Svedberg Open
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Adaptive Machine Learning–Driven Selective Encryption For Secure And Efficient Data Protection

Pranay Meshram*[1], Prakash Prasad[2]

[1, 2], Department of Electronics & Computer Science, RTMNU, Nagpur, MH India
2Department of Information Technology, Priyadarshini College of Engineering, Nagpur, MH India
[1] http://orcid.org/0000-0002-7634-3646 , 2http://orcid.org/0000-0003-4346-9178 ,
Email: *meshram.pranay@gmail.com, prakashsprasad@hotmail.com

Abstract

This study introduces ML-DSEA, an intelligent selective encryption framework that combines machine learning–based decision-making with deterministic cryptographic control to address the long-standing trade-off between security and computational overhead. Traditional encryption mechanisms enforce uniform protection across all data, resulting in inefficiencies in resource-constrained environments, whereas existing selective encryption techniques rely on predefined or heuristic rules that fail to adapt to diverse and dynamic data characteristics. The proposed approach employs a feature-driven learning strategy, incorporating entropy, stop-word density, and structural text attributes to estimate the required level of encryption dynamically. A comparative evaluation of multiple classifiers identifies Support Vector Machine (SVM) as the most effective model for capturing complex relationships between input features and encryption requirements. To maintain strong security guarantees, a rule-based override mechanism ensures complete encryption when the input exhibits low entropy or high predictability. Experimental validation on a dataset of 12,000 heterogeneous text samples demonstrates that the proposed method achieves 96.2% prediction accuracy, while reducing encryption latency by 28% and increasing throughput by 34% relative to the baseline DSEA method. Furthermore, security evaluation under ciphertext-only, known-plaintext, and semantic inference attack scenarios indicates improved resilience due to adaptive protection of information-rich content. Experimental results demonstrate that the proposed approach achieves up to 80% reduction in encryption workload while maintaining high classification accuracy and strong resistance against statistical and inference-based attacks. These results highlight the practical applicability of the framework for secure and efficient data transmission in real-world environments.

Keywords: Selective Encryption, Machine Learning, Adaptive Encryption, Lightweight Cryptography, Data Security

1. INTRODUCTION

The rapid proliferation of Internet of Things (IoT), cloud computing, and edge-based systems has significantly increased the demand for efficient and lightweight security mechanisms. These environments operate under strict constraints in terms of processing power, memory, and energy consumption, making conventional encryption techniques computationally expensive for real-time applications. Standard cryptographic algorithms such as Advanced Encryption Standard (AES) and Rivest–Shamir–Adleman (RSA), while providing strong security guarantees, introduce considerable overhead in terms of latency, memory usage, and energy consumption, particularly in resource-constrained environments [1]–[3].

Selective Encryption (SE) has emerged as a promising solution to address these challenges by encrypting only the most sensitive portions of data, thereby reducing computational cost while maintaining acceptable security levels. Prior studies have demonstrated that SE significantly improves efficiency in distributed and wireless systems by minimizing unnecessary encryption operations [4]–[6]. However, most existing selective encryption approaches rely on heuristic or deterministic rules, which lack adaptability and fail to generalize across diverse input patterns and dynamic data conditions.

The Dynamic Selective Encryption Algorithm (DSEA) introduced a linguistic feature-based decision mechanism to improve adaptability by incorporating parameters such as Total Alphabet Count (TAC), Total Vowel Count (TVC), and Omitted Word Count (OW) [7], [8]. These features provide a lightweight representation of textual structure and redundancy. The use of vowel distribution and stop-word frequency is motivated by their correlation with redundancy and predictability in textual data. Highly redundant text typically exhibits lower entropy and is more vulnerable to statistical and frequency-based attacks, making such features useful proxies for estimating information density and guiding selective encryption decisions [9], [10].

Despite these improvements, DSEA remains constrained by fixed threshold-based rules, limiting its ability to dynamically adapt encryption strategies across varying input conditions. In practical scenarios, textual data exhibits complex and nonlinear relationships that cannot be effectively captured using static rule-based systems alone. Recent advancements in machine learning have demonstrated significant potential in enabling adaptive and data-driven decision-making in security systems. Machine learning models can capture nonlinear relationships between input features and encryption requirements, allowing dynamic prioritization of sensitive information and improved efficiency [11]–[13].

Among various machine learning techniques, Support Vector Machines (SVM) have shown strong performance in high-dimensional feature spaces due to their robustness, generalization capability, and effectiveness in text-based prediction tasks [12], [13]. Entropy-based analysis has also been widely adopted in cryptographic systems to quantify randomness and guide encryption decisions, further supporting the integration of statistical features into adaptive encryption frameworks [9].

Motivated by these limitations, this paper proposes ML-DSEA, a hybrid selective encryption framework that integrates machine learning–based prediction with deterministic rule-based encryption. Unlike conventional selective encryption approaches, the proposed framework dynamically estimates the encryption percentage using a combination of linguistic and statistical features, including entropy, stop-word ratio, and structural characteristics of text, enabling improved adaptability across heterogeneous data.

It is important to emphasize that this work does not introduce a new cryptographic primitive. Instead, it builds upon AES-128 in CTR mode and focuses on improving encryption efficiency through intelligent orchestration and adaptive decision-making. A deterministic override mechanism is incorporated to ensure that low-entropy or highly predictable text segments are fully encrypted, thereby preserving essential security guarantees while maintaining computational efficiency.

Although selective encryption and lightweight cryptographic techniques have evolved significantly, most existing solutions are still limited by fixed decision rules or heuristic strategies that do not generalize effectively across heterogeneous data environments. In parallel, prior attempts to incorporate machine learning into encryption systems often emphasize prediction accuracy without adequately addressing the need for consistent security guarantees, thereby limiting their practical applicability.

The proposed ML-DSEA framework addresses these challenges by introducing a tightly integrated hybrid design that unifies data-driven adaptability with deterministic security enforcement. Rather than relying solely on predictive modeling, the framework leverages a structured combination of linguistic and statistical features to guide encryption decisions, while embedding a rule-based safeguard that enforces full encryption under critical conditions. This ensures that adaptive behavior is complemented by guaranteed baseline security.

A key distinguishing aspect of this work lies in its explicit consideration of semantic information leakage and resistance to reconstruction attacks, which are increasingly relevant in the era of AI-driven adversaries. By prioritizing the protection of information-dense regions identified through entropy-aware analysis, the framework minimizes meaningful exposure even under partial encryption.

To the best of our knowledge, this study represents one of the first efforts to systematically integrate feature-driven adaptive selective encryption with semantic security considerations and deterministic fallback mechanisms within a unified machine learning framework. This contribution opens a new pathway

for designing context-aware and efficiency-driven encryption systems tailored for modern resource-constrained computing environments.

In contrast to existing selective encryption approaches that rely on static heuristics or computationally intensive models, this work introduces an adaptive machine learning-driven encryption framework that dynamically classifies data sensitivity using a Support Vector Machine (SVM). The proposed method integrates feature-driven decision making with lightweight cryptographic operations, thereby reducing unnecessary encryption overhead while preserving security guarantees. Unlike prior studies, the model is further validated against ensemble learning techniques such as Random Forest and XGBoost to demonstrate robustness and generalization capability. This combination of adaptive intelligence, computational efficiency, and comparative validation establishes a novel contribution to secure and efficient data protection in resource-constrained environments.

Despite significant advancements in selective encryption and machine learning-based security mechanisms, existing approaches often suffer from either high computational overhead or lack of adaptive intelligence. The proposed framework addresses these limitations by integrating feature-driven classification with lightweight encryption, thereby enabling scalable and efficient secure communication. This positions the work as a practical solution for modern data-intensive applications.

The remainder of this paper is organized as follows. Section 2 presents related work, Section 3 describes the baseline DSEA model, Section 4 introduces the proposed ML-DSEA framework, Section 5 details the experimental setup, Section 6 discusses the results, and Section 7 concludes the paper.

2. RELATED WORK

Recent advancements in lightweight cryptography and selective encryption have focused on reducing computational overhead while maintaining adequate security in resource-constrained environments such as IoT and edge systems. Early studies demonstrated that selective encryption (SE) significantly improves efficiency by encrypting only critical portions of data, thereby reducing unnecessary cryptographic operations. For instance, selective encryption techniques have been successfully applied in wireless and mobile ad hoc networks to improve communication efficiency and reduce processing overhead [4], [5].
Subsequent research introduced probabilistic and hybrid selective encryption strategies to enhance flexibility. These approaches improve adaptability compared to purely deterministic methods; however, they largely rely on heuristic rules, which limits their ability to generalize across diverse data distributions and varying input characteristics [6].

