KTREND JOURNALS
International Journal of Data Science and Machine Learning (IJDSML)

DOI: 10.5281/zenodo.20651402 **Volume:** 1 **Issue:** 1 **ISSN:** Pending

A Hybrid Algebraic Cryptographic Strength Index and Machine Learning Model for Predicting the Security of Algebraic Structures

Michael Nsikan John

Department of Mathematics, Edo State University, Iyamho, Nigeria

Corresponding Author: john.michael@edouniversity.edu.ng

Abstract

The choice of algebraic structure is a central problem in modern cryptography, especially in the transition from classical public-key schemes to quantum-resistant security models. Classical cryptographic systems such as RSA and elliptic curve cryptography are strongly connected to number theory, finite groups, conjugacy behavior, bilinear maps, and discrete logarithm assumptions; however, there is still no generally accepted data-driven framework for predicting the cryptographic strength of algebraic structures before deployment. This paper proposes a new hybrid modelling framework called the Algebraic Cryptographic Strength Index (ACSI), combined with supervised machine learning, for predicting whether a finite or computationally represented algebraic structure is suitable for cryptographic use. The framework transforms group-theoretic, character-theoretic, elliptic-curve, bilinear, lattice, and number-theoretic invariants into a structured feature space. A simulated dataset of 12,000 algebraic structures was generated using mathematically motivated constraints derived from conjugacy classes, centralizer ratios, derived length, monomial character indices, elliptic curve security margins, RSA modulus complexity, bilinear pairing depth, and lattice hardness. Four models were trained and evaluated: Random Forest, Gradient Boosting, Support Vector Machine, and Neural Network. The Neural Network achieved the best overall accuracy of 78.86% and ROC-AUC of 87.49%, while Gradient Boosting and Random Forest provided more interpretable feature-importance patterns. The results suggest that lattice hardness, elliptic curve security margin, small cancellation score, conjugacy density, and character-codegree entropy are among the most influential predictors. The study contributes a reproducible mathematical-data-science framework for algebraic cryptographic suitability prediction and opens a pathway for explainable machine learning in post-quantum cryptographic structure selection.

Keywords: Algebraic cryptography; Machine learning; Data science; Cryptographic strength modelling; Conjugacy classes; RSA; Elliptic curves; Lattice-based cryptography; Post-quantum security.

1 Introduction

Cryptographic systems are fundamentally mathematical systems. RSA depends on the arithmetic difficulty of integer factorization and modular exponentiation; elliptic curve cryptography relies on the discrete logarithm problem over elliptic curve groups; lattice-based cryptography uses geometric and algebraic hardness assumptions such as the shortest vector problem and learning with errors; and group-based cryptography studies problems involving conjugacy, word problems, centralizers, nilpotent groups, solvable groups, and group actions. The strength of a cryptographic primitive is therefore closely linked to the algebraic structure on which it is constructed.

The increasing interest in post-quantum cryptography has amplified the need for systematic methods of evaluating algebraic structures. Quantum algorithms, particularly Shor's algorithm, show that some classical number-theoretic and discrete-logarithm-based systems are not safe under large-scale quantum computation. This does not make algebra irrelevant; rather, it requires more careful selection of algebraic settings and complexity assumptions. Algebraic structures must now be evaluated according to both classical and quantum attack resistance, efficiency, implementation feasibility, and internal structural complexity.

A major difficulty is that algebraic structure selection is usually performed manually. Researchers inspect properties such as group order, conjugacy class distribution, character degrees, centralizer sizes, factorization complexity, curve parameters, or lattice dimension and then infer security suitability. This process is mathematically rich but computationally expensive and difficult to scale. The present paper therefore asks a data science question: can machine learning models learn patterns from algebraic invariants and predict the cryptographic strength of a structure?

The present study proposes a new hybrid framework that converts algebraic security indicators into a quantitative index called the Algebraic Cryptographic Strength Index (ACSI). The ACSI is not proposed as a replacement for formal cryptanalysis; rather, it is proposed as a screening and decision-support model for ranking algebraic structures before deeper theoretical and computational security analysis. The paper builds on earlier studies on conjugacy classes in finitely generated groups, solvable groups with monomial characters, elliptic-curve groups, RSA number theory, and symmetric bilinear cryptography. These works motivate the feature families used in the present model.

