International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 6, December 2025, pp. 5728~5745
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i6.pp5728-5745

Securing healthcare data and optimizing digital marketing through machine learning: the CAML-EHDS framework

Fathi Abderrahmane[1], Mouyassir Kawtar[1], Ali Waqas[2], Fandi Fatima Zahra[3], Kartit Ali[1]
[1]National School of Applied Sciences, The Information Technology Laboratory at Chouaib Doukkali University El Jadida, Morocco
[2]Department of Computer Science, University of Engineering and Technology, Lahore, Pakistan
[3]LTIM, Hassan II University, Casablanca, Morocco

Article Info

Article history:

Received Feb 1, 2025
Revised Jul 30, 2025
Accepted Sep 15, 2025

Keywords:

CAML-EHDS
Cryptography
Digital marketing
Healthcare data security
Machine learning
Long short-term memory networks

ABSTRACT

Current healthcare data systems face major challenges in preventing unauthorized access, ensuring compliance with data privacy regulations, and enabling intelligent secondary use of patient information. To address these issues, we introduce cluster-based analysis with machine learning for enhanced healthcare data security (CAML-EHDS), a unified framework that combines homomorphic encryption, attribute-based elliptic curve cryptography (ECC), and semantic clustering with machine learning. CAML-EHDS improves upon existing models by offering fine-grained access control, adaptive threat detection, and data-driven insights while preserving privacy. Experimental results show that CAML-EHDS achieves up to 98% classification accuracy with low node count, and maintains 94% accuracy even at high node distribution levels, while ensuring encryption time under 24 seconds and acceptable data loss below 29%. Moreover, in comparative analysis with state-of-the-art models (support vector machine (SVM), random forest (RF), and decision tree (DT)), CAML-EHDS outperforms all in key metrics with an accuracy of 0.96. These results demonstrate CAML-EHDS's potential for real-world deployment in secure, scalable, and intelligent healthcare environments, including privacy-aware digital marketing integration.

Corresponding Author:

Fathi Abderrahmane
National School of Applied Sciences, The Information Technology Laboratory at Chouaib Doukkali University
National Road No. 1, P.O. Box 299, El Jadida 24000, Morocco
Email: Abdou.ft@gmail.com

1. INTRODUCTION

In this paper, we introduce cluster-based analysis with machine learning for enhanced healthcare data security (CAML-EHDS), a robust framework designed to address the escalating challenges of protecting sensitive healthcare data in an increasingly digital landscape. The digitization of medical records, while offering numerous benefits, has also heightened the risk of cyber threats, necessitating advanced analytical techniques to safeguard patient information [1]. Within this framework, cryptographic methods and authorization mechanisms are seamlessly integrated to safeguard healthcare data. Leveraging techniques such as homomorphic encryption and attribute-based elliptic curve cryptography (ECC) schemes [2], this model ensures that sensitive information remains encrypted and accessible only to authorized entities, thus mitigating the risk of unauthorized access or data breaches. This addresses a critical gap identified in prior research, where robust access control and data protection mechanisms are often lacking.

Furthermore, CAML-EHDS incorporates cluster-based analysis with state-of-the-art machine learning algorithms to uncover hidden patterns and identify potential security threats within healthcare data.

By employing semantic clustering, ranking clusters, and computing similarity indices, the model provides invaluable insights into the underlying structure of healthcare datasets, thereby increasing classifier security and threat detection capabilities. Through the utilization of long short-term memory networks (LSTMs) [3] and transfer learning, specifically using pre-trained models like bidirectional encoder representations from transformers (BERT), the model enables healthcare organizations to classify security threats with precision and efficacy, therefore establishing a more secure and resilient healthcare data ecosystem. Compared to existing schemes such as GHZ, J, HZ, XZY, and SCH, which often lack pairing-free operations, ECC-based methods, or key-escrow mechanisms, CAML-EHDS offers a more comprehensive security solution.

In addition to its core security features, the research model also offers seamless integration with digital marketing strategies. By leveraging insights gained from unified threat detection and cluster-based analysis, CAML-EHDS enables organizations to tailor their digital marketing strategies effectively. Through targeted advertisements and personalized content for healthcare products and services based on classified data and detected threats, the model facilitates enriched engagement and customer satisfaction [4]. Moreover, the proposed model ensures that the data used for digital marketing is secure and compliant with privacy regulations, thereby providing organizations with peace of mind while maximizing the effectiveness of their marketing efforts [5]. This integration, demonstrating the practical application of our security framework, is shown through comparative analyses and performance evaluations in the results section, highlighting CAML-EHDS's superior performance in maintaining data security while optimizing marketing strategies.