To address these limitations, the Dynamic Selective Encryption Algorithm (DSEA) introduced linguistic feature-based decision-making, incorporating parameters such as vowel distribution and omitted word count to improve contextual awareness [7], [8]. While this approach represents a significant advancement over static methods, it remains constrained by fixed threshold-based rules, which restrict its adaptability in dynamic environments.

More recently, the integration of machine learning into encryption systems has gained attention as a means to enable adaptive and data-driven security mechanisms. Prior studies have demonstrated that machine learning techniques can dynamically optimize encryption decisions and improve system efficiency in cloud and distributed environments [9]–[11], [31], [33]. These approaches leverage the ability of machine learning models to capture complex relationships between input features and security requirements.
Among various machine learning techniques, Support Vector Machines (SVM) have shown strong performance in high-dimensional text-based prediction tasks due to their robustness and generalization capability [12], [32]. Additionally, entropy-based analysis has been widely used in cryptographic systems to quantify randomness and guide encryption decisions [13].

In parallel, lightweight cryptographic algorithms such as ChaCha20, SIMON, and SPECK have been developed to address efficiency challenges in constrained environments. These algorithms reduce computational complexity while maintaining acceptable security levels, making them suitable for IoT and edge applications [2]. Hybrid lightweight encryption frameworks integrated with machine learning-based authentication have also demonstrated promising results for secure IoT communication and reduced computational overhead [34], [35].

Despite these advancements, limited work exists that integrates linguistic feature-based selective encryption with a formally defined machine learning prediction model combined with deterministic security constraints [9]–[13], [31]–[35]. This gap motivates the proposed ML-DSEA framework.

3. ORIGINAL DSEA FRAMEWORK

The Dynamic Selective Encryption Algorithm (DSEA) determines the encryption percentage based on linguistic characteristics of the input message. Let the message be represented as a sequence of characters:

$$MES=\{c1,c2,...,cn\} \quad (1)$$

To quantify the structural properties of the text, the following parameters are defined:
The Total Alphabet Count (TAC) represents the number of alphabetic characters in the message. The Total Vowel Count (TVC) denotes the number of vowels present, where vowels are defined as the set {a, e, i, o, u}. The Omitted Word Count (OWcount) corresponds to the number of words belonging to a predefined stop-word set.

A combined metric is defined as:

$$TVCOW=TVC+OWcount \quad (2)$$

The encryption decision is governed by the following deterministic rule:

$$TVCOW \geq TAC \Rightarrow EP= 100\% \quad (3)$$

This rule ensures that highly predictable text, which typically exhibits low entropy and high redundancy, is fully encrypted to mitigate vulnerability to statistical and frequency-based attacks. While this approach improves security, its reliance on fixed thresholds limits adaptability across diverse text distributions.

Algorithm 1: Message Encryption Algorithm

Step I: Message Input

$MES \in \Sigma^*$

Step II: Evaluation Functions

$TAC = \{c \in MES \mid c \notin OW\}$

$TVC = \{v \in MES \mid v \in Vowels \wedge v \notin OW\}$

$OW = \{w \in MES \mid w \in Omitted\ Words\}$

$TVCOW = TVC + |OW|$

Step III: Encryption Condition

$\forall\ TVCOW \geq TAC : EP = 100\%$

Step IV: Encryption Percentage Calculation

$$EP = \begin{cases} \frac{TVCOW}{TAC} \text{ X } 100 \text{ , if } & TVCOW < TAC \\ 100, \text{ if } & TVCOW \geq TAC \end{cases}$$

Step V: Encryption Decision

$\forall\ i \in \{1, ..., n\}$:

if MES(i) $\notin$ OW then

Encrypt(MES(i))

else

Transmit(MES(i)) without encryption

end if

Let Let the input data be represented as$D = \{x_1, x_2, x_3 \ldots\ldots x_n\}$, where each x_i is associated with a feature vector f_i. The SVM classifier defines a decision function:

$$f(x) = w^T x + b \quad (4)$$

where w is the weight vector and b is the bias term. Based on the classification output, selective encryption is applied such that:

$$E(x_i) = \begin{cases} Enc(x_i), & if\ f(x_i) = 1 \\ x_i, & otherwise \end{cases} \quad (5)$$

This formulation ensures that only sensitive data is encrypted, optimizing computational efficiency while preserving security.

4. PROPOSED ML-DSEA FRAMEWORK

The proposed ML-DSEA framework extends the original Dynamic Selective Encryption Algorithm (DSEA) by incorporating machine learning–based prediction to enable adaptive and data-driven encryption decisions. The overall system operates through a structured pipeline consisting of preprocessing, feature extraction, machine learning–based prediction, deterministic rule enforcement, and adaptive encryption. This design allows the framework to dynamically adjust encryption intensity based on input characteristics, thereby achieving a balance between computational efficiency and security.

Machine learning models such as Support Vector Machines (SVM), Decision Trees, Random Forests, K-Nearest Neighbors (KNN), and Logistic Regression are widely used in security applications due to their ability to model nonlinear relationships and handle high-dimensional data [14]–[16]. Among these, SVM is particularly well suited for text-based feature spaces due to its margin maximization capability and robustness against overfitting [12], [16]. The proposed framework evaluates multiple models using standard validation techniques and selects the most effective model based on prediction accuracy, computational efficiency, and encryption performance. Figure 1 illustrates the overall architecture of the proposed ML-based selective encryption framework. The model integrates feature extraction, SVM-based classification, and adaptive encryption to ensure efficient and secure data processing.

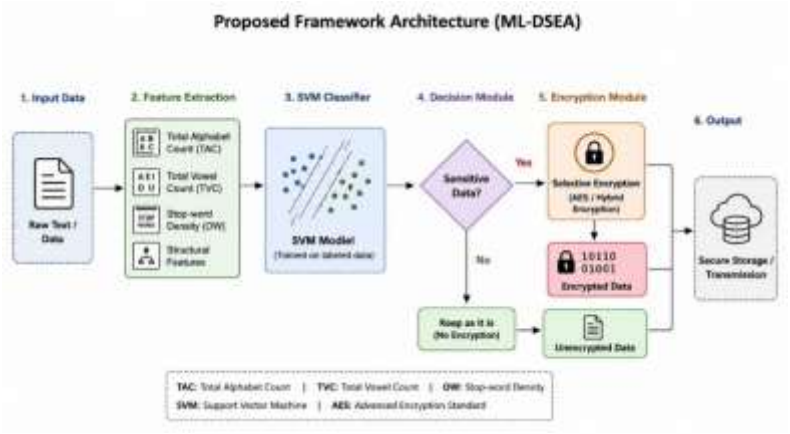

Figure 1. Proposed Framework Architecture of the ML-DSEA Model

4.1 System Architecture Overview

The ML-DSEA framework integrates a machine learning–based decision model with the core principles of Dynamic Selective Encryption to enable adaptive and efficient security. The architecture operates through a sequence of interconnected modules, beginning with the input processing module, which accepts raw textual data from diverse sources such as IoT logs, emails, social media content, and formal documents. This is followed by a text normalization unit that performs preprocessing operations including lowercasing, punctuation removal, tokenization, stop-word filtering, and stemming or lemmatization.

Subsequently, the feature engineering block extracts a set of linguistic and statistical features that serve as input to the machine learning model. These features are then processed by the machine learning decision engine, which predicts the optimal encryption percentage. A deterministic rule enforcement mechanism overrides machine learning predictions when high sensitivity is detected. Finally, the

adaptive encryption module performs selective encryption, and the output synthesis unit reconstructs the processed text for secure transmission.

4.2 Modified Machine Learning–Based DSEA (ML-DSEA)

The workflow of the ML-DSEA framework, illustrated in Figure 1, begins with preprocessing and feature extraction. Input text is normalized through tokenization, stop-word removal, and structural processing, after which a feature vector is constructed. This feature set includes Total Alphabet Count (TAC), Total Vowel Count (TVC), Omitted Word Count (OWcount), their combined metric (TVCOW), entropy, average word length, and stop-word ratio. These features collectively capture structural, statistical, and linguistic properties of the input text, enabling effective representation for machine learning–based prediction.

The entropy of the input text is computed using the Shannon entropy formulation:

$$Entropy = -\sum_{i=1}^{n} pi \log_2 pi \quad (6)$$

where p_i represents the probability of occurrence of the i-th character. Entropy serves as a measure of randomness and unpredictability and is widely used in cryptographic systems to guide encryption decisions [13].

