KnowledgeTrend
Preprints • Research shared early

PREPRINT | Version 1
Posted: 27 September 2026
ID: KTPR/2026/00002-1F98

Certified Conjugacy Learning in Small-Cancellation Groups

Michael Nsikan John[1,*]

[1] Department of Mathematics, Edo State University Iyamho, Edo State, Nigeria

[*]**Corresponding author:** Michael Nsikan John | john.michael@edouniversity.edu.ng

Abstract

The conjugacy decision problem is fundamental in combinatorial group theory and has applications in algebraic cryptography. Previous machine-learning studies demonstrated that conjugacy can be classified in selected non-free groups, but a statistical prediction is not an algebraic proof. This paper introduces a reproducible certificate-guided learning pipeline for random two-generator one-relator presentations satisfying the metric condition $C'(1/6)$. Positive instances are constructed with an explicit conjugator and accepted only after Dehn reduction verifies $g^{-1}ug =_G v$. Negative instances are retained only when the images of u and v in the abelianization are unequal, which is an exact obstruction to conjugacy. A symmetry-preserving feature map combines pairwise length statistics, exponent-sum information and aggregated labelled-walk features of orders one to three. Logistic regression and a random forest are evaluated under a presentation-disjoint split. The principal experiment contains 4,200 certified pairs from six presentations; 2,800 pairs from four presentations are used for training and 1,400 pairs from two unseen presentations are used for testing. Logistic regression obtains accuracy 0.9964 and F1 score 0.9964, while the random forest obtains 1.0000 for accuracy, F1 and ROC–AUC. A second-seed experiment reproduces the result. These high scores characterize the certified pilot distribution and do not solve unrestricted conjugacy. The main contribution is a transparent protocol that separates prediction from proof, prevents unverifiable labels and identifies the construction of difficult certified negative instances as the central next problem. **Keywords:** conjugacy problem; small-cancellation groups; certified machine learning; Dehn reduction; algebraic verification; Cayley graphs

Preprint record: Version 1 | Posted 27 September 2026 | ID: KTPR/2026/00002-1F98

1. Introduction

The conjugacy decision problem asks whether two words u and v represent conjugate elements of a group G. Given a finite presentation

$$G = \langle X \mid R \rangle,$$

the problem is to determine whether an element $g \in G$ exists such that

$$v =_G g^{-1}ug. \tag{1}$$

The associated search problem asks for such a g. Both problems connect algebraic structure, geometry and computational complexity. Their relevance to noncommutative cryptography has motivated algorithmic and experimental investigations [4, 11]. The author's earlier work examined conjugacy classes in finitely generated groups with small-cancellation properties [6] and used group-theoretic structures in several cryptographic settings [7–9].

Machine learning offers fast predictions but ordinarily supplies no proof that Equation (1) holds. This distinction is decisive in computational algebra. A false positive cannot be accepted merely because a classifier is confident. The present study consequently treats learning as a proposal mechanism placed inside a mathematically controlled pipeline. Every positive observation has a witness, and every positive certificate is checked by deterministic reduction. Every negative observation used in training has an exact abelian obstruction.

The study makes four contributions. First, it constructs a reproducible collection of certified conjugacy instances for random $C'(1/6)$ presentations. Second, it defines a symmetric pair representation from length, abelian and local labelled-walk statistics. Third, it evaluates two transparent baseline learners under a presentation-disjoint protocol. Fourth, it reports the limitations of the resulting near-perfect performance rather than interpreting it as a solution of the unrestricted problem.

The qualified novelty is the combination of $C'(1/6)$ sampling, verified conjugator construction, certified negative selection and testing on unseen presentations. Machine learning for conjugacy itself is not new [4]; learning group properties from Cayley graphs is also emerging rapidly [13, 15, 16]. The contribution is therefore a certified experimental protocol rather than a claim that neural or statistical methods originated here.

