Post-quantum cryptographic algorithm identification using machine learning

B.S. Rocha, J. A. M. Xexéo and R. H. Torres

Abstract— **This research presents a study on the identification of post-quantum cryptography algorithms through machine learning techniques. Plain text files were encoded by four post-quantum algorithms, participating in NIST's post-quantum cryptography standardization contest, in ECB mode. The resulting cryptograms were submitted to the NIST Statistical Test Suite to enable the creation of metadata files. These files provide information for six data mining algorithms to identify the cryptographic algorithm used for encryption. Identification performance was evaluated in samples of different sizes. The successful identification of each machine learning algorithm is higher than a probabilistic bid, with hit rates ranging between 73 and 100%.**

Keywords— **Identification of cryptographic algorithm; Data mining; machine learning; post-quantum cryptography, NIST randomness tests.**

I. Introduction

Cryptology can be divided into cryptography and cryptanalysis. Cryptography can be defined as the science of encoded writing, ensuring that only the sender and recipient of a message have access to its content, thus providing confidentiality, irreversibility, authenticity and integrity of information [1]. Cryptanalysis is the science that aims to extract the plaintext from a ciphertext, without prior knowledge of the encryption key used [2]. To achieve its goal, cryptanalysis makes use of different types of attacks, and due to this characteristic it can be used to access the security of a cryptographic algorithm, making it essential for the development of modern cryptography [3].

In a cryptanalytic scenario, little information is available besides ciphertexts and, a priori, it is not known which algorithm was used to encrypt the plaintexts. Therefore, the process of identifying the algorithm used in the encryption process considerably reduces the cryptanalysis effort and is part of the set of activities that contribute to the decoding of the message, which also includes the determination of the key size and the key itself [4].

The development of modern cryptographic algorithms is based on complex mathematical models, which aim to dissipate any patterns that may exist in the ciphertexts produced by them [5] and make difficult the process of determining the algorithm used.

In the literature, there are several studies that analyze the task of determining cryptographic algorithms based on the recognition of patterns in their ciphertexts and on the use of machine learning algorithms, with different methods being proposed for identification through the use of classifier algorithms, as in [26][27][28]. However, no identification research involving post-quantum cryptographic algorithms was found.

Given this gap, the object of this research is the analysis of cryptograms produced by post-quantum cryptographic algorithms, aiming at the subsequent identification of the generator algorithm, through the use of machine learning algorithms, considering a ciphertext-only scenario, in which only ciphertext samples are found available.

Ciphertexts from the post-quantum cryptographic algorithms Frodo, CRYSTALS Kyber, NTRU and Saber – participants in the selective contest implemented by the National Institute of Standards and Technology (NIST) – were analyzed, and useful information extracted from their cryptograms allowed identifying the algorithm's employees with hit rates that ranged between 73.3% and 100%.

Although the scope of this research is post-quantum algorithms, the AES and Blowfish symmetric cryptography algorithms were also analyzed, so that the results obtained could be compared with other results already reported in similar research.

II. Literature Revision

There is a wide variety of cryptographic and machine learning algorithms, among which some were used in this research.

A. Cryptographic Algorithms

Blowfish algorithm was conceived as an alternative to the Data Encryption Standard (DES), due to this algorithm's vulnerabilities to brute force attacks. In [6], Nie, Song and Zhi analyzed the processing speeds and energy consumption of these two algorithms and concluded that Blowfish is significantly faster than DES and that both have similar energy consumption. In research [7], it was concluded that Blowfish provides greater security than Advanced Encryption Standard (AES) and 3DES, due to the key sizes used. Poonia and Yadav analyzed different configurations of the Blowfish algorithm in [8], and presented changes that made it more secure and compact than its original implementation.

The Rijndael block cipher was the winner of the selective competition organized by NIST, between January 1997 and October 2000, which instituted the Advanced Encryption Standard (AES) and replaced DES, in accordance with FIPS

B.S. Rocha, Military Institute of Engineering (IME), Rio de Janeiro, Brazil, rocha.bruno@ime.eb.br
J. A. M. Xexéo, Military Institute of Engineering (IME), Rio de Janeiro, Brazil, xexeo@ime.eb.br
R. H. Torres, University of Pará, renatohidaka@ufpa.br

197. According to [9] and [10], the construction of AES is based on a permutation-replacement network, unlike its predecessor which was based on a Feistel structure. AES employs a fixed-size block of 128 bits and keys of 128, 192 or 256 bits. The key size specifies the number of rounds that convert the input – the plaintext – to the final output – the ciphertext – : 10, 12, and 14 rounds for 128, 192, and 256-bit keys, respectively.