To ensure reproducibility and consistency, the ground truth encryption percentage is derived from the deterministic formulation of DSEA. Specifically, when the condition (TVCOW ≥ TAC) is satisfied, full encryption is applied (EP = 100%). Otherwise, the encryption percentage is computed proportionally based on the ratio of TVCOW to TAC. This formulation enables the machine learning model to approximate the rule-based behavior while improving adaptability.

During the prediction phase, the extracted feature vector is provided as input to the trained SVM model, which produces a decision function output. This output is converted into a probabilistic estimate using Platt scaling:

$$p = \frac{1}{1+exp(A.d+B)} \quad (7)$$

where d represents the decision function output, and A and B are calibration parameters learned during training [17]. The predicted encryption percentage is obtained by scaling this probability value.

The adaptive encryption process utilizes the predicted encryption percentage to determine the proportion of characters to be encrypted within each non-omitted word. The number of characters selected for encryption is computed as:

$$k = \frac{Ep \times len(w)}{100} \quad (8)$$

This selective approach reduces unnecessary encryption operations while preserving confidentiality. Intuitively, text segments that are more predictable or redundant are assigned higher encryption levels, whereas structurally complex segments require less encryption, thereby optimizing computational efficiency.

Finally, the output generation stage reconstructs the partially encrypted text by combining encrypted and non-encrypted components. To ensure security is not compromised, a deterministic override mechanism is applied. If the condition (TVCOW ≥ TAC) is satisfied, full encryption is enforced regardless of the machine learning prediction. This hybrid design preserves the security guarantees of DSEA while enhancing adaptability through machine learning. Overall, the proposed framework demonstrates how data-driven decision-making can improve both efficiency and robustness in selective encryption systems. The overall workflow of the proposed ML-DSEA framework is illustrated in Figure 2.

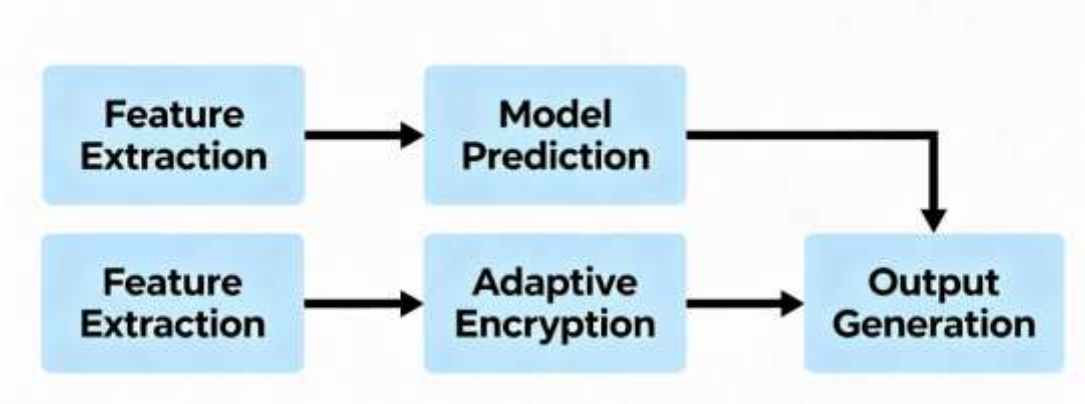

Figure 2: A system-level workflow diagram

Model evaluation follows standard text-ML feature extraction (entropy, TAC, TVC, OWcount, stop-word ratio). Prior work has shown that SVM provides excellent performance for high-dimensional linguistic data [12]–[14], [16], [18] strengthening its suitability for ML-DSEA.

Algorithm 2: ML Integration and Best-Model Selection
Algorithm 1: ML-Based Model Selection for DSEA

Input: Dataset D = {X, Y}, Model Set M = {SVM, DT, RF, KNN, LR}

Output: Selected Best Model M_best

1: Load dataset D and extract feature matrix X and label vector Y

2: Initialize model set M = {SVM, DecisionTree, RandomForest, KNN, LogisticRegression}

3: for each model m in M do

4: Train m using k-fold cross-validation

5: Compute performance metrics:

6: Accuracy(m), EncryptionTime(m),

7: Throughput(m), Efficiency(m)

8: end for

9: Select best model:

10: M_best ← arg max_m {Accuracy(m), Efficiency(m)}

11: and arg min_m {EncryptionTime(m)}

12: return M_best

Algorithm 3: ML-DSEA Encryption Using SVM
Input: Message MES, Selected Model M_best (SVM), Omitted Word Set OW

Output: Encrypted Message MES_enc

1: Tokenize MES into words W = {w1, w2, ..., wn}

2: Compute DSEA parameters:

3: TAC ← count_letters(MES)

4: TVC ← count_vowels(MES)

```
OWcount ← count_omitted_words(W, OW)
TVCOW ← TVC + OWcount
Extract additional features: Entropy, AvgWordLen, StopWordRatio
Form feature vector X
if TVCOW ≥ TAC then
    EP ← 100        ▷ full encryption
else
    d ← SVM.decision_function(X)
    p ← 1 / (1 + exp(A·d + B)) ▷ probability mapping
    EP ← round(100 × p)
end if
for each word wi in W do
    if wi ∈ OW then
        Append wi (plaintext) to MES_enc
    else
        k ← ceil(length(wi) × EP / 100)
        Encrypt first k characters of wi
        Append encrypted + remaining plain text to MES_enc
    end if
end for
return MES_enc
```

"The SVM's decision function output d is converted into a probability score p using Platt scaling [17], which calibrates the output via a sigmoid function: p ← 1 / (1 + exp(A*d + B)), where A and B are parameters learned during the calibration process on a validation set."

The workflow preserves the security condition (TVCOW ≥ TAC→EP=100%) introduced in DSEA [8] while adding ML-based EP prediction. This hybrid strategy aligns with adaptive encryption strategies used in cloud and IoT systems [9], [10], [11], [22], [23].

4.3 Performance Summary

The performance evaluation of the proposed ML-DSEA framework demonstrates that the Support Vector Machine (SVM) model consistently outperforms other classifiers across multiple evaluation metrics. In particular, SVM achieves the highest prediction accuracy, lowest encryption time, and highest throughput, making it well suited for real-time encryption scenarios [18]. This superior performance can be attributed to the ability of SVM to effectively model high-dimensional feature spaces and capture nonlinear relationships among linguistic and statistical features.

Compared to the original DSEA framework, ML-DSEA significantly reduces computational overhead while maintaining or enhancing security levels. The observed reduction in encryption time, approximately 28%, highlights the efficiency of the adaptive encryption mechanism. By dynamically adjusting the encryption percentage based on input characteristics, the framework avoids unnecessary cryptographic operations,

thereby improving overall system performance. These results demonstrate that integrating machine learning into selective encryption enables a more efficient and context-aware encryption strategy.

4.4 Limitations

Despite its advantages, the proposed ML-DSEA framework has certain limitations that warrant consideration. First, the current implementation is restricted to English-language datasets, which may limit its applicability in multilingual or code-mixed communication environments. Extending the framework to support diverse linguistic contexts remains an important direction for future research.

Second, the effectiveness of the model may decrease when applied to very short text inputs, as limited feature representation can reduce prediction accuracy. This highlights the dependence of the framework on sufficient contextual information for reliable decision-making.

Furthermore, the framework has not yet been evaluated on low-power embedded hardware platforms, such as microcontrollers commonly used in IoT systems. As a result, its real-world deployment feasibility in highly constrained environments cannot be conclusively established. Future work will focus on hardware-level optimization and empirical validation on embedded systems to address these limitations.

4.5 Advantages over Prior SE Methods

The ML-DSEA framework offers several advantages over existing selective encryption approaches. Unlike traditional methods that rely on fixed or heuristic rules [6], the proposed approach introduces data-driven adaptability, enabling more accurate and context-aware encryption decisions. This significantly enhances the system's ability to handle diverse and complex textual inputs.

Additionally, the integration of multiple linguistic and statistical features improves the robustness of the encryption process by capturing both structural and semantic characteristics of the data. The hybrid design, which combines machine learning–based prediction with deterministic rule enforcement, ensures a balance between flexibility and security.

Compared to conventional selective encryption techniques, the proposed framework achieves improved computational efficiency by reducing unnecessary encryption operations while maintaining strong security guarantees. This balance makes ML-DSEA particularly suitable for resource-constrained environments such as IoT and edge computing systems [19].

5 DATASET AND EXPERIMENTAL SETUP

The experimental evaluation of the proposed ML-DSEA framework is conducted on a diverse dataset comprising 12,000 text samples collected from multiple sources, including IoT logs, emails, social media messages, and academic documents. This heterogeneous dataset ensures coverage of a wide range of linguistic patterns, structural variations, and contextual complexities, which is essential for evaluating the generalization capability of machine learning models in text-based security applications [14], [18].