The main contributions of this paper are as follows:

(i) A new algebraic-data-science framework is proposed for predicting cryptographic suitability from group-theoretic and number-theoretic invariants.

(ii) A hybrid numerical score, the Algebraic Cryptographic Strength Index, is formulated

using conjugacy, character, elliptic curve, RSA, bilinear, lattice, and small cancellation indicators.

(iii) A reproducible simulated dataset of 12,000 algebraic structures is constructed for machine learning evaluation.

(iv) Four supervised learning models are compared using accuracy, precision, recall, F1-score, ROC-AUC, confusion matrices, and feature importance.

(v) The paper identifies the algebraic feature groups that contribute most strongly to cryptographic suitability prediction in the simulated model.

2 Related Literature

The relationship between algebra and cryptography is well established. Classical public-key cryptography uses number-theoretic structures, while modern post-quantum approaches explore lattices, codes, multivariate polynomial systems, hash functions, and group-based platforms. In RSA, the security mechanism arises from modular arithmetic and integer factorization, as discussed in the number-theoretic treatment of RSA encryption systems by John et al. [4]. Elliptic curve cryptography uses the algebraic group structure of points on elliptic curves over finite fields, a direction discussed in John, Udoaka, and Nwala [3].

Group-based cryptography has also attracted attention because group operations may generate hard computational problems. John, Musa, and Udoaka [1] examined conjugacy classes in finitely generated groups with small cancellation properties. Their work is relevant to the present study because conjugacy distribution, centralizer size, and small cancellation indicators are used as structural features in the proposed model. Similarly, John, Udoaka, and Musa [2] studied solvable groups with monomial characters of prime power codegree and monolithic characters, motivating the character-codegree and derived-length features used in this paper.

Bilinear and elliptic-curve constructions also provide useful algebraic indicators. John, Udoaka, and Musa [5] discussed symmetric bilinear cryptography on elliptic curves and Lie algebras. In this study, bilinear pairing depth and elliptic curve security margin are treated as measurable cryptographic indicators. Other works by John and collaborators on computational group theory, lattice-based cryptography, finite groups, and algebraic structures provide additional background for connecting algebraic invariants with cryptographic use [6, 7, 8, 9, 10].

Machine learning has been widely used in classification, anomaly detection, malware analysis, network security, side-channel analysis, and cryptographic implementation testing. However, less attention has been given to using machine learning as a decision-support tool for selecting algebraic structures for cryptographic design. The present paper fills this gap by treating algebraic cryptographic evaluation as a supervised learning problem.

3 Mathematical Background

3.1 Conjugacy Classes and Centralizers

Let G be a group and let $g \in G$. The conjugacy class of g in G is

$$\mathrm{Cl}_G(g) = \{xgx^{-1} : x \in G\}. \tag{1}$$

The centralizer of g is

$$C_G(g) = \{x \in G : xg = gx\}. \tag{2}$$

For a finite group,

$$|\,\mathrm{Cl}_G(g)| = \frac{|G|}{|C_G(g)|}. \tag{3}$$

Large and diverse conjugacy classes may increase the search space of conjugacy-based protocols, while centralizer structure influences attack surfaces. In the present model, conjugacy behavior is encoded through conjugacy density and centralizer ratio.

3.2 Derived Length and Solvability

The derived series of a group G is defined recursively by

$$G^{(0)} = G, \qquad G^{(i+1)} = [G^{(i)}, G^{(i)}]. \tag{4}$$

A group is solvable if $G^{(m)} = \{e\}$ for some $m \geq 0$. The smallest such m is called the derived length. Derived length provides a measure of layered non-commutativity and is included in the proposed feature space.