The N-th degree Truncated polynomial Ring Units (NTRU) post-quantum public-key encryption algorithm has a simple deployment, high encryption and decryption speeds, and reasonably small keys whose sizes range from 699 to 2401 bytes. According to [11], the NTRU encryption and decryption processes are based on the combination of polynomial algebra with a clustering principle based on elementary probability theory. The security of NTRU derives from the interaction of the polynomial combination system with the independence of the reduction modulus of two prime numbers.

The post-quantum cryptographic algorithm SABER has indistinguishability level IND-CCA2 and is based on lattices. This algorithm comprises a public key encryption scheme and a key encapsulation mechanism, respectively called SABER.PKE and SABER.KEM. According to research [12], the SABER public-key encryption scheme employs the Module Learning with Rounding (MLWR) mathematical problem, a variant of the Learning with Errors (LWE) problem that differentiates by rounding the samples to create noise instead of of adding errors to them.

The post-quantum public-key algorithm Cryptographic Suite for Algebraic Lattices (CRYSTALS) Kyber is based on the Module Learning with Errors (MLWE) mathematical problem and has an IND-CPA level of indistinguishability. Combining the use of the Fujisaki–Okamoto (FO) transformation and the Kyber.PKE public key scheme, the Kyber.KEM key encapsulation mechanism has IND-CCA2 degree of indistinguishability and shares 32-byte session keys [13]

Another participant in the selective competition organized by NIST to define a new US standard for cryptographic algorithms resistant to quantum attacks, Frodo comprises a public key encryption scheme and a key encapsulation mechanism, and its security lies in the standard, lower lattice problem. to the RLWE and MWE employed respectively by Saber and CRYSTALS Kyber, resulting in limited practical applications [14].

B. Machine Learning Algorithms

Due to the results from Support Vector Machines (SVM), there are records in the literature of the application of this machine learning algorithm in different areas, such as pattern recognition in texts [17] and in bioinformatics [18]. According to the theory developed by Vapnik [19], SVMs are based on statistical learning, whose principles allow the correct prediction of data classes belonging to the same set in which the learning took place.

According to [20], k-Nearest Neighbors is a non-parametric algorithm, whose accuracy is related to the analyzed dataset. It does not need a training set to perform the learning and conducts the classification process from the test set, not performing any transformation or calculations on these data. The classification takes place based on a certain number of neighbors, whose value is variable and which, from a threshold, causes the classification error to increase substantially.

The NaiveBayes classifier [21] uses a probabilistic model in which there are no hidden attributes that influence the prediction of a class. The prediction result will indicate the class with the highest probability of occurrence given a set of attributes. However, this classification process has several flaws, such as the inadequate treatment of the superposition of classes and the sensitivity to the difference in the number of samples of the classes in the training and test sets. ComplementNaiveBayes was developed to mitigate the faults of distorted training of the NaiveBayes classifier [22] and to increase its processing speed and accuracy.

After the random selection of classes and samples, the Random Forest (RF) classification algorithm builds several different decision trees and integrates them to obtain the best decision result. When the sample to be classified is entered, the classification result is determined by most of the classification results from each decision tree.

The Logistic Regression (LR) algorithm is a statistical classifier that from a set of independent variables allows the prediction of a certain category, often binary, as a function of one or more continuous and/or binary variables [23].

C. Related Research

In research [24], SVMs were used to distinguish cryptograms from DES algorithms in Electronic Code Book (ECB) and Cipher Block Chaining (CBC), Triple DES (3DES), Blowfish and AES modes from 4,000-bit cleartexts using different configurations of encryption keys, and recorded accuracy between 26.79% and 97.78%.

In [25], R. Manjula and R. Anitha designed a system capable of identifying the DES, 3DES, AES, Blowfish, RC2, RC4, IDEA, RSA and ECC encryption algorithms through the identification of characteristics based on the entropy of the cryptograms by them. For each algorithm, 30 text files of 512 KB were encrypted in ECB mode. The designed system used the C4.5 classifier, which is based on pruning methods in decision trees to accelerate and improve the classification process, and the results obtained varied between 70% and 75%.

Chou et al. [26] analyzed cryptograms produced after 3000 text, audio and image files were encrypted by AES and DES algorithms in ECB and CBC modes, and used SVM to identify their generating algorithms. The conclusion of the research indicates that the SVM classifier obtained better performance in the ciphers generated in the ECB mode, presenting accuracies that varied between 48.49% and 100%.