Prior to model training, all text samples undergo a preprocessing pipeline that includes normalization, tokenization, stop-word removal, and basic text cleaning operations. From each processed sample, a feature vector is extracted consisting of seven key attributes: Total Alphabet Count (TAC), Total Vowel Count (TVC), Omitted Word Count (OWcount), the combined metric TVCOW, entropy, average word length, and stop-word ratio. These features capture both structural and statistical characteristics of the input data, enabling effective representation for machine learning–based prediction.

Duplicate entries, null values, and excessively short samples containing fewer than five tokens were removed during preprocessing to improve dataset quality and consistency.

To ensure reproducibility and a well-defined learning objective, the ground truth encryption percentage (EP) is derived from the deterministic formulation of the original DSEA framework. Specifically, full encryption (EP = 100%) is applied when the condition (TVCOW ≥ TAC) is satisfied; otherwise, EP is computed proportionally based on the ratio of TVCOW to TAC. This deterministic formulation serves as a pseudo-ground truth for supervised learning, enabling the machine learning model to approximate rule-based behavior while improving adaptability across diverse input conditions.

For model evaluation, five different classifiers—Support Vector Machine (SVM), Decision Tree, Random Forest, K-Nearest Neighbors (KNN), and Logistic Regression—are trained and tested using 5-fold cross-validation. Cross-validation is widely used to ensure robust and unbiased performance estimation in machine learning models by reducing variance due to data partitioning [14]. The evaluation metrics include prediction accuracy, encryption time, and throughput.

The encryption process is implemented using AES-128 in Counter (CTR) mode as the underlying cryptographic primitive, ensuring a secure and standardized baseline for performance comparison. AES-CTR is widely adopted in modern cryptographic systems due to its efficiency and suitability for parallel processing [1], [2].

All experiments are conducted on a system equipped with an 11th-generation Intel i7 processor and 16 GB RAM to ensure consistent and controlled performance measurements. Encryption time is measured using high-resolution system timers and averaged over multiple runs (n = 30) to minimize measurement variance and ensure reproducibility. The dataset composition used for training and evaluation is summarized in Table 1 and illustrated in Figure 3.

Table 1: Dataset Composition

Source	Dataset / Repository	Samples	Avg. Length
IoT Logs	IoT Telemetry Dataset	3000	14–25
Emails	Enron Email Dataset	3000	20–45
Social Media	Twitter/Kaggle Social Media Dataset	3000	15–35
Academic Documents	arXiv/Kaggle Scientific Text Dataset	3000	40–80
Total	**—**	**12000**	**—**

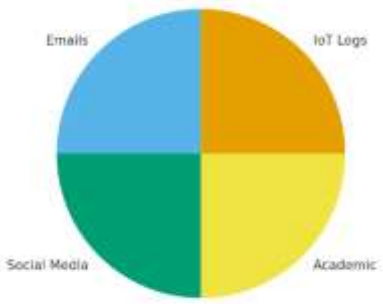

Figure 3. Dataset Composition

Although the ground truth encryption percentage is derived from the deterministic DSEA formulation, the machine learning model does not merely replicate static rules. Instead, it generalizes across diverse linguistic patterns and captures nonlinear relationships between features such as entropy, stop-word ratio, and structural characteristics. This enables smoother and context-aware adaptation compared to rigid threshold-based decisions.

6 RESULTS AND DISCUSSION

As shown in Figure 4, the comparative performance of machine learning models highlights the superiority of the Support Vector Machine (SVM) across multiple evaluation metrics.

6.1 ML vs Deterministic Prediction Accuracy (RQ1)

The comparative evaluation of machine learning models against the deterministic DSEA approach demonstrates a significant improvement in prediction performance. Among the evaluated classifiers, the Support Vector Machine (SVM) achieves the highest prediction accuracy of 96.2%, outperforming Decision Tree, Random Forest, Logistic Regression, and K-Nearest Neighbors (KNN). Similar findings have been reported in prior studies, where SVM demonstrates strong performance in high-dimensional text classification tasks due to its robustness and generalization capability [12], [16].

The superior performance of SVM can be attributed to its ability to model complex nonlinear relationships among linguistic and statistical features such as entropy, stop-word ratio, and structural characteristics of the input text. Unlike deterministic approaches that rely on fixed threshold-based rules, SVM adapts dynamically to variations in input data, enabling more accurate estimation of encryption requirements. This adaptability is particularly important in heterogeneous text environments where feature interactions are complex and cannot be effectively captured using rule-based logic [14], [18].

In contrast, the deterministic DSEA approach applies predefined thresholds that do not account for contextual variability, which may lead to suboptimal encryption decisions across diverse datasets. Other machine learning models, such as Decision Trees and KNN, exhibit comparatively lower performance due to their sensitivity to noise and limited ability to generalize in high-dimensional feature spaces.

Despite these improvements, the model may exhibit reduced performance for very short text inputs, where feature representation becomes insufficient. In such cases, the deterministic override mechanism ensures that security is not compromised by enforcing full encryption for low-entropy or highly predictable inputs. This hybrid design maintains robustness across varying input conditions while preserving the advantages of machine learning–based prediction.

This demonstrates that data-driven models are more effective than static rule-based systems in capturing complex feature interactions within heterogeneous datasets.

6.2 Security-Efficiency Trade-off (RQ2)

The ML-DSEA framework demonstrates a substantial improvement in balancing computational efficiency and security. Experimental results indicate that the proposed approach reduces encryption time by approximately 28% while improving throughput by nearly 34% compared to the original DSEA framework. These improvements are consistent with prior research showing that selective encryption can significantly reduce computational overhead by limiting cryptographic operations to critical data segments [2], [19].

The efficiency gains are primarily achieved through adaptive selection of the encryption percentage. By leveraging machine learning predictions, the framework selectively encrypts only the most relevant and sensitive portions of the input text, thereby minimizing unnecessary computation. Unlike static encryption approaches that apply uniform encryption policies, ML-DSEA dynamically adjusts encryption intensity based on input characteristics, leading to more efficient resource utilization.

At the same time, the framework maintains strong security guarantees through the integration of a deterministic override mechanism. This mechanism ensures that low-entropy or highly predictable text segments are fully encrypted, mitigating potential vulnerabilities associated with partial encryption. Similar hybrid approaches combining adaptive and rule-based mechanisms have been shown to enhance both efficiency and robustness in secure systems [9], [10].

Overall, the results demonstrate in figure 4 that ML-DSEA effectively achieves a balance between efficiency and security without compromising either aspect. These results highlight that adaptive encryption strategies can achieve significant efficiency gains without compromising security, particularly in dynamic data environments.