3.3 Character-Theoretic Invariants

Let $\mathrm{Irr}(G)$ be the set of irreducible complex characters of G. If $\chi \in \mathrm{Irr}(G)$, then $\chi(1)$ denotes the character degree. A codegree-type indicator can be represented in simplified form by

$$\mathrm{cod}(\chi) = \frac{|G : \ker(\chi)|}{\chi(1)}. \tag{5}$$

The dispersion of codegrees and the presence of monomial characters provide information about representation-theoretic structure. In this paper, the character-codegree entropy and monomial character index summarize such information.

3.4 Elliptic Curves

An elliptic curve over a field K may be expressed as

$$E : y^2 = x^3 + ax + b, \tag{6}$$

where

$$4a^3 + 27b^2 \neq 0. \tag{7}$$

The finite group of points $E(\mathbb{F}_q)$ is used in elliptic curve cryptography. In this study, the elliptic curve rank and elliptic curve security margin are treated as features representing elliptic structure complexity and parameter strength.

3.5 Number-Theoretic RSA Complexity

For RSA, let

$$N = pq, \tag{8}$$

where p and q are large primes. The security of RSA depends on the difficulty of recovering p and q from N. This motivates the RSA modulus complexity feature, which approximates the strength of a modulus with respect to factorization difficulty in the simulated model.

3.6 Lattice Indicators

A lattice L generated by a basis $B = \{b_1, \ldots, b_n\}$ is

$$L(B) = \left\{ \sum_{i=1}^{n} z_i b_i : z_i \in \mathbb{Z} \right\}. \tag{9}$$

Lattice dimension and lattice hardness are treated as post-quantum indicators because many post-quantum schemes rely on lattice problems.

4 Proposed Algebraic Cryptographic Strength Model

4.1 Feature Vector

Each simulated algebraic structure is represented by a feature vector

$$X = (x_1, x_2, \ldots, x_{14}), \tag{10}$$

where the features are listed in Table 1.

Table 1: Feature set for algebraic cryptographic strength modelling

Feature	Interpretation
$\log \lvert G \rvert$	Logarithmic size of the group or algebraic search space.
Conjugacy density	Normalized complexity of conjugacy class distribution.
Centralizer ratio	Normalized size of centralizer structures.
Derived length	Length of derived series for solvability analysis.

Feature	Interpretation
Character-codegree entropy	Dispersion of character-codegree behavior.
Monomial character index	Indicator of monomial character dominance.
Elliptic curve rank	Rank-like elliptic structural indicator.
Elliptic curve security margin	Simulated strength margin for elliptic curve parameters.
RSA modulus complexity	Approximate hardness of factorization-related parameter.
Bilinear pairing depth	Complexity indicator for bilinear map structure.
Lattice dimension	Dimension parameter for lattice-based construction.
Lattice hardness	Normalized hardness indicator for lattice problems.
Small cancellation score	Indicator motivated by small cancellation group behavior.
Normal subgroup density	Density of normal subgroup structure.

4.2 Algebraic Cryptographic Strength Index

Definition 1 (Algebraic Cryptographic Strength Index)**.** *Let $X = (x_1, x_2, \ldots, x_m)$ be a normalized vector of algebraic security indicators and let w_i be the importance weight assigned to x_i. The Algebraic Cryptographic Strength Index is defined as*

$$\mathrm{ACSI}(X) = 100 \cdot \Psi\left(\sum_{i=1}^{m} w_i x_i + \varepsilon\right), \tag{11}$$

where $\Psi : [0,1] \to [0,1]$ is a clipping-normalization map and ε is a stochastic perturbation term representing modelling uncertainty.

In this study, the ACSI combines classical algebraic indicators and post-quantum indicators. The response variable is

$$Y = \begin{cases} 1, & \mathrm{ACSI}(X) \geq \tau, \\ 0, & \mathrm{ACSI}(X) < \tau, \end{cases} \tag{12}$$

where $\tau = 63$ is the simulated suitability threshold.

