAESynthesizer. Synthetic Data Vault documentation. https://docs.sdv.dev/sdv/modeling/single-table-synthesizers/tvaesynthesizer (accessed 28 October 2026).

[10] DataCebo. Quality Report API. SDMetrics documentation. https://docs.sdv.dev/sdmetrics/data-metrics/quality/quality-report-api (accessed 28 October 2026).

[11] Esteban, C.; Hyland, S.L.; Ratsch, G. (2017). Real-valued (medical) time series generation with recurrent conditional GANs. arXiv:1706.02633. https://doi.org/10.48550/arXiv.1706.02633

[12] Kapoor, S.; Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. Patterns, 4(9), 100804. https://doi.org/10.1016/j.patter.2023.100804

[13] Cawley, G.C.; Talbot, N.L.C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. Journal of Machine Learning Research, 11, 2079-2107. https://jmlr.org/papers/v11/cawley10a.html

[14] Alaa, A.; Van Breugel, B.; Saveliev, E.S.; van der Schaar, M. (2022). How faithful is your synthetic data? Sample-level metrics for evaluating and auditing generative models. Proceedings of Machine Learning Research, 162, 290-306. https://proceedings.mlr.press/v162/alaa22a.html

[15] Roberts, D.R.; Bahn, V.; Ciuti, S.; et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40(8), 913-929. https://doi.org/10.1111/ecog.02881

[16] Stadler, T.; Oprisanu, B.; Troncoso, C. (2022). Synthetic data - Anonymisation groundhog day. 31st USENIX Security Symposium, 1451-1468. https://www.usenix.org/conference/usenixsecurity22/presentation/stadler