In the research [27] published in 2016, Mello and Xexéo analyzed cryptograms from the ARC4, Blowfish, DES, Rijdael, RSA, Serpent and Twofish ciphers, after texts in seven different languages were encrypted in ECB mode. The authors used C4.5, Complement Naive Bayes, PART, Multilayer Perceptron, FT and WiSARD classifiers to classify blocks whose sizes ranged from 2 to 34 bits. When 30-bit blocks were classified, the research concluded that a large portion of the classifiers distinguished the generating algorithms with 100% accuracy.

In the scheme proposed by Mishra et al in [28], cryptograms generated by the AES, DES and Blowfish algorithms are submitted to three distinct blocks that work simultaneously. The first checks the length of the block/bit stream, the second analyzes the entropy and recurrence of the samples and the last one employs Decision Trees. The research analyzed 10, 200, 700 and 2000 samples of 128, 256, 512 and 1024 bits, ranking them at 83%, 64%, 87.3% and 89.1%.

William et al proposed a distinction attack to identify block ciphers in [29], combining neural networks with linguistic patterns that generate signatures in ciphertexts. Employing a single 128-bit key, 240 plaintexts of 6144 and 8192 bytes in eight different languages were encrypted by the MARS, RC6, Rijndael, Serpent and Twofish algorithms. The grouping processes allowed the formation of well-defined groups, allowing the total distinction and classification of cryptograms for samples of 8192 bytes.

Wu et al [30] selected 1000 plaintexts of 1.1 MB from the Open American National Corpus (OANC) and encrypted them with AES-128, KASUMI, 3DES, PRESENT, RSA and ElGamal algorithms, in CBC mode. After the generated cryptograms were subjected to three of the fifteen tests contained in the NIST SP 800-22 battery of statistical tests, a deep learning algorithm was employed to distinguish the generating algorithms. The authors obtained identification rates of around 90%.

In the research by Zhao et al published in [31], 10 NIST randomness tests were used to extract useful information from 500 plaintext files with sizes of 1, 8, 64, 256 and 512 KB encrypted by the AES, Blowfish, Camellia, DES, 3DES and IDEA in ECB mode. Employing a hybrid model composed of the Random Forest and Logistic Regression classifiers, the authors achieved identification rates of 80% in certain cases.

III. Cryptographic Algorithm Identification Scheme

A. Statistical Tests and Data Mining

The security of a cryptographic algorithm can be evaluated through the use of statistical tests that, based on probabilistic indices, determine whether a binary sequence has characteristics of a random sequence. Considering the large number of existing tests, no set can be considered a “complete” package, in order to specify, without any margin of error, whether a sequence is random or not. Therefore, all the results obtained must be interpreted with caution to avoid erroneous conclusions [32].

Developed by NIST to validate the use of random or pseudorandom number generators in cryptographic applications, the NIST SP 800-22rev1a suite is a package composed of the following statistical tests: frequency, frequency within a block, runs, longest-run-of-ones in a block, binary matrix rank, discrete fourier transform, non-overlapping template matching, overlapping template matching, maurer's, linear complexity, serial, approximate entropy, cumulative sums, random excursions and random excursions variant.

Data mining is the process of extracting patterns in large masses of data [15][27], through the use of algorithms that identify connections and extract useful information from this mass, helping decision-making and analysis of future trends through prediction of values or classes [16].

Employing the methodology presented in [30] [31], this research submitted ciphertexts generated by post-quantum cryptosystems FrodoKEM-1344, CRYSTALS Kyber1024, NTRU-HRSS-701 and FireSaber in ECB operating mode to the 15 statistical tests that are components of the NIST SP 800-22rev1a suite, according to figures 01 and 02. In order to allow the comparison of the results of this research with related works, in addition to the post-quantum algorithms, cryptograms from the AES and Blowfish algorithms, which were widely analyzed by other researchers, were also analyzed.

To form the data set, 100 plain texts of 20, 40, 60, 80 and 100 KB, randomly selected and without repetition, are encrypted by NTRU, CRYSTALS Kyber, Saber, Frodo, AES and Blowfish cryptosystems, totaling 500 files for each algorithm and 3.000 for the six figures. Then, the generated cryptograms are analyzed by the open source tool SP 800-22-tests-master, which was used in [30] and [31] to generate the representative vectors constituting the set of metadata, which make explicit characteristics of the cryptograms.