2. Related work

Small-cancellation theory provides presentations in which long overlaps between distinct relators are prohibited. The classical results of Greendlinger and subsequent treatments establish strong geometric and algorithmic consequences, including Dehn-type solutions of the word problem [3, 10]. Conjugacy and powers in small-cancellation groups have also received direct study [1]. The author's previous investigation discussed conjugacy classes, residual finiteness and quasi-isometric considerations in finitely generated groups with small-cancellation properties [6].

Gryak, Haralick and Kahrobaei applied decision trees, random forests and N-tuple neural networks to the conjugacy decision problem in several finitely presented non-free groups [4]. Their work established the feasibility of statistical conjugacy classification. The present study differs in its selected group class, presentation-disjoint evaluation and insistence upon exact certificates for dataset membership.

Equivariant learning incorporates known symmetries into an architecture or feature map. LieTransformer provides equivariant self-attention for Lie groups [5], while graph-automorphism equivariant networks formalize parameter sharing with respect to graph symmetries [13]. Recent work has learned solvability and other algebraic properties from finite Cayley graphs and explored transfer to infinite groups [14–16]. Other studies use Cayley graphs to improve message propagation [2, 17] or guide pathfinding [12]. None of these contributions, as located in the literature search conducted for this study, combines verified conjugator construction with certified training labels for random $C'(1/6)$ presentations.

3. Algebraic setting

Let $X = \{a, b\}$ and let

$$\mathcal{A} = X \cup X^{-1} = \{a, a^{-1}, b, b^{-1}\}.$$

A word is freely reduced if it has no adjacent inverse pair. If r is a cyclically reduced relator, its symmetrized closure contains all cyclic shifts of r and r^{-1}.

Definition 1. *A nonempty word p is a* piece *if it is an initial segment of two distinct elements of the symmetrized relator set. A presentation satisfies $C'(\lambda)$ when every piece p contained in a relator r satisfies $|p| < \lambda|r|$.*

Only one cyclically reduced relator of length 42 is used in each sampled presentation. Every cyclic shift of the relator and its inverse is compared, and a presentation is admitted only if

$$6p_{\max} < 42, \tag{2}$$

where $p_{\max}$ is the greatest observed piece length. This direct finite check avoids assuming the required condition from random generation alone.

Write $\mathcal{S}(r)$ for the symmetrized closure of r. For distinct $s, t \in \mathcal{S}(r)$, define

$$\ell(s,t) = \max\{j \geq 0 : s_1 \cdots s_j = t_1 \cdots t_j\}.$$

The program evaluates

$$p_{\max} = \max_{s \neq t \in \mathcal{S}(r)} \ell(s,t). \tag{3}$$

Since $|\mathcal{S}(r)| \leq 2|r|$, the direct test uses $O(|r|^3)$ symbol comparisons in the worst case. At the fixed length used in the experiment, this exhaustive calculation is preferable to a probabilistic shortcut.

For a symmetrized relator written as $r = st$ with $|s| > |r|/2$, the Dehn rule replaces s by t^{-1}. Free reduction is applied after every replacement. Under $C'(1/6)$, this procedure solves the word problem for the presentations considered here [3, 10].

Proposition 1 (Termination of the implemented reduction)**.** *Every successful Dehn replacement strictly decreases word length. Reduction therefore terminates after at most $|w|$ successful replacements for an input word w.*

Proof. If $r = st$ and $|s| > |r|/2$, then $|t| = |r| - |s| < |s|$. Replacing s by t^{-1} changes the current length by $-|s| + |t| < 0$. Free reduction can only shorten the word further. Since word length is a nonnegative integer, an infinite sequence of replacements is impossible. □

For current word length N and relator length L, the explicit table has $O(L^2)$ left-hand sides. The transparent list-based scan has a coarse worst-case bound $O(N^2L^3)$. Trie matching could reduce this cost, but is unnecessary at the tested lengths.