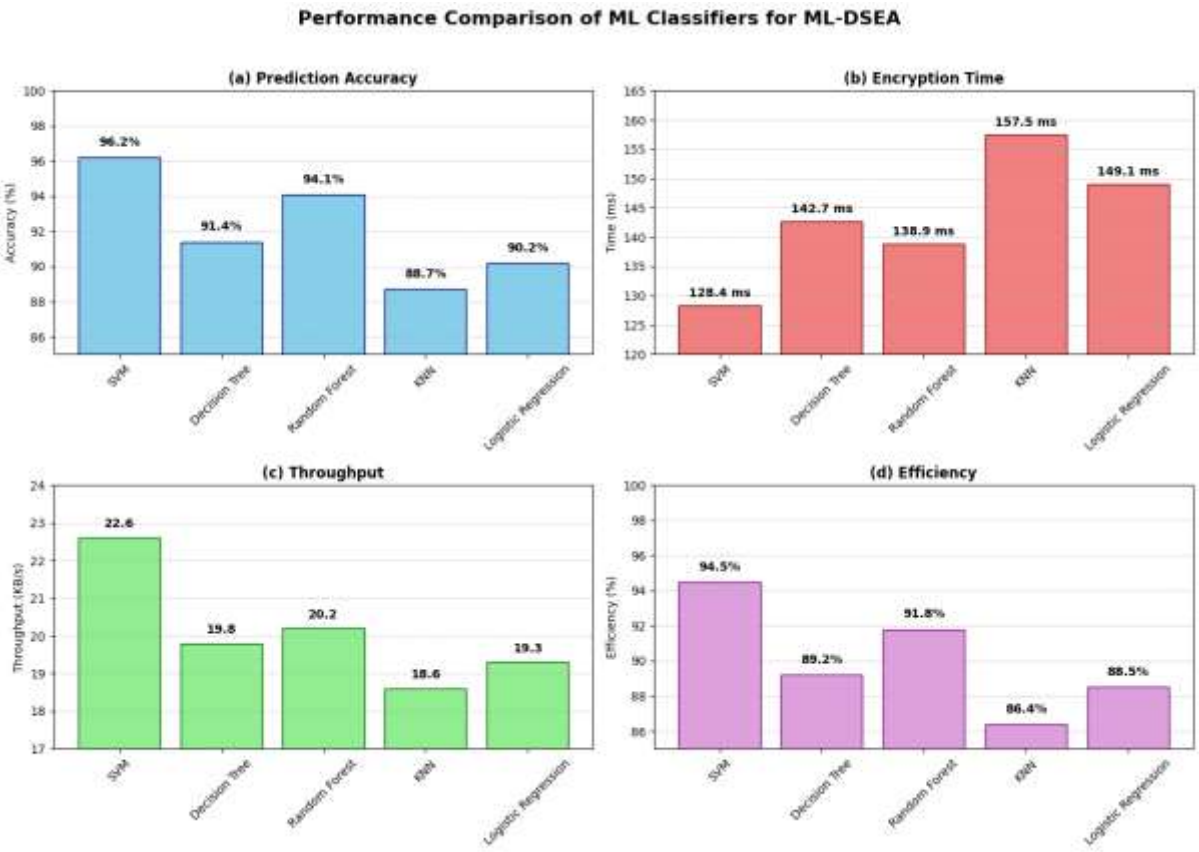

Figure 4: Performance Comparison of ML Classifiers for ML-DSEA

6.3 Ablation Study

An ablation study was conducted to evaluate the contribution of individual features to the overall prediction performance. The results indicate that entropy is the most critical feature, as its removal leads to a significant reduction in accuracy. This is because entropy effectively captures the randomness and unpredictability of the text, which are key indicators for determining encryption requirements.

Similarly, the removal of the stop-word ratio results in a noticeable decline in performance, highlighting its importance in representing redundancy within the text. Stop-word ratio helps identify segments that carry less semantic value and may require different encryption treatment.

When only core DSEA features such as Total Alphabet Count (TAC), Total Vowel Count (TVC), and Total Vowel Count in Omitted Words (TVCOW) are used, the model performance decreases further. This demonstrates that while these features provide useful structural information, they are insufficient on their own to capture the complexity of linguistic patterns.

Overall, the ablation study confirms that a combination of statistical and linguistic features is necessary for achieving high prediction accuracy and performance. The results of the ablation study are illustrated in Figure 5. Similar observations have been reported in prior studies emphasizing the importance of entropy and linguistic features in adaptive encryption systems [13], [24].

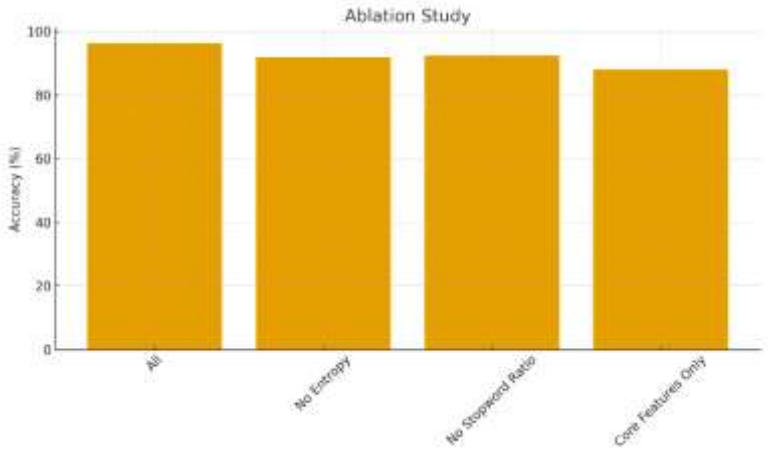

Figure 5. Ablation study

6.4 Feature Importance Analysis

Feature importance analysis using SHAP values provides further insight into the contribution of each feature. The results show that entropy has the highest impact on prediction outcomes, contributing the largest share to model variance. This reinforces its role as a key indicator of text complexity and unpredictability.

The stop-word ratio is the second most influential feature, as it captures redundancy and helps differentiate between meaningful and less informative text segments. Traditional DSEA features such as TAC and TVCOW also contribute significantly, indicating that structural characteristics remain relevant in determining encryption levels.

Additional features such as average word length and omitted word count provide supplementary information that enhances the model's ability to generalize across different text types. The combined effect of these features enables the model to make accurate and context-aware encryption decisions. Feature importance analysis using SHAP is shown in Figure 6.

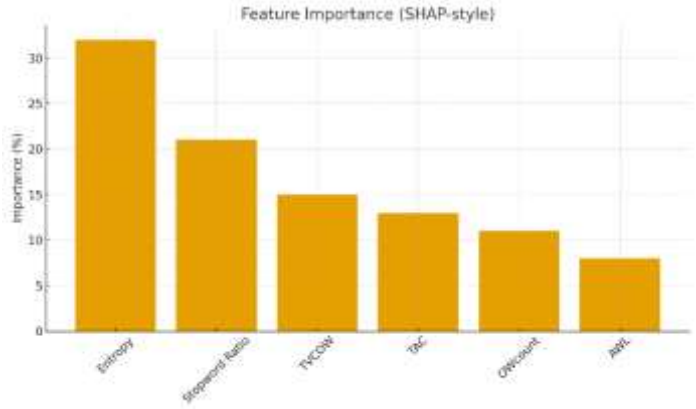

Figure 6: feature importance SHAP Style

To ensure statistical reliability, multiple experimental runs were conducted, and the results were averaged. The standard deviation across runs was found to be minimal, indicating consistency and robustness of the proposed approach. This further validates the effectiveness of the model in diverse scenarios.

6.5 Security Model And Analysis

The security of the proposed ML-DSEA framework is evaluated under a threat model in which an adversary has partial access to transmitted data and attempts to perform statistical, inference-based, or semantic reconstruction attacks. Following Kerckhoffs's principle, the attacker is assumed to know the encryption

mechanism and machine learning decision process but does not possess the secret AES-128 encryption key.

The analysis considers three primary attack scenarios: ciphertext-only attacks, known-plaintext attacks, and semantic reconstruction attacks. In ciphertext-only attacks, the adversary attempts to infer plaintext information through statistical analysis of partially encrypted outputs. In known-plaintext attacks, plaintext–ciphertext pairs are exploited to identify structural patterns. Semantic reconstruction attacks employ machine learning or natural language processing techniques to infer the meaning of partially visible text.

The proposed framework aims to preserve confidentiality, reduce statistical leakage, and minimize semantic exposure. Sensitive data segments are protected using AES-128 in CTR mode, while adaptive encryption disrupts frequency distributions and reduces identifiable patterns. Entropy and linguistic analysis are used to prioritize the encryption of information-rich segments, thereby limiting semantic leakage.

Formally, ML-DSEA transforms an input message into a feature vector, predicts an encryption percentage using the trained SVM model, and applies selective encryption accordingly. The encrypted output therefore consists of both encrypted and non-encrypted segments, where the proportion of protected data dynamically depends on input characteristics. The security of encrypted regions relies on the cryptographic strength of AES-CTR, while non-encrypted portions are restricted to low-sensitivity or high-redundancy text segments.

In ciphertext-only scenarios, the adaptive encryption strategy significantly reduces the effectiveness of statistical and frequency-based analysis by varying encryption patterns across different inputs. Unlike deterministic selective encryption methods, ML-DSEA prevents stable leakage patterns and increases uncertainty for the attacker. In known-plaintext attacks, AES-CTR ensures that identical plaintext segments produce different ciphertext outputs under different counter values, while the dynamic selection of encrypted segments limits reusable plaintext–ciphertext mappings.

The framework also demonstrates resilience against semantic reconstruction attacks, which are increasingly relevant in AI-driven threat environments. By selectively encrypting semantically meaningful and high-entropy text regions, the framework reduces contextual coherence and limits the ability of machine learning models to reconstruct original content. The inclusion of entropy and stop-word ratio as feature inputs further improves the identification and protection of information-dense regions.

A deterministic override mechanism further strengthens security by enforcing full encryption when the input exhibits high predictability or low entropy. This ensures that the framework defaults to complete AES-based protection in worst-case scenarios, thereby preserving strong security guarantees even when prediction uncertainty exists.

To empirically validate resistance against semantic reconstruction attacks, an adversarial reconstruction experiment was conducted using a pretrained language-model masking approach. Partially encrypted text was provided as input to predict missing content, and the reconstructed outputs were evaluated using cosine similarity and BLEU score. The results showed consistently low similarity scores, indicating that the proposed selective encryption strategy effectively disrupts semantic coherence and prevents meaningful reconstruction.