The methodological novelty of this study lies in the integration of homomorphic encryption, attribute-based ECC) and semantic clustering with machine learning (LSTM and transfer learning via BERT) within a unified framework tailored for healthcare data security. CAML-EHDS addresses critical gaps in existing frameworks, notably the lack of secure, privacy-compliant models capable of real-time threat detection and encrypted data processing. Unlike conventional models, CAML-EHDS simultaneously enhances data confidentiality, improves anomaly classification accuracy (96%), and supports secure digital marketing strategies aligned with GDPR and health insurance portability and accountability act (HIPAA). The comparative results presented in this paper highlight its superior encryption efficiency, reduced computational overhead, and increased resilience to cyberattacks. These contributions offer promising implications for future applications in secure healthcare infrastructures, including cloud-based systems, IoT environments, and AI-driven medical data services.

2. LITERATURE REVIEW

The escalating digitization of medical records and the increasingly sophisticated landscape of cyber threats have underscored the critical need for robust healthcare data security. While prior research has explored various facets of this challenge, significant limitations persist, which CAML-EHDS is designed to address. Acar *et al.* [6] demonstrated the promise of homomorphic encryption for safeguarding sensitive medical records. However, their approach lacked the fine-grained access control necessary in collaborative healthcare environments, leaving data vulnerable to internal breaches. Similarly, Imam *et al.* [7] proposed an attribute-based ECC scheme, but they did not adequately address the complexities of key management in dynamic healthcare settings, which CAML-EHDS tackles with its robust key-escrow system. Cluster-based analysis has also been explored for anomaly detection, as evidenced by Festag *et al.* [8], who investigated semantic clustering algorithms. However, their work did not integrate real-time machine learning for dynamic threat detection, a critical component of CAML-EHDS. Prasad *et al.* [9] explored similarity-based clustering, but their approach lacked the temporal analysis capabilities provided by CAML-EHDS's LSTM-based component. Moreover, Balhareth and Ilyas [10] utilized CNNs for security breach detection in medical imaging, and Rajkomar *et al.* [11] employed LSTMs for temporal pattern recognition in electronic health records; however, these studies focused on isolated aspects of data security and did not offer a comprehensive framework that integrates multiple security layers.

CAML-EHDS, in contrast, combines cryptographic methods, advanced clustering, and sophisticated machine learning, including both LSTM and BERT, to provide a multi-layered security approach. Furthermore, a significant gap exists in the literature regarding the integration of security measures with digital marketing strategies. Prior research has largely overlooked this intersection. CAML-EHDS addresses this gap by leveraging insights from unified threat detection and cluster-based analysis to optimize digital marketing efforts while ensuring data security compliance. By tailoring targeted advertisements and personalized content based on classified data and detected threats, CAML-EHDS improves customer engagement while adhering to stringent privacy regulations. This integration of security and marketing, coupled with its robust cryptographic and machine learning components, distinguishes CAML-EHDS as a comprehensive and innovative solution, surpassing the limitations of previous models.

3. ARCHITECTURE OF CAML-EHDS MODEL

CAML-EHDS architecture represents a meticulously engineered fortress, designed to provide unparalleled healthcare data security while simultaneously optimizing digital marketing strategies. The process initiates with a fortified data collection and preprocessing phase, ensuring the secure gathering and meticulous preparation of healthcare data, thereby establishing an impregnable foundation for subsequent analyses [12]. Following this, a dual-layered cryptographic shield is deployed, incorporating homomorphic encryption and ECC-based authorization. This robust combination guarantees impenetrable data encryption and enforces stringent, granular access policies, effectively thwarting unauthorized access and mitigating potential breaches. Next, CAML-EHDS employs an advanced cluster-based analysis, leveraging semantic clustering, ranking clusters with precision, computing similarity indices, and executing domain transformations [13]. This sophisticated process uncovers hidden patterns and bolsters classifier security, transforming raw data into actionable intelligence.