3.1 Positive certificates

Choose freely reduced words w, h, k and form

$$u = h^{-1}wh, \qquad v = k^{-1}wk. \tag{4}$$

Then $g = h^{-1}k$ is a conjugator because

$$g^{-1}ug = k^{-1}hh^{-1}whh^{-1}k = k^{-1}wk = v.$$

The implementation does not rely only on this symbolic construction. It reduces

$$g^{-1}ugv^{-1} \tag{5}$$

with the Dehn procedure and admits the pair only when the result is the empty word.

Proposition 2. *Every positive observation admitted by the generator is accompanied by a machine-checked conjugator certificate.*

Proof. The generator stores $g = h^{-1}k$ and evaluates Equation (5). Admission requires its Dehn normal form to be empty. Since Dehn reduction solves the word problem under the checked $C'(1/6)$ condition, the accepted equality holds in G. □

3.2 Negative certificates

Conjugate elements have identical images in every abelian quotient. Let

$$\pi : G \longrightarrow G_{\mathrm{ab}} = G/[G, G]$$

be the abelianization map. For a two-generator one-relator presentation, exponent-sum vectors represent elements of $\mathbb{Z}^2/\langle \mathbf{e}(r) \rangle$, where $\mathbf{e}(r)$ is the exponent-sum vector of r. A pair is admitted as negative only when

$$\mathbf{e}(u) - \mathbf{e}(v) \notin \langle \mathbf{e}(r) \rangle. \tag{6}$$

Equation (6) proves that $\pi(u) \neq \pi(v)$ and therefore proves that u and v are not conjugate.

Let

$$\mathbf{e}(w) = \begin{pmatrix} \sigma_a(w) \\ \sigma_b(w) \end{pmatrix},$$

where $\sigma_x(w)$ counts occurrences of x minus occurrences of x^{-1}. In the one-relator case,

$$G_{\mathrm{ab}} \cong \mathbb{Z}^2/\langle \mathbf{e}(r) \rangle.$$

Thus u and v have the same abelian image precisely when some $q \in \mathbb{Z}$ satisfies

$$\mathbf{e}(u) - \mathbf{e}(v) = q\mathbf{e}(r). \tag{7}$$

The implementation handles zero coordinates separately and requires every nonzero-coordinate ratio to give the same integer.

Proposition 3 (Soundness of negative certification)**.** *If Equation (7) has no integer solution, then u and v are not conjugate in G.*

Proof. If $v = g^{-1}ug$, then $\pi(v) = -\pi(g) + \pi(u) + \pi(g) = \pi(u)$ in the abelianization. Equality of the two cosets is equivalent to Equation (7); failure of that equation contradicts conjugacy. □

Remark 1. *This negative certificate is intentionally conservative. It excludes nonconjugate pairs having equal abelian images. The resulting benchmark measures learning on a certified subset and is not a complete model of the conjugacy decision problem.*

4. Feature construction and learning models

For each pair (u, v), the feature map begins with

$$(|u|, |v|, ||u| - |v||, \min\{|u|, |v|\}, \max\{|u|, |v|\}).$$

It then appends the absolute difference and sum of the exponent vectors. For $q \in \{1, 2, 3\}$, let $H_q(w)$ denote the normalized histogram of directed labelled walks of length q along the word path. The pair representation contains

$$\Phi_q(u, v) = (|H_q(u) - H_q(v)|, H_q(u) + H_q(v)). \tag{8}$$

This sum-and-absolute-difference readout is invariant under exchanging u and v. Aggregation inside each histogram is invariant under the enumeration of occurrences. It is best understood as a fixed message-passing baseline, not as a fully learned equivariant GNN.

For an alphabet of size four, $H_q(w) \in \mathbb{R}^{4^q}$. Consequently, the complete feature vector has

$$5 + 2 + 2 + 2(4 + 16 + 64) = 177$$

coordinates. The two exponent-vector blocks have dimension two each. Histogram entries are divided by $\max\{1, |w| - q + 1\}$, so the corresponding coordinates represent relative rather than raw frequencies. If P exchanges the two input words, then

$$|H_q(v) - H_q(u)| = |H_q(u) - H_q(v)|, \quad H_q(v) + H_q(u) = H_q(u) + H_q(v),$$

and hence $\Phi_q(P(u, v)) = \Phi_q(u, v)$. This proves pair-exchange invariance of the local-walk component.