The same number of samples is generated for all analyzed cryptographic algorithms, in order to eliminate the possibility of occurrence of privileges in the identification of one algorithm over another, and each sample of clear text is associated with a different key, in order to avoid possible influence on the data mining process. According to [27], the reuse of keys in cryptography can induce biases in classification algorithms and consequently mask results.

Due to its exploratory nature, the corpora of this research consist of texts in Portuguese from the literary works: Elite da Tropa (vol. 1 and 2), Fogo & Sangue and Holy Bible. Literary works that have different linguistic constructions were chosen in order to minimize the possible presence of language defects in plain texts, and the corpora is constituted by a single language because, according to the conclusion presented in [27], different languages do not influence data mining, nor are they relevant to the classification process of machine learning tools.

After producing the set of metadata, it is divided into two portions. The first consists of 70% of the samples in the set and is intended for training the classifiers; the second, composed of the remaining 30%, is used as a test set. Both sets are submitted to the SVM, KNN, NB, RF, LR classifiers and to a hybrid model, based on the ensemble learning concept, called Hybrid Logistic Regression and Random

Forest (HLRNRF). The confusion matrices of the classifiers are obtained after completing the data mining and classification processes of the machine learning tools. The results obtained by the proposed method are evaluated according to the criteria of accuracy, precision and recall.

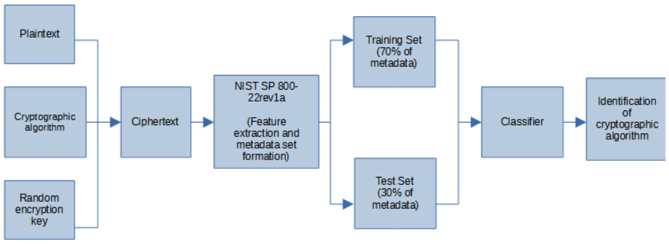

Figure 01: Post-quantum cryptographic algorithm identification method.

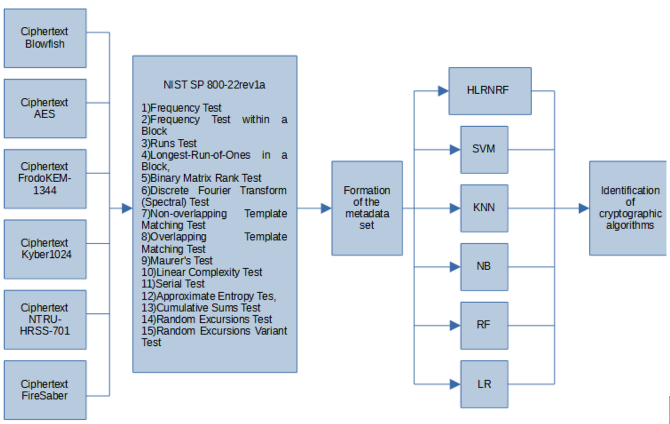

Figure 02: NIST SP 800-22rev1a suite extracting features

B. Hybrid Classifiers

Most of the cryptographic algorithm identification schemes present in the literature employ single-layer machine learning classifiers. In [27], Complement Naive Bayes was used to distinguish the DES, Blowfish, RSA, ARC4, Rijndael, Serpent and Twofish algorithms. In [33], Fan & Zhao employed three classifiers - RF, LR and SVM - to perform a distinction attack on DES, 3DES, AES-128, AES-256, IDEA, SMS4, Blowfish and Camellia-128 block ciphers. However, single-layer classifiers may present low accuracies, overfitting and difficulties to find adequate parameters according to [34].

In order to minimize possible problems that may exist in single-layer classifiers, this research evaluated the use of a classifier based on ensemble learning, which was called hybrid Logistic regression and Random Forest algorithm (HLRNRF).

Ensemble learning, also called cluster learning, is based on combining several single-layer predictors to produce a more complex and effective clustered model. To perform multilayer integration, stacking is performed, which consists of using the metadata as input to the first classifier, and using its output as input to the classifier of the next layer.

IV. Results and Performance Analysis

Commonly, the most used evaluation criteria in classification tasks are accuracy, precision and recall. Accuracy can be defined as the proportion that indicates, of the positive and negative classifications of the model, how many were correct. Precision is the proportion that indicates, of the positive classifications of the model, how many were correct, and recall is the proportion that indicates, of the existing positive samples, how many the model was able to classify correctly.