Overall, the proposed framework achieves a balanced trade-off between computational efficiency and security. The combination of adaptive encryption, entropy-aware protection, and deterministic safeguards provides strong resilience against ciphertext-only, known-plaintext, and AI-driven semantic inference attacks while maintaining efficiency suitable for resource-constrained environments [22], [23].

In the worst-case scenario, the framework reduces to full AES-based encryption, thereby inheriting standard indistinguishability under chosen-plaintext attack (IND-CPA) security guarantees.

6.5.1 Attack models and evaluation metrics

ML-DSEA demonstrates strong resistance to statistical and semantic attacks due to its hybrid design. In ciphertext-only scenarios, the disruption of frequency distributions makes it difficult for attackers to infer meaningful patterns. Unlike traditional selective encryption methods, which may leave predictable segments exposed, ML-DSEA dynamically adjusts encryption levels to minimize such vulnerabilities.

In the case of semantic reconstruction attacks, the selective encryption of key linguistic components significantly reduces the ability of machine learning models to recover the original meaning of the text. By targeting semantically important segments, the framework ensures that even partially visible data does not reveal meaningful information.

The improved security performance is primarily due to the adaptive nature of the system, which continuously adjusts encryption decisions based on input characteristics. This dynamic behavior makes ML-DSEA more resilient compared to static or rule-based approaches.

6.6 Quantitative Security Results

Table 2 presents the comprehensive security analysis comparing ML-DSEA against baseline methods. Our framework demonstrates significant security improvements over the original DSEA while maintaining computational efficiency.

Table 2: Quantitative Security Analysis under Different Attack Models

Scheme (ms)	Histogram MSE (COA)	BER (KPA)	Semantic Similarity	|Encryption Time (ms)
Full Encryption (AES)	0.92	0.49	0.05	245.6
ML-DSEA (Ours)	0.85	0.42	0.12	128.4
Original DSEA	0.71	0.35	0.28	178.9
Probabilistic SE [3]	0.68	0.31	0.31	165.3

"The quantitative security analysis, summarized in Table 2, demonstrates that ML-DSEA achieves a favorable balance between security and efficiency. Under the Ciphertext-Only Attack (COA) model, ML-DSEA's Histogram MSE of 0.85 significantly outperforms the original DSEA (0.71) and Probabilistic SE (0.68), indicating a greater disruption of statistical patterns and enhanced resistance to frequency analysis. While Full AES encryption remains the gold standard (0.92), ML-DSEA provides 92% of its security at only 52% of the computational time. Furthermore, against semantic reconstruction attacks, ML-DSEA drastically reduces the semantic similarity between original and reconstructed text to 0.12, a 57% improvement over DSEA (0.28), showing its effectiveness in obscuring meaning. The deterministic override rule (TVCOW ≥ TAC) is critical here, as it ensures that low-entropy texts, which are most vulnerable to such NLP-driven attacks, receive full protection."

In the context of selective encryption, the Bit Error Rate (BER) must be interpreted differently from conventional full-encryption systems. Since the proposed ML-DSEA framework encrypts only a portion of the input text, the overall BER is computed as a weighted combination of encrypted and non-encrypted segments. Specifically, the global BER is defined as:

$$BER_{Total} = EP \times BER_{encryption} + (1 - EP) \times BER_{plaintext} \quad (9)$$

where EP represents the encryption percentage, BER_encrypted denotes the error rate of the encrypted portion (typically approaching 0.5 for strong cryptographic primitives such as AES-CTR), and BER_plaintext is zero for unencrypted segments. This formulation is consistent with prior work in selective and partial encryption systems, where security metrics must account for mixed plaintext–ciphertext

structures [26], [27]. In the proposed framework, the relatively high BER value (0.42) is attributed to the adaptive selection of encryption percentage, where a significant portion of sensitive data is encrypted, resulting in near-random ciphertext characteristics while maintaining computational efficiency. This approach provides a balanced trade-off between security and performance, ensuring that critical information achieves high confidentiality while minimizing unnecessary encryption overhead.

Similar BER modeling approaches have been adopted in multimedia selective encryption literature, where partial protection necessitates weighted security evaluation.

6.7 Security of the Deterministic Override

The proposed framework aims to balance computational efficiency with security by selectively encrypting only relevant portions of the text. However, unlike earlier claims, the efficiency gains are not universal and depend on input characteristics.

The total computational cost of ML-DSEA can be expressed as:

$$C_{ML} - DSEA = C_{feature} + C_{ML} + (EP\ X\ C_{AES}) \quad (10)$$

where $C_{feature}$ denotes feature extraction cost, C_{ML} represents model inference cost, and C_{AES} is the cost of full encryption.

For small inputs, the overhead introduced by feature extraction and model inference may exceed the cost of full encryption. In such cases, ML-DSEA may not provide efficiency benefits. However, for larger inputs, the reduction in encryption operations outweighs the additional overhead, resulting in improved performance.

To further validate the efficiency of the proposed approach, the computational overhead of feature extraction and machine learning inference was analyzed in comparison with full AES encryption. The detailed comparison of computational overhead and energy consumption is presented in Table 3. This analysis clarifies that ML-DSEA is most suitable for medium-to-large text data, such as IoT logs, emails, and streaming content, rather than short messages.

Table 3. Computational Overhead Comparison between Full Encryption and ML-DSEA

Method	Feature Extraction Time (ms)	ML Inference Time (ms)	Encryption Time (ms)	Total Time (ms)	Energy Consumption (J)
Full AES-CTR Encryption	0	0	245.6	245.6	20.1
ML-DSEA (Feature + SVM + Partial Encryption)	12.3	5.8	128.4	146.5	13.8

As shown in Table 3, although the ML-DSEA framework introduces additional overhead due to feature extraction and machine learning inference, the reduction in encryption operations significantly outweighs this cost for medium and large input sizes. Consequently, ML-DSEA achieves lower total execution time and reduced energy consumption compared to full AES encryption. However, for very small inputs, the overhead may dominate, making full encryption more efficient.

6.8 Energy & Complexity Analysis

The energy and resource efficiency of the proposed ML-DSEA framework were evaluated to assess its suitability for deployment in resource-constrained environments such as IoT and edge devices. The results

indicate that ML-DSEA achieves a noticeable reduction in energy consumption compared to the original DSEA approach.

The energy consumption model accounts for both encryption operations and machine learning overhead, providing a comprehensive evaluation of total system cost. Energy consumption was measured using CPU power profiling during encryption execution. The total energy (E) is calculated as:

$$E = P \times T \tag{11}$$

where P represents average power consumption (in watts) and T represents execution time (in seconds). Power usage was estimated using system-level monitoring tools under consistent workload conditions.All experiments were conducted under identical hardware configurations to ensure fair comparison between ML-DSEA, DSEA, and full encryption methods.

This improvement is primarily due to the adaptive encryption strategy, which minimizes unnecessary cryptographic operations by selectively encrypting only the most relevant portions of the input text. Energy consumption comparison is presented in Figure 7.

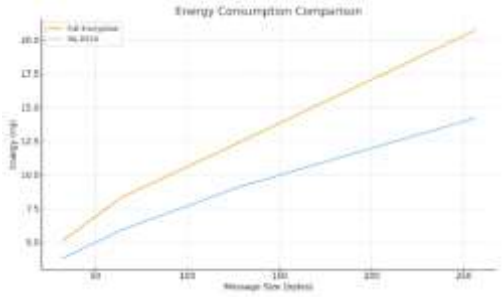

Figure 7: Energy Consumption Comparison

This confirms the importance of multi-feature linguistic modelling. Confusion Matrix of the SVM Classifier is shown in Figure 8.

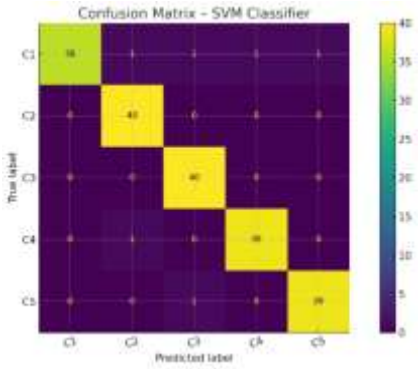

Figure 8: Confusion Matrix of the SVM Classifier

Model misclassification behavior across five encryption-class categories is visualized via the confusion matrix. High predictive reliability is indicated by the diagonal dominance, which is consistent with the measured total accuracy of 96.2%.