Two classifiers are fitted. The first is logistic regression with a maximum of $3,000$ optimization iterations. The second is a random forest with 400 trees, minimum leaf size two, balanced class weights and a fixed random seed. These baselines were selected because their behaviour is easier to audit than that of a large neural system. A learned graph architecture is deferred until a harder certified dataset is available.

For logistic regression, the conditional estimate has the form

$$\widehat{\Pr}(Y = 1 \mid z) = \frac{1}{1 + \exp[-(\beta_0 + \beta^{\mathsf{T}} z)]},$$

where $z \in \mathbb{R}^{177}$. Classification uses the standard threshold $1/2$. Each random-forest tree recursively partitions the same feature space; the forest probability is the mean of the class probabilities produced by the 400 trees.

5. Reproducible experimental protocol

Six independent presentations are generated with master seed 20,260,927. Each presentation contributes 350 positive and 350 negative observations. Relators have length 42; core words have length 5–16; conjugating words have length 2–8; and initial negative words have length 9–28. The code restricts the absolute length difference of negative pairs to at most six so that gross length imbalance alone is not sufficient.

All pairs from four presentations form the training collection. All pairs from the remaining two presentations form the test collection. This division is stricter than a random pair-level split because no test relator occurs during fitting. Table 1 summarizes the design.

The implementation is written in Python and uses NumPy for arrays and scikit-learn for the classifiers and metrics. Relator generation, free reduction, Dehn rules, abelian certification and feature construction are implemented directly rather than delegated to a computer-algebra system. This choice makes every algebraic transformation visible in the accompanying source. The principal run completes in approximately 8.08 seconds in the execution environment, excluding document compilation.

Table 1: *Certified dataset and split.*

Partition	Presentations	Positive pairs	Negative pairs
Training	4	1,400	1,400
Testing	2	700	700
Total	6	2,100	2,100

The maximum piece lengths of the six accepted relators are $4, 5, 6, 6, 5, 5$. Each satisfies Equation (2). All $2,100$ positive conjugator certificates and all $2,100$ negative abelian certificates pass verification.

Algorithm 1: Certificate-guided instance generation

1. Supply seed s, number of presentations m and pairs per class n.
2. For each presentation, sample a cyclically reduced relator r_i until $C'(1/6)$ is verified, then construct its Dehn replacement table.
3. For each positive pair, sample w, h, k; set $u = \mathrm{red}(h^{-1}wh)$, $v = \mathrm{red}(k^{-1}wk)$ and $g = \mathrm{red}(h^{-1}k)$; retain the pair only if $\mathrm{red}(g^{-1}ugv^{-1}) = \varepsilon$.
4. For each negative pair, sample and reduce u, v; retain the pair only if their abelian images differ.

Accuracy, balanced accuracy, precision, recall, F1 score, ROC–AUC and confusion matrices are reported. A second run uses seed 20,260,928 and 150 pairs per class per presentation. Five unit tests examine inverse reduction, the $C'(1/6)$ test, relator reduction, positive-certificate verification and abelian separation.

6. Results

Table 2 gives the presentation-disjoint test results. Logistic regression misclassifies five negative observations as positive and recognizes every positive observation. The random forest classifies all $1,400$ test pairs correctly.