To evaluate the classification results, the confusion matrix presented in table 01 can be used, which presents the four possible results: True Positive (VP), True Negative (VN), False Positive (FP) and False Negative (FN). The mathematical expressions presented below allow the calculation of the presented evaluation criteria — accuracy, precision and recall — based on the possible results of the confusion matrix:

$$\text{Accuracy} \text{ ¿ } \frac{(TP+TN)}{(TP+TN+FP+FP)}$$

$$\text{Precision} \text{ ¿ } \frac{TP}{(TP+FP)}$$

$$\text{Recall} = \frac{TP}{(TP+FN)}$$

Confusion matrix of ranking results.		
Real situation	Prediction result	
	Positive	Negative
Positive	TP (True Positive)	FN (False Negative)
Negative	FP (False Positive)	TN (True Negative)

Table 01: Results of Confusion matrix

In the researches of distinction of cryptographic algorithms studied, the accuracy of the classifier was the main evaluation parameter adopted. Therefore, this research adopted accuracy as the main criterion to evaluate the performance of the classifier.

To evaluate the performance of the proposed ensemble learning model in the identification of post-quantum algorithms, they were compared with five classical machine learning models: K — Nearest Neighbors (KNN), Logistic regression (RL), Complement Naive Bayes (CNB), Random Forest and Support Vector Machine (SVM).

In this research, the KNN classifier was experimented with different distance metrics (Hamming, Manhattan, Minkowski and Euclidean) and different K values (1,3,5 and 10). All models were applied to the same datasets, and classification performance was evaluated at different ciphertext file sizes (20, 40, 60, 80 and 100 KB). After analyzing the results obtained, it was verified that the best configuration was for the value of k equal to one and the Euclidean distance.

The distinction model based on the 15 types of useful resources extracted was employed. Then, the accuracy, precision and recall values of the HLRNRF, KNN, LR, NB, RF and SVM models were calculated for different ciphertext file sizes encrypted by the classical algorithms AES, Blowfish and post-quantum FrodoKEM-1344, CRYSTALS Kyber1024, NTRU-HRSS-701 and FireSaber. The results obtained are shown in table 02.

Evaluation criteria	Classifier	File Size (KB)				
		20	**40**	**60**	**80**	**100**
Accuracy	HLRNRF	**0.733**	**0.607**	**0.717**	**0.817**	**0.941**
	KNN	**0.667**	**0.738**	**1.000**	**0.933**	**0.882**
	LR	**0.667**	**0.557**	**0.966**	**0.850**	**0.863**
	NB	**0.617**	**0.623**	**0.877**	**0.833**	**0.627**
	RF	**0.583**	**0.623**	**1.000**	**0.850**	**0.765**
	SVM	**0.550**	**0.485**	**0.667**	**0.746**	**0.569**
Precision	HLRNRF	**0.729**	**0.685**	**0.797**	**0.841**	**0.956**
	KNN	**0.703**	**0.786**	**1.000**	**0.941**	**0.885**
	LR	**0.820**	**0.545**	**0.970**	**0.897**	**0.868**
	NB	**0.649**	**0.707**	**0.886**	**0.825**	**0.622**
	RF	**0.632**	**0.744**	**1.000**	**0.865**	**0.773**
	SVM	**0.616**	**0.558**	**0.667**	**0.767**	**0.601**
Recall	HLRNRF	**0.733**	**0.607**	**0.717**	**0.817**	**0.941**
	KNN	**0.667**	**0.738**	**1.000**	**0.933**	**0.882**
	LR	**0.667**	**0.557**	**0.967**	**0.850**	**0.863**
	NB	**0.617**	**0.623**	**0.883**	**0.833**	**0.627**
	RF	**0.583**	**0.623**	**1.000**	**0.850**	**0.765**
	SVM	**0.550**	**0.508**	**0.667**	**0.767**	**0.569**

Table 02: Classification results

The first column of table 02 presents the evaluation criteria of the identification process and the second column shows the sizes of the ciphertext files. The average accuracies of the HLRNRF, KNN, LR, NB, RF and SVM classifiers in the different sizes of ciphertext files are respectively 76.3%, 84.4%, 78%, 71.5%, 76.4% and 60 .3%.

It was observed that the accuracy of the classifiers varies according to the size of the ciphertext files, indicating the influence of this parameter on the prediction result. According to the results obtained, it can be affirmed that the hybrid model HLRNRF presented better global performance for samples of 20KB (73.3%) and 100KB (94.1%). For samples of 40 KB, 60 KB and 80 KB, the KNN presented greater accuracy, having correctly classified 73.8%, 100% and 93.3% of the samples, respectively.