The model's analytical prowess is further amplified by a powerful machine learning core. Individual models, including the impressive LSTMs and transfer learning models, undergo rigorous training to classify security threats with unmatched accuracy [14]. A pivotal model fusion stage then integrates the outputs of these models through weighted averaging and ensemble prediction, generating a final, exceptionally robust output. This fusion creates a synergistic defense, exceeding the capabilities of any single model and providing a unified, impenetrable threat detection system. Finally, CAML-EHDS seamlessly integrates digital marketing strategies, leveraging the alarming insights derived from unified threat detection and cluster-based analysis. This integration enables the implementation of targeted advertisements and personalized content, ensuring both marketing effectiveness and unwavering compliance with data security regulations [15]. This comprehensive architecture, with its multi-layered defenses and integrated intelligence, establishes CAML-EHDS as a paragon of robust and secure healthcare data management.

The overall architecture of the CAML-EHDS model is visually summarized in Figure 1. This figure illustrates the end-to-end pipeline of the framework, including data preprocessing, cryptographic methods, cluster-based analysis, machine learning integration, unified threat detection, and digital marketing applications. The diagram highlights how each component interacts to enhance healthcare data security while supporting privacy-compliant marketing strategies.

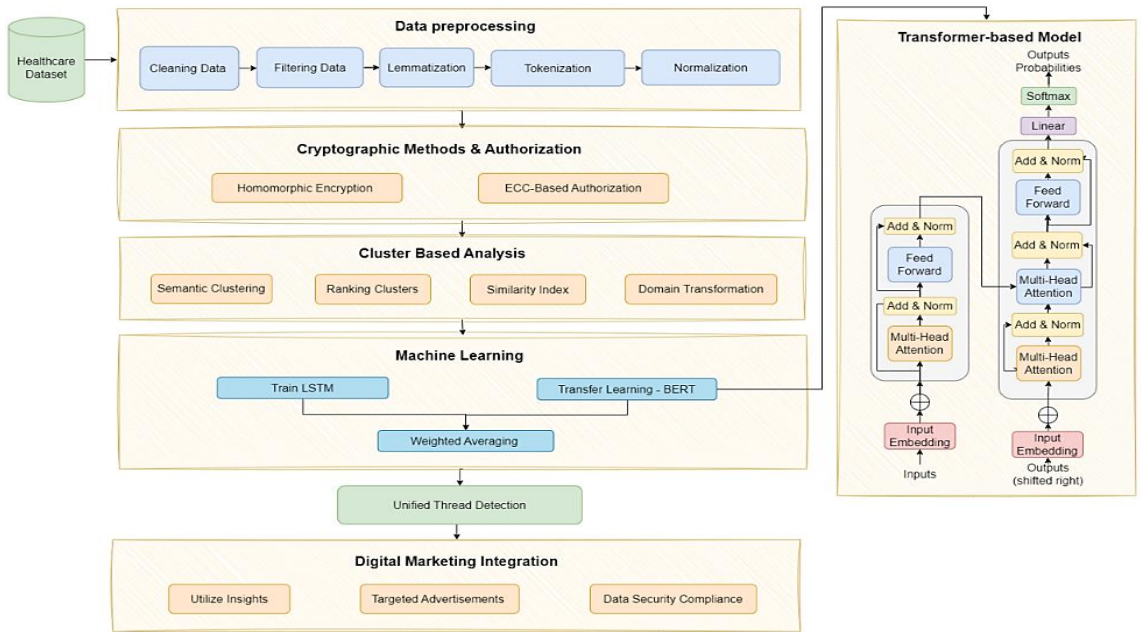

Figure 1. CAML-EHDS model's architecture for securing healthcare data and optimizing digital marketing

4. METHOD

4.1. Data collection and experimental setup

The healthcare data, encompassing patient information on various diseases sourced from healthcare websites, was meticulously gathered. However, a crucial bias analysis revealed potential demographic