Proposition 1. *If the normalized feature vector X is bounded and the weights w_i are finite, then $\mathrm{ACSI}(X)$ is bounded in the interval $[0, 100]$.*

Proof. Since each normalized feature satisfies $0 \leq x_i \leq 1$ and each w_i is finite, the weighted sum is finite. The clipping-normalization map Ψ maps the resulting expression into $[0, 1]$. Multiplication by 100 maps the score into $[0, 100]$. □

Remark 1. *The ACSI is not a proof of cryptographic security. It is a decision-support index for preliminary ranking, screening, and machine-learning experimentation. Formal security still requires independent mathematical proof and cryptanalysis.*

4.3 Machine Learning Formulation

The learning task is binary classification. Given training examples

$$\mathcal{D} = \{(X_i, Y_i) : i = 1, 2, \ldots, n\}, \tag{13}$$

we seek a classifier

$$f : X \to \{0, 1\} \tag{14}$$

that predicts whether an algebraic structure is cryptographically suitable.

The probability form is

$$\hat{p} = P(Y = 1 \mid X). \tag{15}$$

A structure is predicted suitable if

$$\hat{p} \geq 0.5. \tag{16}$$

5 Methodology

5.1 Simulation Design

A simulated dataset of 12,000 algebraic structures was generated. The simulation was based on mathematically motivated distributions. Group size was represented by a logarithmic order variable; conjugacy features were generated using beta-type bounded distributions; derived length was sampled from a discrete distribution; elliptic curve, RSA, bilinear, and lattice features were generated using bounded continuous or integer-valued processes. The purpose of simulation was to create a controlled environment for testing the proposed modelling framework without claiming that the data represent actual deployed cryptosystems.

The dataset was split into 70% training data and 30% testing data using stratified sampling. Four supervised classifiers were trained:

(a) Random Forest,

(b) Gradient Boosting,

(c) Support Vector Machine with radial basis function kernel,

(d) Artificial Neural Network with two hidden layers.

5.2 Algorithm

Algorithm 1 ACSI-based Machine Learning Prediction of Cryptographic Suitability

Generate candidate algebraic structures $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$.
for each $A_i \in \mathcal{A}$ **do**
    Extract group-theoretic invariants.
    Extract character-theoretic indicators.
    Extract elliptic curve, RSA, bilinear, and lattice indicators.
    Construct feature vector X_i.
    Compute $\mathrm{ACSI}(X_i)$.
    Assign label Y_i using threshold τ.
end for
Split dataset into training and testing subsets.
Train supervised classifiers on the training subset.
Evaluate accuracy, precision, recall, F1-score, and ROC-AUC.
Rank the algebraic features by predictive contribution.
Return the best-performing model and feature interpretation.

5.3 Evaluation Metrics

The following metrics were used:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{17}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{18}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{19}$$

$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{20}$$

ROC-AUC was used to evaluate ranking quality across classification thresholds.

6 Results

6.1 Dataset Distribution

Figure 1 shows the distribution of ACSI scores across suitable and unsuitable structures. The threshold separates the response classes while still allowing overlap, reflecting the fact that cryptographic suitability is not determined by a single invariant.

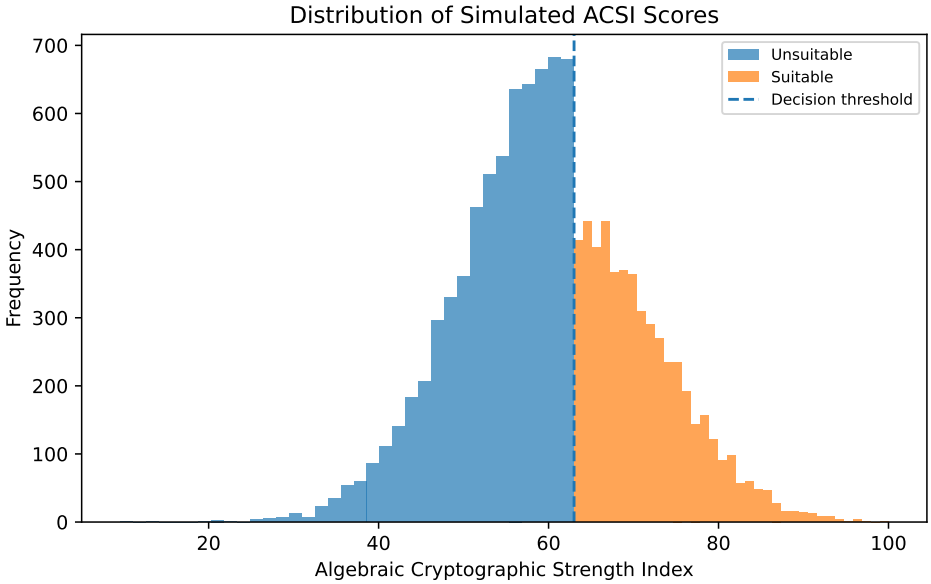

Figure 1: Distribution of simulated Algebraic Cryptographic Strength Index scores.