Considering that the HLRNRF classifier presented better results for samples of 20KB and 100KB, as well as the KNN presented greater accuracies for 40 KB, 60 KB and 80 KB, tables 03, 04, 05, 06 and 07 present the confusion matrices obtained by it.

		Predict					
		AES	**Blowfish**	**Frodo**	**Kyber**	**NTRU**	**Saber**
Rea l	**AES**	75%			25%		
	Blowfish		100%				
	Frodo		12,5%	87,5 %			
	Kyber	80%			20%		
	NTRU					42,9%	57,1 %

	Saber					30,8%	69,2 %

Table 03: HLRNRF classifier confusion matrix - 20KB samples

		Predict					
		AES	**Blowfish**	**Frodo**	**Kyber**	**NTRU**	**Saber**
Real	**AES**	61,5 %			38,5%		
	Blowfish		100%				
	Frodo			100%			
	Kyber	37,5 %			50%		12,5 %
	NTRU					66,7%	33,3 %
	Saber					42,9%	57,1 %

Table 04: KNN classifier confusion matrix - 40KB samples

		Predict					
		AES	**Blowfish**	**Frodo**	**Kyber**	**NTRU**	**Saber**
Real	**AES**	100 %					
	Blowfish		100%				
	Frodo			100%			
	Kyber				100%		
	NTRU					100%	
	Saber						100%

Table 05: KNN classifier confusion matrix - 60KB samples

		Predict					
		AES	**Blowfish**	**Frodo**	**Kyber**	**NTRU**	**Saber**
Real	**AES**	100 %					
	Blowfish		100%				
	Frodo			100%			
	Kyber				100%		
	NTRU					85,7%	14,3 %
	Saber					23,1%	76,9 %

Table 06: KNN classifier confusion matrix - 80KB samples

		Predict					
		AES	**Blowfish**	**Frodo**	**Kyber**	**NTRU**	**Saber**
Real	**AES**	100 %					
	Blowfish		100%				
	Frodo			100%			
	Kyber				100%		
	NTRU					100%	
	Saber					33,3%	66,7 %

Table 07: HLRNRF classifier confusion matrix - 100KB samples

In this research, the cryptograms of the AES and Blowfish algorithms in ECB encryption mode were submitted to the fifteen component tests of the NIST SP 800-22rev1a suite. The results obtained here are significantly superior to those obtained by YUAN, Ke et al. in [34] and [35], where ciphertext files of 1, 8, 64, 256 and 512 KB were submitted to ten tests contained in this test battery. Special attention must be paid to the KNN classifier, which showed total accuracy in samples of 60KB. It can be inferred that the total number of tests used contributed directly to the increase in the accuracy obtained by the machine learning classifiers.

V. Conclusion and Future work

This research is primarily intended for the identification of post-quantum cryptographic algorithms in a ciphertext-only scenario. It was possible to distinguish ciphertext files of different sizes encrypted by the classical AES, Blowfish and post-quantum algorithms FrodoKEM-1344, CRYSTALS Kyber1024, NTRU-HRSS-701 and FireSaber - in ECB mode - through the use of traditional classifiers present in machine learning and an ensemble learning-based model called HLRNRF.

The results obtained indicate that the size of the ciphertext file and the difference in the cryptographic algorithm used in the encryption process are factors that influence the identification accuracy, which in some cases reached total accuracy of 100%.

The model and scheme proposed in this article are mainly suitable for the identification of cryptographic algorithms. All the results shown above are superior to the random choice index, whose approximate value is 16.67%, and indicate that the scheme based on ensemble learning has a higher accuracy compared to the scheme based on a single-layer classifier on samples of 20KB and 100KB. For samples of 40KB, 60KB and 80KB, the KNN classifier proved to be more favorable.

In the ciphertext-only scenario, in the future we will delve deeper into the research of extracting useful information from post-quantum cryptographic algorithms operating in different block cipher encryption modes, especially in the CBC mode. Additionally, the set learning-based identification scheme is worthy of further exploration and has certain positive significance for future research on block cipher algorithm identification.

References

[1] STALLINGS, William. Criptografia e segurança de redes. Princípios e práticas, ch. 6. 2006.

[2] SCHNEIER, Bruce. Applied cryptography: protocols, algorithms, and source code in C john wiley & sons. Inc: California, 1996.

[3] PAAR, Christof; PELZL, Jan. Understanding cryptography: a textbook for students and practitioners. Springer Science & Business Media, 2009.