overrepresentation within the dataset. We employed data augmentation, fairness-aware machine learning algorithms, and sensitivity analyses, though we acknowledge that inherent biases may persist. Future work will focus on expanding dataset diversity and exploring advanced bias mitigation to ensure fair and generalizable model performance. CAML-EHDS model incorporates strategies to maintain computational efficiency, the framework employs optimized preprocessing techniques to reduce data dimensionality and complexity, and is designed to leverage parallel processing to handle large datasets without significant performance degradation. For the implementation and testing of the research model, Python was selected due to its versatility and the extensive range of libraries [16], including TensorFlow, scikit-learn, NumPy, and Pandas. These libraries are crucial for developing and testing machine learning models. In terms of simulation parameters, symmetric key encryption was employed to secure data during experiments [17], highlighting the critical role of efficient key management in maintaining data security. The dataset consists of medical data with a mean value of 507k and a standard deviation of 12.5k. For health camp IDs, the mean is calculated as 6.57k with a standard deviation of 13.2k. Similarly, for patient data, the mean is 387k with a standard deviation of 39.6k. The overall description of the dataset is detailed in Table 1.

- *https://www.kaggle.com/datasets/mehradaria/covid19-lung-ct-scans;*
- *https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database;*
- *https://www.kaggle.com/datasets/mathchi/diabetes-data-set;*
- *https://www.kaggle.com/competitions/diabetes-classification/data;*
- *https://www.kaggle.com/datasets/saurabh00007/diabetescsv;*
- *https://www.kaggle.com/code/paultimothymooney/predict-diabetes-from-medical-records/data;*
- *https://www.kaggle.com/datasets/kumargh/pimaindiansdiabetescsv;*
- *https://www.kaggle.com/datasets/rischan/diabetes-dataset;*
- *https://www.kaggle.com/datasets/jillanisofttech/diabetes-disease-updated-dataset;*
- *https://www.kaggle.com/code/mathchi/diagnostic-a-patient-has-diabetes/data;*
- *https://www.kaggle.com/datasets/paultimothymooney/blood-cells;*
- *https://www.kaggle.com/datasets/draaslan/blood-cell-detection-dataset;*
- *https://www.kaggle.com/competitions/3md3070-dlmi/data;*

Table 1. Dataset distribution

Dataset	Mean	Standard deviation
Healthcampus	507k	12.5
Covid - 19	387k	39.6
Lung	5856	1.28
Heart	54.4	9.07
Iris	75.5	43.3

4.2. Data cleaning and filtering

Post-data collection, a comprehensive cleaning and filtering protocol was implemented to ensure data integrity and consistency. This protocol encompassed the removal of irrelevant information, error correction, and format standardization. Given the critical nature of missing data in healthcare analytics, a multifaceted imputation strategy was adopted. For numerical variables, such as patient age, mean imputation was utilized to provide statistically representative values. For categorical variables, including patient gender, mode imputation was applied, assigning the most frequent category. In instances where missing data was deemed analytically significant or where simple imputation could introduce substantial bias, k-nearest neighbors (k-NN) imputation was utilized, leveraging similar data point values to estimate missing values. This approach was selected to minimize data loss and preserve dataset integrity, particularly in cases where missing data patterns could yield valuable insights. To address the issue of imbalanced data, where certain security threat categories were less frequent than others, the Synthetic Minority Over-sampling Technique (SMOTE) was subsequently applied. SMOTE was chosen to generate synthetic instances of the minority classes, creating a more balanced dataset for model training. This technique helps prevent the model from being biased towards the majority class and improves its ability to accurately detect rare but critical security threats. Furthermore, textual data underwent tokenization, lowercasing, stop word removal, stemming, and lemmatization [18], preparing it for effective and reliable analysis [19].

4.3. Cryptographic methods and authorization

4.3.1. Cryptographic processes and key management

The CAML-EHDS model implements a robust cryptographic protocol to safeguard healthcare data, featuring key elements such as the data owner (DO), key generation center (KGC), cloud storage (CS),

decryption server (DS), and data receiver (DR). The DO oversees data exchange and encryption for cloud storage, ensuring the secure transfer of patient data. The KGC coordinates key generation and integrates private keys based on user attributes, facilitating encrypted information exchange within the cloud [20]. Serving as a semi-trusted entity, the CS enables data sharing and storage while generating secret keys for users [21]. The DS enables decryption of transmitted information, determining decryption capabilities at the receiver's end [22]. Meanwhile, the DR ensures secure data analysis by integrating attribute sets during decryption and accommodating resource constraints for lightweight mobile devices [23].