Table 2: *Main test results on two unseen presentations.*

Model	Accuracy	Balanced acc.	Precision	Recall	F1	ROC–AUC
Logistic regression	0.9964	0.9964	0.9929	1.0000	0.9964	1.0000
Random forest	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

The confusion matrices, with rows denoting true labels $0, 1$ and columns denoting predicted labels $0, 1$, are

$$C_{\text{logistic}} = \begin{pmatrix} 695 & 5 \\ 0 & 700 \end{pmatrix}, \qquad C_{\text{forest}} = \begin{pmatrix} 700 & 0 \\ 0 & 700 \end{pmatrix}.$$

The robustness run contains $1,800$ certified observations, of which $1,200$ are used for training and 600 for testing. Logistic regression attains accuracy 0.9950, F1 0.9950 and ROC–AUC 1.0000. Its confusion matrix is

$$\begin{pmatrix} 297 & 3 \\ 0 & 300 \end{pmatrix}.$$

The random forest again attains 1.0000 on every reported classification metric. All five unit tests pass.

7. Discussion

The experiment establishes that a complete, reproducible and certificate-controlled pipeline can be constructed for small-cancellation presentations. It also shows that a model fitted to four presentations can classify the certified distribution in two unseen presentations. The repeated result across two seeds indicates that the observation is not caused by one accidental train–test draw.

Nevertheless, the perfect random-forest result must be interpreted cautiously. Negative labels are defined through an abelian obstruction, and exponent-sum features are supplied to the classifiers. The learner can therefore recover much of the same separation used by the certificate generator. The result validates implementation consistency and cross-presentation transfer on this controlled distribution; it

does not show that the model has learned the deeper geometry of conjugacy classes.

This apparent weakness produces a useful methodological conclusion. Reporting a high score without stating how negative pairs were certified would be misleading. Removing exponent-sum features would not repair the fundamental issue because short labelled-walk statistics may still correlate with the same obstruction. A genuinely demanding second-generation benchmark needs nonconjugate pairs satisfying

$$\pi(u) = \pi(v)$$

and matching other inexpensive invariants. Such pairs require a complete conjugacy algorithm, a theoretically justified bounded search, or independently checkable nonconjugacy certificates.

The current fixed aggregation is also not a fully trainable graph neural network. It provides symmetry-respecting local statistics and an auditable baseline. A future architecture can operate on paired, rooted neighbourhoods of $\mathrm{Cay}(G, X)$, share parameters across generator-labelled edges and generate candidate conjugators autoregressively. Every generated candidate should still pass exact verification before acceptance.

8. Threats to validity and limitations

Four limitations delimit the conclusions. First, only two-generator, one-relator presentations of a fixed relator length are studied. Second, negative certification through abelianization selects comparatively easy cases. Third, the experiment predicts the decision label but does not train a sequence model to recover a previously unseen conjugator. The generator nevertheless records and verifies the known witness for every positive instance. Fourth, six presentations are adequate for a pilot but insufficient for claims about the entire class of $C'(1/6)$ groups.

The Dehn reducer is an exact word-problem verifier under the checked presentation condition, but the implementation is optimized for clarity rather than asymptotic speed. The code scans a finite rule list repeatedly. More efficient automata or indexed matching would be preferable for large relator sets and long words.

9. Conclusion

This paper presents a reproducible certificate-guided machine-learning pilot for conjugacy decisions in random $C'(1/6)$ small-cancellation presentations. Every positive observation has a verified conjugator, every negative observation has an exact abelian obstruction, and the train–test split is disjoint at the presentation level. Logistic regression achieves 0.9964 accuracy and the random forest achieves 1.0000 on 1,400 test pairs from unseen presentations. A second-seed run reproduces these findings, and all algebraic unit tests pass.

The central result is not that unrestricted conjugacy has been solved. The result is that prediction and algebraic proof can be separated cleanly, with exact certificates governing dataset construction and acceptance. The next substantive advance is to construct hard certified negatives with equal abelian images and to train an equivariant graph model that proposes conjugators rather than only class labels.

Data and code availability

The accompanying Python program generates every presentation and observation from a recorded seed, verifies all certificates, fits both models and writes machine-readable JSON results. A separate unit-test file checks the principal algebraic operations. No private or externally collected dataset is used.

Version history

Version 1 (27 September 2026): Initial public preprint.