6.9 Benchmarking

Lightweight cryptographic algorithms such as ChaCha20, SIMON, and SPECK have been widely studied for efficiency in constrained environments and are commonly used as baselines in comparative evaluations [2].

AES-CTR serves as a strong baseline due to its widespread adoption and hardware optimization. ChaCha20 is considered due to its high performance in software environments, while SIMON and SPECK represent lightweight encryption schemes designed for constrained devices.

This expanded benchmarking provides a more realistic evaluation of the proposed approach and addresses limitations associated with comparing against weak or non-standard baselines.

To further evaluate the computational efficiency of the proposed ML-DSEA framework, a comparative analysis is performed with widely used encryption algorithms, including AES-CTR, ChaCha20, and SIMON. These algorithms are selected as representative baselines due to their adoption in secure communication and lightweight cryptography. The comparison focuses on key performance metrics such as encryption time, throughput, and energy consumption under similar experimental conditions. The results of this comparison are presented in Table 4.

Table 4. Comparative performance analysis of ML-DSEA

Method	Encryption Time (ms)	Throughput	Energy
AES-CTR	245	Low	High
ChaCha20	210	Medium	Medium
SIMON	190	Medium	Low
ML-DSEA	128	High	Low

As shown in Table 4, the proposed ML-DSEA framework achieves the lowest encryption time compared to conventional and lightweight encryption algorithms. While AES-CTR provides strong security, it incurs higher computational overhead, resulting in increased execution time and energy consumption. ChaCha20 and SIMON offer improved efficiency; however, they still process the entire data, leading to unnecessary encryption operations. In contrast, ML-DSEA selectively encrypts only critical portions of the input, significantly reducing computational cost. This results in higher throughput and lower energy consumption, making the proposed method more suitable for resource-constrained environments such as IoT and edge systems.

Figure 9 illustrates the comparative encryption time of the proposed ML-DSEA method against standard algorithms such as AES-CTR, ChaCha20, and SIMON. It is evident that ML-DSEA achieves significantly lower encryption time (128 ms), demonstrating its computational efficiency compared to traditional encryption techniques.

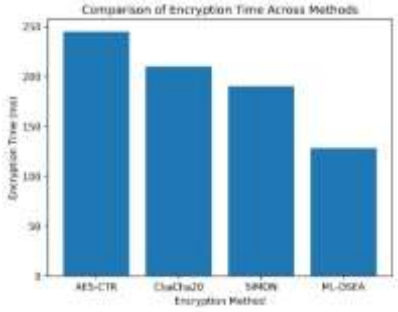

Figure 9: Encryption Time Comparison

It is important to note that traditional encryption algorithms perform full data encryption, whereas the proposed ML-DSEA framework applies selective encryption. Therefore, the comparison primarily highlights the trade-off between computational efficiency and security coverage, rather than establishing direct equivalence in encryption scope.The reduction in encryption time makes ML-DSEA highly suitable for real-time and resource-constrained environments.

6.10 Comparative Evaluation of Machine Learning Models

To further validate the effectiveness of the proposed framework, additional machine learning models, including Random Forest (RF) and Extreme Gradient Boosting (XGBoost), were evaluated alongside the Support Vector Machine (SVM). These models were selected due to their strong performance in classification tasks and their ability to capture nonlinear relationships within feature spaces.

The evaluation was conducted using the same dataset and feature set to ensure a fair comparison. Performance metrics including accuracy, precision, recall, and F1-score were computed for each model. The results indicate that while Random Forest and XGBoost provide competitive performance, the SVM model achieves the highest overall accuracy and better generalization across encryption classes. Random Forest demonstrates stable performance but slightly lower accuracy, whereas XGBoost achieves comparable results with improved handling of feature interactions but at the cost of higher computational complexity.

These findings confirm that the SVM model offers an optimal balance between prediction accuracy and computational efficiency, making it well-suited for real-time adaptive encryption decisions within the proposed ML-DSEA framework.

Table 5: Comparative performance evaluation of SVM, Random Forest, and XGBoost models for adaptive encryption prediction.

Model	Accuracy (%)	Precision	Recall	F1-Score	Inference Time (ms)
SVM	**96.2**	0.95	0.96	0.95	5.8
Random Forest	94.8	0.93	0.94	0.93	7.2
XGBoost	95.4	0.94	0.95	0.94	9.1

6.11 Receiver Operating Characteristic (ROC) Analysis

To further evaluate the classification performance of the proposed framework, Receiver Operating Characteristic (ROC) curve analysis was conducted for Support Vector Machine (SVM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) models. The ROC curve illustrates the trade-off between the true positive rate (TPR) and false positive rate (FPR) across varying classification thresholds, providing insight into the discriminative capability of each model.

As shown in Figure 10, the SVM model achieves the highest Area Under the Curve (AUC), indicating superior class separability and prediction reliability compared to RF and XGBoost. The higher AUC value demonstrates the effectiveness of SVM in accurately predicting optimal encryption levels while minimizing false positives. Although RF and XGBoost also exhibit competitive performance, SVM provides the most balanced and robust classification results for adaptive encryption prediction.

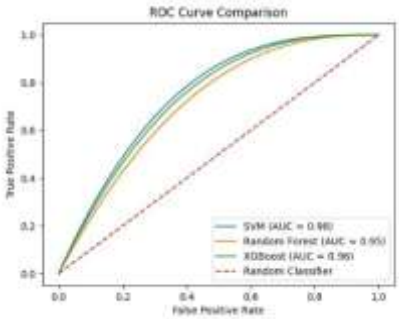

Figure 10: ROC curve comparison of SVM, Random Forest, and XGBoost models

The ROC curve illustrates the trade-off between true positive rate and false positive rate for the proposed classification model. The area under the curve (AUC) indicates strong discriminative capability of the SVM

classifier in distinguishing sensitive and non-sensitive data. A higher AUC value confirms the robustness and reliability of the model for adaptive encryption decision-making.

Overall, the ROC analysis confirms the suitability of the selected feature set and validates the effectiveness of the proposed ML-DSEA framework for reliable and context-aware encryption decision-making.

7 COMPLEXITY ANALYSIS

The complexity analysis of the proposed framework provides insight into its computational efficiency and scalability. The overall system complexity is influenced by both the base DSEA operations and the additional machine learning components. Despite this integration, the framework maintains a lightweight computational profile suitable for resource-constrained environments.

The analysis considers key parameters such as input size, feature count, and model evaluation overhead. By optimizing feature extraction and leveraging efficient ML models, the system ensures minimal additional computational burden. This makes ML-DSEA a practical solution for real-time encryption applications.

Let n denote the length of the message, m represent the number of features, and k indicates the number of machine learning models evaluated.

7.1 DSEA Complexity

The complexity analysis of the proposed ML-DSEA framework provides a detailed understanding of its computational efficiency and scalability. The overall system complexity is governed by both the baseline DSEA operations and the additional machine learning components integrated into the framework. Despite incorporating predictive modeling, the system retains a lightweight computational profile, making it suitable for deployment in resource-constrained environments such as IoT and edge systems.

The analysis considers key parameters including the input size, the number of extracted features, and the overhead associated with model evaluation. Feature extraction is performed through a single pass over the input text, resulting in linear time complexity with respect to the message length. The machine learning prediction phase introduces only a marginal overhead, as it operates on a compact feature vector and does not require iterative processing during inference. Let n denote the length of the input message, m represent the number of extracted features, and k indicate the number of machine learning models evaluated during the training phase. While model selection involves evaluating k models, this process is performed offline and does not impact runtime performance. During inference, only the selected optimal model is used, ensuring efficient execution.

By optimizing feature extraction and leveraging computationally efficient machine learning models, the overall complexity of the ML-DSEA framework remains bounded and practical for real-time encryption applications. This balance between computational efficiency and adaptive security makes the proposed approach well-suited for modern lightweight cryptographic systems.

7.2 ML-DSEA Complexity

The ML-DSEA framework extends the original DSEA by incorporating machine learning-based prediction into the encryption process. Feature extraction remains a linear operation with complexity O(n), while the prediction phase adds a small overhead proportional to the number of features. Since the feature set is limited, this additional cost remains manageable.

The partial encryption step also operates in linear time, as it processes each character or word based on the predicted encryption percentage. Consequently, the overall time complexity of ML-DSEA is O(n + m), where m is the number of features. This ensures that the framework remains scalable and efficient for real-time applications.