4.3.2. Homomorphic encryption of CAML-EHDS model

a. Setup

The CAML-EHDS model utilizes attribute-based ECC to safeguard healthcare data, alongside a key-based approach and a lightweight model tailored for attribute-based ECC processing [24]. The security features encompass domain-specific feature parameters, leveraging an elliptical curve model for estimating and computing public parameters. In this setup, the CS is integrated with the KGC [25], where a random number is computed as $a_i \in Z_q^*$ with the authorization of $i \in \omega_i$, where ω represented as the attributes set for the authorization of $PP = \{a_1, a_2, a_3, \ldots\ldots, a_n\}$, i = 1 to m and $i \in \omega$. With the setup of KGC the secret ley for the master is computed as k ∈ Z*q with ECC for the computation of the public key stated as:

$$PP_{KGC} = k.G \text{ i.e., } \{MK_{KGC} = k, PP_{KGC} = k.G\} \quad (1)$$

CS Setup: The master secret key is elected based on c ∈ Z*q and the public key is estimated as:

$$PP_{CS} = c.G; \{MK_{CS} = c, PP_{CS} = c.G\} \quad (2)$$

The public parameter output is denoted as $params = \{PP, P_{KGS}, PP_{CS}\}$.

b. Encryption and re-encryption

Within DO, data is uploaded for message sharing and execution, using a structure with defined attributes for authorization, denoted as ω. This phase includes computing and estimating the encrypted message m for the data input Λ. The access tree Λ is represented as T, with the message encryption of m using a random number estimated as $s \in Z* q$, for the encryption and integrity of symmetric data computation [26]. The CS execution process involves distributing and storing ciphertext data for the data generated by the DO. The ciphertext data parameters are calculated based on the input data and the CS master key ciphertext [27], with the master key c generated as (3):

$$CS = (T, Cn, Ci, MACm, CmCS = Enc(Cm, cx)) \quad (3)$$

Here $c.G = (cx, cy)$.

c. Key generation, key update and decryption

The key generation phase centers on producing the KGC key K, associated with the attribute set S for the receiver data. The private key for the KGC [28], derived using a random number r ∈ Z*q, is expressed as (4):

$$PKinit = ai.r, \forall i \in S \quad (4)$$

In this equation, the random number generated for the ai takes into account the setup phases. The CAML-EHDS model comprises of three phases such as KGC, CS and DR for the estimation of CS and KGC. The key generation of the components comprises of the following steps that are stated as below:

- Initially, the secret key is generated as k and r with the generation of the secret key as CS represented as c.
- Based on the estimated values of k, r and c the computation process is performed with the information transferred through the CS.
- With the value of reception with the CS the random number is generated $d \in Z * q$ computation of $\left(\frac{x}{d}\right).G$ for the KGC values.
- The KGC values are computed with the estimation of value $B = A.k^2$is conversion of value B within the CS.
- The estimated CS value for the components is denoted as:

$$K' = B.d = A.k^2.d = \left(\frac{x}{d}\right).G.k^2.d = \left(\frac{c}{k} + r\right) * \frac{1}{k}k^2.G = (c + kr).G \quad (5)$$

The decryption process is evaluated through the integration of DS and DR, emphasizing lightweight operations. The key components involved in this process are denoted as $SK_{KGC}, SK_{CS}, SK_{DR}$ corresponding to each key element required for decrypting the DR process.

4.3.3. Attribute-based ECC for authorization

Elliptic curve cryptography (ECC) utilizes specific parameters to define the elliptic curve and the cryptographic operations executed on it. The most commonly employed ECC parameters include the curve equation, the prime modulus, the base point [29], and the order of the base point. The ECC parameters for the commonly used NIST P-256 curve are provided: The curve parameter selected for the analysis is shown in equation:

$$y^2 = x^3 - 3x + b \quad (6)$$

The Prime modulus (field size) with the CAML-EHDS model is presented as

$$p = 2^{256} - 2^{224} + 2^{192} + 2^{96} - 1 \quad (7)$$

Through the equation the coordinates and value b is computed as

$$b = 41058363725152142129326129780047268409114441015993725554835256314039467401291$$

with based generator of G = (x, y) where:

$$x = 48439561293906451759052585252797914202762949526041747995844080717082404635286$$

$$y = 36134250956749795798585127919587881956611106672985015071877198253568414405109$$

The order of pair is computed as:

$$n = 115792089210356248762697446949407573529996955224135760342422259061068512044369$$

These parameters define the elliptic curve and are used in ECC operations like key generation, point multiplication, and digital signatures.