[4] PFLEEGER, Shari Lawrence. Engenharia de software: teoria e prática. 2. ed. São Paulo: Prentice Hall, 2004. 537 p.

[5] SCHNEIER, B. Applied Cryptography. 2. ed. New York, NY, USA: John Wiley & Sons, 1996.)

[6] NIE, T.; SONG, C.;ZHI, X.Performance Evaluation of DES and Blowfish Algorithms, International Conference on Biomedical Engineering and Computer Science (ICBECS), pp.1-4, Wuhan, 2010

[7] VERMA, O. P.; AGARWAL, R.;DAFOUTI, D.;TYAGI, S. Performance Analysis Of Data Encryption Algorithms, 3rd International Conference on Electronics Computer Technology (ICECT), pp. 399-403, Kanyakumari, 2011

[8] POONIA, V.;YADAV, N. S. Analysis of modified Blowfish Algorithm in different cases with various parameters, International Conference on Advanced Computing and Communication Systems, pp. 1-5, Coimbatore, 2015.doi: 10.1109/ICACCS.2015.7324114

[9] DAEMEN, Joan; RIJMEN, Vincent. The Design of Rijndael, Berlin, Springer, 2002.doi: 10.1007/978-3-662-04722-4

[10] Abdullah, A. M. (2017). Advanced encryption standard (AES) algorithm to encrypt and decrypt data. Cryptography and Network Security, 16, 1-11.

[11] HOFFSTEIN, Jeffrey; PIPHER, Jill; SILVERMAN, Joseph H. NTRU: A ring-based public key cryptosystem. In: International algorithmic number theory symposium. Springer, Berlin, Heidelberg, 1998. p. 267-288.

[12] BEIRENDONCK, Michiel Van et al. A side-channel-resistant implementation of SABER. ACM Journal on Emerging Technologies in Computing Systems (JETC), v. 17, n. 2, p. 1-26, 2021.

[13] BOS, Joppe et al. CRYSTALS-Kyber: a CCA-secure module-lattice-based KEM. In: 2018 IEEE European Symposium on Security and Privacy (EuroS&P). IEEE, 2018. p. 353-367.

[14] GUO, Qian; JOHANSSON, Thomas; NILSSON, Alexander. A key-recovery timing attack on post-quantum primitives using the Fujisaki-Okamoto transformation and its application on FrodoKEM. In: Annual International Cryptology Conference. Springer, Cham, 2020. p. 359-386.

[15] FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From data mining to knowledge discovery in databases. AI magazine, v. 17, n. 3, p. 37-37, 1996.

[16] Witten, I. H., Frank, E., Hall, M. A. Data Mining Practical Machine Learning Tools and Techniques, 3rd edition, Morgan Kaufmann, Burlington, 2011.

[17] K. I. Kim, K. Jung, S. H. Park, and H. J. Kim. Support vector machines for texture classification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(11):1542–1550, 2002.

[18] B. Schölkopf, I. Guyon, and J. Weston. Statistical learning and kernel methods in bioinformatics. In P. Frasconi and R. Shamir, editors, Artificial Intelligence and Heuristic Methods in Bioinformatics, pages 1–21. IOS Press, 2003.

[19] V. N. Vapnik. The nature of Statistical learning theory. Springer-Verlag, New York, 1995.

[20] Aha, D.W., Kibler, D., Albert, M.K.: Instance-Based learning algorithms. Kluwer Academic Publishers, 1991

[21] JOHN, George H.;LANGLEY, Pat. Estimating Continuous Distributions in Bayesian Classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, 338-345, 1995.

[22] RENNIE, J. D. M.;SHIH, L.;TEEVAN, J.;KARGER, D. R. Tackling the Poor Assumptions of Naive Bayes Text Classifiers, Proceedings of the Twentieth International Conference on Machine Learning, Washington DC, 2003.

[23] LAVALLEY, Michael P. Logistic regression. Circulation, v. 117, n. 18, p. 2395-2399, 2008.