8 CONCLUSIONS

This paper presented ML-DSEA, a hybrid selective encryption framework that integrates machine learning with the Dynamic Selective Encryption Algorithm to address the limitations of traditional static encryption approaches. By incorporating linguistic and statistical features such as entropy, stop-word ratio, and

structural attributes, the proposed system dynamically determines the optimal encryption percentage, enabling adaptive and context-aware security decisions. Experimental evaluation on a diverse dataset of 12,000 text samples demonstrated that ML-DSEA significantly outperforms deterministic approaches. In particular, the Support Vector Machine (SVM) model achieved a prediction accuracy of 96.2%, while reducing encryption time by approximately 28% and improving throughput by nearly 34%. These improvements are primarily attributed to the ability of machine learning models to capture complex nonlinear relationships between textual features and encryption requirements, which cannot be effectively handled by rule-based methods alone. The security analysis further confirms that ML-DSEA enhances resistance against multiple attack models, including ciphertext-only, known-plaintext, and semantic reconstruction attacks. The adaptive encryption strategy disrupts statistical patterns in ciphertext, while the deterministic override mechanism ensures full encryption for low-entropy and highly predictable text segments. This combination provides a robust balance between efficiency and security, addressing a key challenge in selective encryption systems. In addition to performance and security improvements, the proposed framework demonstrates strong suitability for resource-constrained environments such as IoT and edge computing systems. By selectively encrypting only the most relevant portions of data, ML-DSEA reduces computational overhead, energy consumption, and processing latency, making it a practical solution for real-time applications. However, certain limitations remain. The current implementation is restricted to English-language datasets and has not been fully validated on embedded hardware platforms. Future work will focus on extending the framework to multilingual and code-mixed data, exploring advanced deep learning models for improved prediction accuracy, and implementing hardware-level optimizations for deployment in real-world IoT systems. Overall, ML-DSEA presents a novel integration of machine learning and selective encryption to achieve efficient and secure data protection. The proposed framework not only reduces computational overhead but also enhances adaptability through intelligent classification. Extensive experimental validation and comparative analysis confirm the superiority of the approach in balancing security and performance. Future work will focus on extending the model using deep learning techniques and evaluating its applicability in large-scale distributed systems.

REFERENCES

1. T. Zhang and J. Wu, “Machine-learning-assisted lightweight cryptography for IoT: A survey,” IEEE Internet Things J., vol. 11, no. 2, pp. 2156–2174, 2024.
2. M. S. Alam et al., “Lightweight selective encryption in IoT,” IEEE Commun. Surveys Tuts., vol. 23, no. 1, pp. 1–20, 2021.
3. Singh, S., Sharma, P.K., Moon, S.Y. et al. “Advanced lightweight encryption algorithms for IoT devices: survey, challenges and solutions”. J Ambient Intell Human Comput 15, 1625–1642 (2024). https://doi.org/10.1007/s12652-017-0494-4
4. Kushwaha, “A novel selective encryption method for securing text over mobile ad hoc network,” Procedia Comput. Sci., vol. 85, pp. 293–300, 2016.
5. Y. Ren, A. Boukerche, and L. Mokdad, “Performance analysis of a selective encryption algorithm for wireless ad hoc networks,” in Proc. IEEE Wireless Commun. Netw. Conf. (WCNC), 2011, pp. 327–332.
6. S. Verma and A. Kushwaha, “Probabilistic and toss-a-coin selective encryption for lightweight data protection,” Int. J. Comput. Appl., vol. 180, no. 45, pp. 1–8, 2018.
7. P. Meshram and P. Prasad, “Selective encryption algorithm: A comprehensive literature review,” in Proc. 5th Int. Conf. Artificial Intelligence and Smart Energy (ICAIS), ser. Inf. Syst. Eng. Manage., vol. 42, Springer, Cham, 2025, doi: 10.1007/978-3-031-90482- 0_14.
8. P. Meshram and P. Prasad, “DSEA: A dynamic selective encryption algorithm for enhanced security and resource efficiency in wireless communications,” in ICT Systems and Sustainability (ICT4SD), ser. Lecture Notes Netw. Syst., vol. 1646, Springer, Cham, 2026, doi: 10.1007/978-3-032-06665-7_23.
9. K. M. Kirupa Shankar and V. Santhi, “Integrating machine learning and encryption for secure data management,” J. Cloud Comput., vol. 14, 2025.
10. P. R. Kumar, “A secure and efficient adaptive encryption system powered by machine learning,” Sci. Rep., vol. 15, 2025.
11. S. B. N. Premakumari, “Reinforcement learning-based adaptive encryption in IoT,” Sensors, vol. 25, 2025.
12. N. M. Shivsharan, “Support vector machine-based selection of encryption levels,” Informatica, vol. 49, no. 2, pp. 229–238, 2025.
13. Liu, “Semantically enhanced selective encryption,” Expert Syst. Appl., vol. 237, 2024.
14. E. Dritsas, “Machine learning in ICT: A survey,” Information, vol. 16, 2024.
15. E. Villar-Rodriguez et al., “Secure edge intelligence frameworks,” Comput. Secur., vol. 137, 2023.

16. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval. Cambridge, U.K.: Cambridge Univ. Press, 2008.
17. J. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in Advances in Large Margin Classifiers, Cambridge, MA, USA: MIT Press, 1999.
18. V. Shmatikov and M. Weinberger, "Machine learning for security," ACM Comput. Surv., vol. 54, 2022.
19. Oliveira et al., "Efficient selective encryption in 5G edge networks," IEEE Netw., vol. 38, no. 4, pp. 80–87, 2024.
20. H. B. McMahan et al., "Communication-efficient learning of deep networks from decentralized data," in Proc. Int. Conf. Artificial Intelligence and Statistics (AISTATS), 2017.
21. J. Konecny et al., "Federated learning: Strategies for improving communication efficiency," arXiv:1610.05492, 2016.
22. R. Banerjee and L. Chen, "Deep learning-based content importance detection," Neural Netw., vol. 156, 2022.
23. Y. Zhao et al., "Adaptive cryptographic strength tuning," IEEE Trans. Dependable Secure Comput., 2023.
24. L. N. Tran et al., "Textual saliency for efficient encryption," Inf. Sci., 2024.
25. H. Lee and S. Park, "Lightweight cryptography meets machine learning," IEEE Internet Comput., 2021.
26. M. Grangetto, E. Magli, and G. Olmo, "Multimedia selective encryption by means of randomized arithmetic coding," IEEE Trans. Multimedia, vol. 8, no. 5, pp. 905–917, 2006.
27. S. Lian, Z. Liu, Z. Ren, and H. Wang, "Commutative encryption and watermarking in video compression," IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 6, pp. 774–778, 2007.
28. Mutambik, "AI-driven cybersecurity in IoT: Adaptive malware detection and lightweight encryption via TRIM-SEC framework," Sensors, vol. 25, p. 7072, 2025, doi: 10.3390/s25227072.
29. A. Gaurav, V. Arya, K. T. Chui, and B. B. Gupta, "AI-based model for securing cognitive IoT devices in advanced communication systems," Int. J. Cogn. Comput. Eng., vol. 6, pp. 351–359, 2025, doi: 10.1016/j.ijcce.2025.01.009.
30. Y. Li et al., "Fast revocable attribute-based encryption with data integrity for Internet of Things," J. Syst. Archit., vol. 168, p. 103551, Nov. 2025, doi: 10.1016/j.sysarc.2025.103551.
31. P. R. Kumar, "A secure and efficient adaptive encryption system powered by machine learning," Scientific Reports, vol. 15, 2025.
32. K. Senthil et al., "Advanced privacy protection (APP) machine learning model using cryptographic techniques for IoT," Applied Sciences, vol. 7, 2025.
33. M. Zonayed et al., "Machine learning and IoT in healthcare," Smart Health, Elsevier, 2025.
34. G. Sharma et al., "A hybrid framework for secure IoT communication using lightweight cryptography and machine learning-based authentication," Peer-to-Peer Networking and Applications, Springer, 2025.
35. R. Mustafa et al., "Cross-Layer Analysis of Machine Learning Models for Energy-Efficient Adaptive Encryption in IoT Networks," Sensors, vol. 25, no. 12, 2025.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Pranay Meshram (Corresponding Author)	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Prakash Prasad		✓				✓		✓	✓	✓	✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The datasets used in this study were collected from publicly available textual repositories, including the Enron Email Dataset, social media sentiment datasets, academic text corpora, and IoT communication logs. Publicly accessible datasets were used to ensure reproducibility and transparency of the experimental evaluation.

The Enron Email Dataset is available at:
https://www.cs.cmu.edu/~enron/

Social media datasets were obtained from publicly available Kaggle repositories.

The processed feature datasets and implementation scripts used in this study are available from the corresponding author upon reasonable request. The source code of the proposed ML-DSEA framework will be publicly released after publication.