4.4. Cluster-based analysis

The construction of the cluster is assessed by considering the observed semantic domains. By computing the CAML-EHDS model, clusters are ranked based on the estimation of the mean value within the cluster group. The i^{th} cluster relationship is evaluated based on the length of the cluster model in the domain as ($i - 1^{th}$ and $i + 1^{th}$). With computation of the similarity index in the i^{th} cluster is designed with $M_s^{p,q}$. Within the domain of i^{th}cluster with domain p and q values is measured as 1 [30]. Similarly, for the domain p and q the assigned values is stated as 0.5 other it is assigned as the 0. The transformation of the source domain is evaluated by mapping the target function with the latent space d of attacks [31]. Through the conversion of the attacker's domain, the transformation of the latent space is assessed using abundant label instances to classify the healthcare target domain for security. To enhance the security of healthcare data, labelling is applied to the target instances with the training of the classifier [32]. With the proposed model deep learning focused on the assignment of the score to the cluster group for the attack prevention. Initially, each cluster source is assigned as the "normal" or "attacker" with the assigned labels to the cluster. The domain source comprises of the target domain denoted as D_1^t and D_2^t with the Euclidean distance. The label for the source in the i^{th} cluster is ranked as the $r_i, r_i + 1, r_i - 1$ in this model those are labelled as the follows:

- Step 1: Initially, set the value as zero for the label
- Step 2: Upon the ranking of the source cluster r_i and attacker is denoted as α with the elimination of the cluster value.
- Step 3: With the source node cluster is ranked as $r_i + 1$ with the attack denoted as $\frac{\alpha}{2}$ will be included in the cluster else it will be eliminated.
- Step 4: With the source nodes the rank of cluster is stated as $r_i - 1$ and attacker is defined as $\frac{\alpha}{2}$ included within the system else node is eliminated from the cluster group.

Through the estimation of the score target as 0 and 1 the cluster instances are normalized with the normal or attack. With the assigned soft labels the instances for the threshold is T_1 is considered as the attack else the threshold T_2 is considered as the threshold defined as the normal". The instances for the target are defined as: $T_1 = \alpha$ \\ Set as attack label, and $T_2 = 1 - \alpha$ \\ Set as a normal label in this label, the assignment scheme with labelled instances attacks is classified and eliminate the incorporation of the attacks in the network by soft labelling. Within the cluster group, labels are assigned to each cluster, incorporating three components of healthcare data security and classification. The node clusters should include various factors such as prior knowledge, probability of edges, and conditional probability table (CPT) [33]. Our process focuses on estimating network attacks by computing causality and integrating it with the ML-based transfer learning process. The CAML-EHDS process, combined with the transfer learning process for assigning labels and detecting attacks, is described in (8). Through the assigned label instance T_1 and T_2 unknown attacks are computed and estimated with consideration of CPT attacks $U_{ij} = P(X = x_j | U = u_i)$. The process flow of our model for attack detection and prevention is evaluated with the ML model for the training and computation of the trust values in the database.

$$U_{ij}^{t} = \begin{cases} \delta + (1-\delta)U_{ij}^{t-1}, & P(u_i|y_t) = 1 \quad P(u_j|y_t) = 1 \\ (1-\delta)U_{ij}^{t-1}, & P(u_i|y_t) = 1 \quad P(u_j|y_t) = 0 \\ U_{\text{ij}}^{\text{t-1}} & \text{otherwise} \end{cases} \tag{8}$$

4.5. CAML-EHDS algorithm for key management strategies

With the assigned label instances of the attack data eliminated that was identified as $D(D = \{y_1, y_2, y_3 \ldots\ldots\})$ for the attack data estimation denoted as y_t. The CAML-EHDS attack scenario is estimated as $S = (I_1, I_2, \ldots\ldots I_n)$ with the assigned label of ML based deep learning model for the estimation of the attacks. The model attacks for the estimation of the variables are computed for our model is presented in Algorithm 1.