Status and citation

This document is a preprint and has not been peer reviewed. The recommended citation is: John, M. N. (2026). *Certified Conjugacy Learning in Small-Cancellation Groups* (Version 1) [Preprint]. KnowledgeTrend Preprints. Record KTPR/2026/00002-1F98.

Author contributions

Michael Nsikan John: conceptualization, methodology, software, validation, formal analysis, investigation, writing of the original draft, review and editing.

Funding

This research received no external funding.

Conflict of interest

The author declares no competing interests.

Ethical approval

Ethical approval was not required because the study used synthetically generated mathematical data and involved no human participants, animals or personal information.

References

[1] L. P. Comerford Jr., Powers and conjugacy in small cancellation groups, *Journal of the Australian Mathematical Society* 19 (1975), 349–356.

[2] A. Deac, M. Lackenby and P. Veličković, Expander graph propagation, in *Learning on Graphs Conference*, PMLR 198 (2022), 38:1–38:18.

[3] M. D. Greendlinger, On Dehn's algorithms for the conjugacy and word problems, with applications, *Communications on Pure and Applied Mathematics* 13 (1960), 641–677.

[4] J. Gryak, R. M. Haralick and D. Kahrobaei, Solving the conjugacy decision problem via machine learning, *Experimental Mathematics* 29 (2020), 66–78.

[5] M. J. Hutchinson, C. Le Lan, S. Zaidi, E. Dupont, Y. W. Teh and H. Kim, LieTransformer: Equivariant self-attention for Lie groups, in *Proceedings of the 38th International Conference on Machine Learning*, PMLR 139 (2021), 4533–4543.

[6] M. N. John, A. Musa and O. G. Udoaka, Conjugacy classes in finitely generated groups with small cancellation properties, *European Journal of Statistics and Probability* 12(1) (2023), 1–9. doi:10.37745/ejsp.2013/vol12n119.

[7] M. N. John, O. G. Udoaka and A. Musa, Nilpotent groups in cryptographic key exchange protocol for $n \geq 1$, *Journal of Mathematical Problems, Equations and Statistics* 4(2) (2023), 32–34. doi:10.22271/math.2023.v4.i2a.103.

[8] M. N. John, O. G. Udoaka and B. O. Nwala, Elliptic-curve groups in quantum-era cryptography, *ISAR Journal of Science and Technology* 1(1) (2023), 21–24. doi:10.5281/zenodo.10207536.

[9] M. N. John, O. G. Udoaka and A. Musa, Symmetric bilinear cryptography on elliptic curve and Lie algebra, *GPH–International Journal of Mathematics* 6(10) (2023), 1–15. doi:10.5281/zenodo.10200179.

[10] R. C. Lyndon and P. E. Schupp, *Combinatorial Group Theory*, Springer, Berlin, 1977.

[11] A. Mahalanobis, The Diffie–Hellman key exchange protocol and non-abelian nilpotent groups, *Israel Journal of Mathematics* 165 (2008), 161–187.

[12] A. D. Miasnikov and A. V. Ushakov, Pathfinding and reinforcement learning on Cayley graphs, arXiv:2502.18663 (2025).

[13] E. Pearce-Crump and W. Knottenbelt, Graph automorphism group equivariant neural networks, in *Proceedings of the 41st International Conference on Machine Learning*, PMLR 235 (2024), 40051–40077.

[14] T. Weissblat, Graph neural networks for predicting solvability of finite groups from Cayley graph representations, arXiv:2606.07619 (2026).

[15] T. Weissblat, A general framework for learning algebraic properties from Cayley graphs using graph neural networks, arXiv:2606.26212 (2026).

[16] T. Weissblat, From finite Cayley graphs to growth of infinite groups, arXiv:2607.02102 (2026).

[17] J. J. Wilson, M. Bechler-Speicher and P. Veličković, Cayley graph propagation, in *Proceedings of the Third Learning on Graphs Conference*, PMLR 269 (2025), 8:1–8:20.

KnowledgeTrend Preprints | Record: KTPR/2026/00002-1F98