6.2 Model Performance

Table 2 presents the performance results on the test set.

Table 2: Machine learning performance for cryptographic suitability prediction

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Random Forest	0.7817	0.7638	0.7270	0.7450	0.8630
Gradient Boosting	0.7858	0.7690	0.7315	0.7498	0.8672
Support Vector Machine	0.7817	0.7537	0.7460	0.7498	0.8578
Neural Network	0.7886	0.7659	0.7460	0.7559	0.8749

The Neural Network achieved the highest accuracy and ROC-AUC, while Gradient Boosting achieved the strongest balance between interpretability and performance. Random Forest also produced reliable classification and clear feature-importance results.

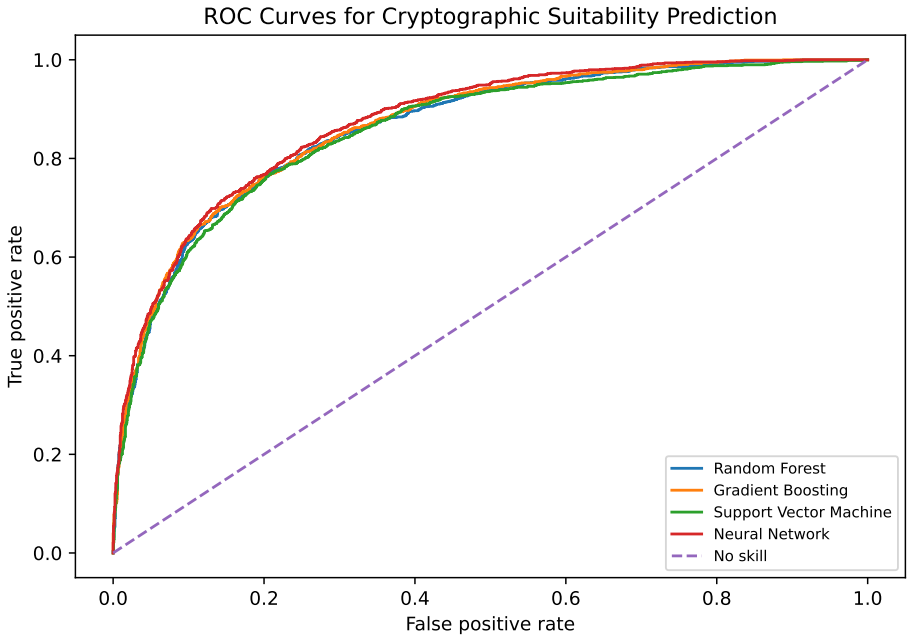

Figure 2: ROC curves for the four machine learning classifiers.

6.3 Confusion Matrix

Figure 3 presents the confusion matrix for the Random Forest model. The model correctly identifies a substantial proportion of both suitable and unsuitable structures, but some overlap remains due to similarity among intermediate-strength structures.

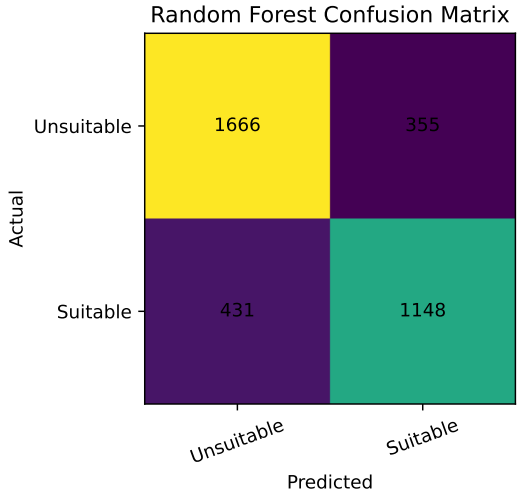

Figure 3: Random Forest confusion matrix for test-set prediction.