[24] Dileep AD, Sekhar CC (2006) Identification of block ciphers using support vector machines. In: The 2006 IEEE International Joint Conference on Neural Network Proceedings, pages 2696–2701. IEEE. https://doi.org/10.1109/IJCNN.2006.247172

[25] Manjula R, Anitha R (2011) Identification of encryption algorithm using decision tree. In: Communications in Computer and Information Science, volume 133, pages 237–246. Springer. https://doi.org/10.1007/978-3-642-17881-8_23

[26] Chou JW, Lin SD, Cheng CM (2012) On the effectiveness of using state-of-the-art machine learning techniques to launch cryptographic distinguishing attacks. In: Acm Workshop on Security and Artificial Intelligence, pages 105–110

[27] DE MELLO, Flavio Luis; XEXEO, Jose Antonio Moreira. Cryptographic algorithm identification using machine learning and massive processing. IEEE Latin America Transactions, v. 14, n. 11, p. 4585-4590, 2016

[28] Mishra S, Bhattacharjya A (2013) Pattern analysis of cipher text: A combined approach. In: 2013 International Conference on Recent Trends in Information Technology (ICRTIT), pages 393–398.

[29] De Souza WAR, Tomlinson A (2013) A distinguishing attack with a neural network. In: 2013 IEEE 13th International Conference on Data Mining Workshops, pages 154–161
[30] Yang W, Tao W, Jindong L (2015) Research on a new method of statistical detection of block cipher algorithm ciphertext. Journal of Ordnance Engineering College 000(003):58–64. https://doi.org/10.3969/ j.issn.1008-2956.2015.03.011
[31] Zhicheng Z, Yaqun Z, Fengmei L (2019) Recognition scheme of block cipher system based on randomness test. Journal of Cryptography 6(2):177–190
[32] BASSHAM III, Lawrence E. et al.Sp 800-22 rev. 1a. a statistical test suite for random and pseudorandom number generators for cryptographic applications. National Institute of Standards & Technology, 2010.
[33] FAN, SiJie; ZHAO, YaQun. Analysis of cryptosystem recognition scheme based on Euclidean distance feature extraction in three machine learning classifiers. In: Journal of Physics: Conference Series. IOP Publishing, 2019. p. 012184.
[34] YUAN, Ke et al. A block cipher algorithm identification scheme based on hybrid k-nearest neighbor and random forest algorithm. PeerJ Computer Science, v. 8, p. e1110, 2022.
[35] YUAN, Ke et al. A Block Cipher Algorithm Identification Scheme Based on Hybrid Random Forest and Logistic Regression Model. Neural Processing Letters, p. 1-19, 2022.

Bruno dos Santos Rocha graduated in Naval Sciences, with an emphasis on electronics, from Escola Naval (2012) and post-graduated in Electronic Warfare from the Center for Electronic Warfare and Acoustics of the Navy. He worked aboard ships belonging to the Brazilian Navy and at the Operating Systems Support Center. He is currently pursuing a master's degree in Systems and Computing at the Military Institute of Engineering.

Jose Antonio Moreira Xexeo holds a degree in Communications Engineering from the Instituto Militar de Engenharia (1972), a master's degree in Systems and Computing from the Instituto Militar de Engenharia (1983) and a PhD in Systems and Computer Engineering from the Federal University of Rio de Janeiro (2001) . He was an institutional and undergraduate course evaluator at INEP. He is currently professor emeritus of the undergraduate and master's degrees in Computer Engineering at the Instituto Militar de Engenharia (IME), where he conducts research in cryptology. He led the team that designed and implemented at IME, in 1985, the first Computer Engineering course in Brazil, of which he was its first coordinator. He taught in the undergraduate courses in Business Administration and Production Engineering at Universidade Veiga de Almeida. He created, coordinated and taught, for 10 years, the Computer Science course at Bennett Colleges. He created, coordinated and taught, for four years, the Computer Engineering course at Faculdade Salesiana de Macaé. He worked for more than 10 years with technological development in the industrial area of computing. He has academic background in engineering and computer science, particularly cryptology. He has been working in Higher Education as a teacher and course coordinator for almost 40 years.

Renato Hidaka Torres holds a bachelor’s degree in Computer Science from the University Center of Pará (2010), a Master's in Systems and Computing from the Military Institute of Engineering (2012) and a PhD in Computer Science from the Federal University of Pará (2019). He was a professor at the Instituto de Estudos Superiores da Amazônia (2012-2015), professor at the Federal Institute of Education, Science and Technology of Pará (2015-2020). Since 12/14/2018 he has been part of the Bank of SINAES Evaluators (BASiS) of INEP / MEC. Since 03/17/2020 he has been a permanent professor at the Institute of Exact and Natural Sciences at the Federal University of Pará, and since 08/17/2021 he has been a permanent professor at the Postgraduate Program in Public Security at the Institute of Philosophy and Sciences. Humanities at the Federal University of Pará. Main lines of research: Information Security, Data Mining and Cyber Crime.