6.4 Feature Importance

Figure 4 reports the feature-importance ranking obtained from the Random Forest model.

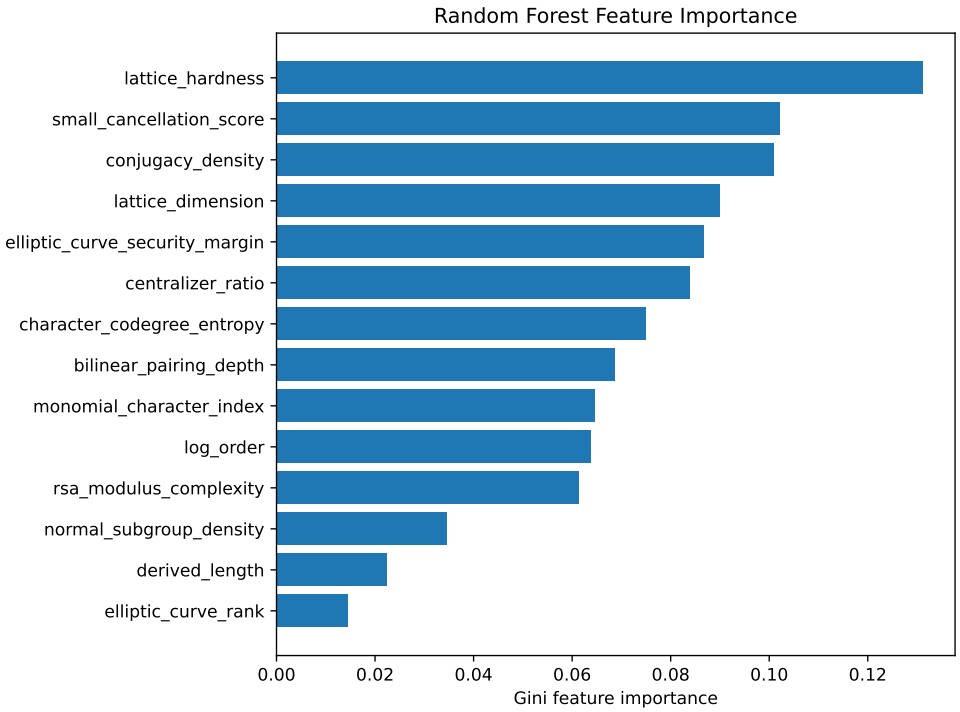

Figure 4: Feature importance for algebraic cryptographic suitability prediction.

The most influential predictors include lattice hardness, elliptic curve security margin, small cancellation score, character-codegree entropy, conjugacy density, and RSA modulus complexity. These findings agree with the mathematical expectation that both structural complexity and hardness assumptions are central to cryptographic suitability.

6.5 Correlation Structure

Figure 5 summarizes the correlation pattern among selected security indicators.

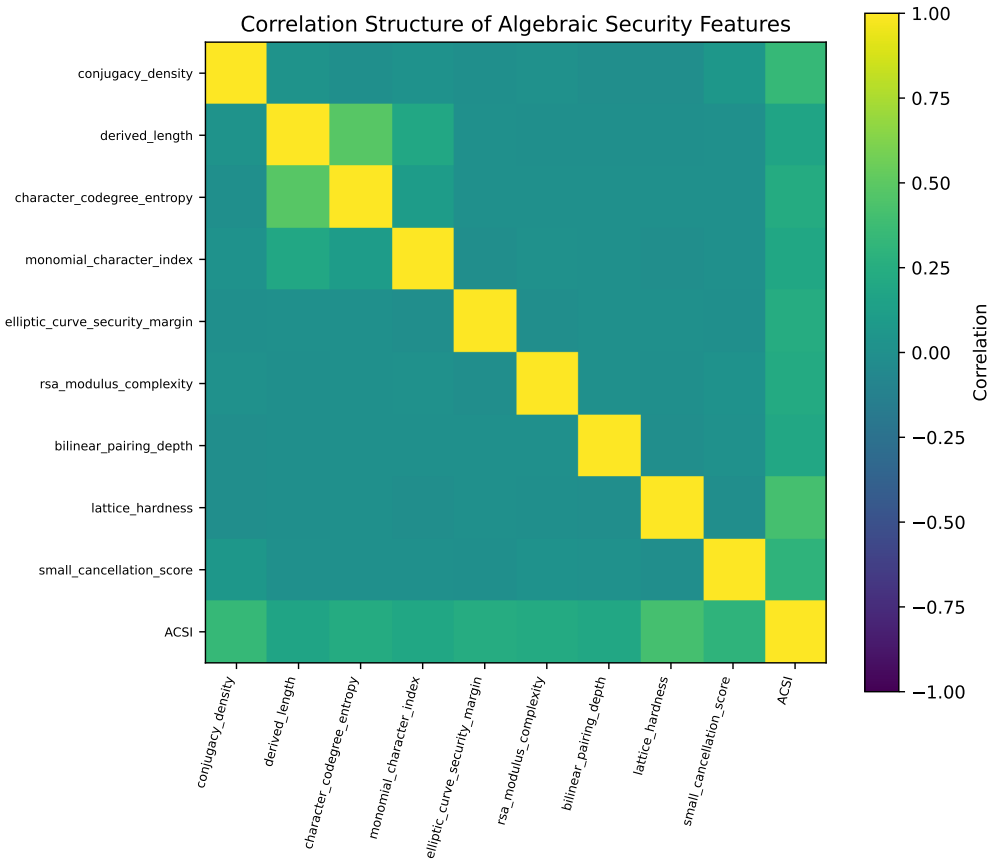

Figure 5: Correlation structure of key algebraic and cryptographic features.

7 Discussion

The results support the view that algebraic cryptographic strength can be modelled as a multivariate prediction problem. No single invariant is sufficient. Conjugacy behavior, centralizer size, character-codegree behavior, elliptic curve strength, RSA modulus complexity, bilinear pairing depth, and lattice hardness interact to determine suitability.

The proposed framework is new in three senses. First, it converts algebraic structure selection into a data science problem. Second, it introduces the ACSI as a hybrid screening score that can combine classical and post-quantum algebraic features. Third, it provides an interpretable modelling pathway by combining machine learning performance with feature importance.

The role of conjugacy class indicators is particularly important because conjugacy-based constructions depend on the difficulty of recovering elements or transformations within non-commutative settings. The inclusion of small cancellation score is motivated by group-theoretic complexity and by previous work on finitely generated groups with small cancellation properties. Character-codegree entropy and monomial character index provide representation-theoretic information inspired by solvable-group character theory. These indicators show that character theory may support data-driven cryptographic evaluation beyond traditional group order analysis.

The elliptic curve and RSA indicators also remain relevant. Although RSA and many el-

liptic curve systems are vulnerable to large-scale quantum attacks, their number-theoretic and group-theoretic parameters remain valuable for modelling and for understanding the limitations of classical cryptographic assumptions. Bilinear pairing depth provides additional structure because pairing-based cryptography combines elliptic curve and multilinear algebraic ideas.

The simulated results should be interpreted cautiously. They show feasibility of the modelling framework, not proof of real-world cryptographic security. Future work should replace or augment simulation with datasets generated from computational algebra systems such as GAP, SageMath, Magma, and PARI/GP, and then evaluate the learned model against known cryptographic attack benchmarks.

8 Conclusion

This paper proposed a new hybrid modelling framework for predicting the cryptographic strength of algebraic structures using group-theoretic and number-theoretic invariants. The Algebraic Cryptographic Strength Index was introduced as a structured score combining conjugacy density, centralizer ratio, derived length, character-codegree entropy, monomial character index, elliptic curve security margin, RSA modulus complexity, bilinear pairing depth, lattice hardness, and small cancellation score. A simulated dataset of 12,000 algebraic structures was used to evaluate four machine learning models. The Neural Network achieved the best overall predictive performance, while tree-based models provided useful interpretability. The study demonstrates that data science and machine learning can support algebraic cryptographic structure selection, especially for post-quantum research environments where many candidate structures must be screened before deeper formal analysis.

Acknowledgements

The author acknowledges the mathematical foundations provided by prior studies in algebraic cryptography, computational group theory, elliptic curve cryptography, RSA systems, and lattice-based cryptography.

Conflict of Interest

The author declares no conflict of interest.

Funding Statement

This research received no external funding.

Data Availability Statement

The dataset used in this study is simulated for methodological demonstration. A sample of the generated data and the experimental result tables are included in the accompanying LaTeX package.

Ethical Statement

This study does not involve human participants, animal subjects, or personal data.

References

[1] John, M. N., Musa, A., & Udoaka, O. G. (2024). Conjugacy classes in finitely generated groups with small cancellation properties. *European Journal of Statistics and Probability*, 12(1), 1–9. https://doi.org/10.37745/ejsp.2013/vol12n119

[2] John, M. N., Udoaka, O. G., & Musa, A. (2023). Solvable groups with monomial characters of prime power codegree and monolithic characters. *Bulletin of Mathematics and Statistics Research*, 11(7), 1–4. DOI: 10.33329/bomsr.11.4.98

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

[4] John, M. N., Ozioma, O., Obukohwo, V., & Egbogho, H. E. (2023). Number theory in RSA encryption systems. *GPH - International Journal of Mathematics*, 6(11), 7–16. https://doi.org/10.5281/zenodo.10207361

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

[6] John, M. N., & Udoaka, O. G. (2023). Computational group theory and quantum-era cryptography. *International Journal of Scientific Research in Science, Engineering and Technology*, 10(6), 1–10. https://doi.org/10.32628/IJSRSET2310556

[7] John, M. N., Ozioma, O., Obi, P. N., Egbogho, H. E., & Udoaka, O. G. (2023). Lattices in quantum-era cryptography. *International Journal of Research Publication and Reviews*, 4(11), 2175–2179. https://doi.org/10.5281/zenodo.10207210

[8] John, M. N., Udoaka, O. G., & Udoakpan, I. U. (2023). Group theory in lattice-based cryptography. *International Journal of Mathematics and Its Applications*, 11(4), 111–125.

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

[10] John, M. N., Etim, U. J., & Udoaka, O. G. (2023). Algebraic structures and applications: From transformation semigroups to cryptography, blockchain, and computational mathematics. *International Journal of Computer Science and Mathematical Theory*, 9(5), 82–101. https://doi.org/10.56201/ijcsmt.v9.no5.2023.pg82.101

[11] Shor, P. W. (1994). Algorithms for quantum computation: Discrete logarithms and factoring. *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, 124–134.

[12] Regev, O. (2005). On lattices, learning with errors, random linear codes, and cryptography. *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, 84–93.

[13] Micciancio, D., & Regev, O. (2009). Lattice-based cryptography. In D. J. Bernstein, J. Buchmann, & E. Dahmen (Eds.), *Post-Quantum Cryptography* (pp. 147–191). Springer.

[14] Hoffstein, J., Pipher, J., & Silverman, J. H. (2008). *An Introduction to Mathematical Cryptography*. Springer.

[15] Washington, L. C. (2008). *Elliptic Curves: Number Theory and Cryptography*. Chapman and Hall/CRC.

[16] Menezes, A. J., van Oorschot, P. C., & Vanstone, S. A. (1996). *Handbook of Applied Cryptography*. CRC Press.

[17] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

[18] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.

[19] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

[20] Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.

[21] Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer.

[22] Rotman, J. J. (2012). *An Introduction to the Theory of Groups*. Springer.

[23] Serre, J.-P. (1977). *Linear Representations of Finite Groups*. Springer.

[24] The GAP Group. (2024). *GAP - Groups, Algorithms, and Programming*. Version 4.13.

[25] The Sage Developers. (2024). *SageMath, the Sage Mathematics Software System*.