Algorithm 1. Parameter estimation

```
Input: Network Attack = {(al1, al2.....)(al3, al4.....)....}
Output: δ^(n+1) = (C^(n+1), J^(n+1), λ^(n+1))
// Start
For n = 0 estimate δ_i^0
For a_ij^0 = δ_i^0 set b_j(k)^0
Compute the attacks those are unknown as n =0,1,2...
   do
Compute using (7)
Compute using (8)
     End for
   End for
Set the values for estimation
Set values for the comparison
   If P(Z_i = 1|I_i = 1) > trustValue
     then
Calculate the T1 and T2 based estimated values
   End if
       for value(Z_i) set as the attack value
       If value(Z_i) > trustedevaluethen
      Calculate the set Z_i
      End if
     End for
  End for
```

The ML framework focuses on generating source mappings and constructing the target domain within the latent space. Upon converting the latent space, the source domain consists of probable instance labels for attack classification. The accuracy of the training label classifier for the targeted instances is evaluated using the assigned soft labels. The CAML-EHDS solution involves generating, distributing, and updating encryption keys for various entities, including the data owner, the cloud server, and the data recipient. Here are some considerations for key management in this environment:

- Key generation: Encryption keys are paramount for safeguarding healthcare data [34]. Secure methods such as reliable random number generators or trusted key management systems are essential for their generation [35]. Furthermore, these keys must possess adequate strength to withstand brute-force attacks and adhere to recommended key size guidelines specified for the encryption algorithm in use [36].

- Key distribution: Once encryption keys are generated, secure and reliable methods are imperative for their distribution to designated entities [37]. This process often involves the utilization of secure channels such as encrypted email or secure file transfer protocols [38]. Ensuring the secure transmission and protection of keys during distribution is critical to prevent unauthorized access [39].
- Key updates: In the dynamic healthcare landscape, regular updates to encryption keys may become necessary due to various factors [40]. These factors include key expiration, compromised keys, or changes in user access permissions. A well-defined process must manage key updates effectively, incorporating mechanisms for revoking and replacing keys as required [41].
- Access control: Proper access control mechanisms are vital to restrict access to encryption keys to only authorized entities [42]. This may entail implementing role-based access control, cryptographic key management systems, or other access control policies to safeguard sensitive information from unauthorized access [43].
- Key storage: Secure storage of encryption keys is crucial to prevent unauthorized access and potential breaches [44] utilizing hardware security modules (HSMs) or other secure storage solutions can help safeguard keys from both physical and logical attacks, thereby enhancing overall security [45].
- Key backup and recovery: Regular backups of encryption keys are necessary to mitigate the risk of data loss in the event of key compromise or system failures [46]. Establishing a robust key recovery process is vital to restore access to encrypted data promptly if keys are lost or become inaccessible [47].
- Compliance and auditing: Key management processes must comply with relevant regulatory requirements, such as health insurance portability and accountability act (HIPAA) for healthcare data [48]. Regular audits and continuous monitoring should be conducted to ensure compliance and identify any potential vulnerabilities in the key management system [49].

Implementing a comprehensive key management strategy is essential for maintaining the security and confidentiality of healthcare data in a dynamic environment [50]. Consulting with security experts and adhering to industry best practices is recommended to design and implement an effective key management system that meets the specific security needs of healthcare organizations [51].

4.6. Machine learning model

4.6.1. Long short-term memory networks (LSTMs)

LSTMs are employed to analyze the inherent temporal dependencies within patient records, essential for detecting evolving security threats that manifest over time [52]. To analyze the temporal dependencies inherent in-patient records, crucial for detecting evolving security threats, long short-term memory networks (LSTMs) were selected for their optimized ability to process sequential time-series data, a common format in electronic health records. Unlike transformers, which excel at capturing long-range dependencies across entire sequences but are computationally intensive, or convolutional neural networks (CNNs), which are effective for spatial data but less suited for temporal patterns, LSTMs offer a balance of efficiency and effectiveness in identifying subtle anomalies and patterns that emerge over time. Their recurrent architecture allows them to maintain memory across sequences, enabling the detection of threats that manifest as changes in patient data over extended periods, making them a more practical and efficient choice for this specific application.

For our LSTM implementation, hyperparameters were meticulously selected through a combination of grid search and validation set performance evaluation [53]. We utilized a multi-layered LSTM architecture with 128 hidden units per layer, determined to balance model complexity and computational efficiency. The sequence length was set to 50, exhibited efficient memory usage, requiring approximately 4 GB of GPU memory during training. This configuration resulted in an average training time of 3 hours on our dataset. The Adam optimizer was chosen with a learning rate of 0.001, and batch size was set to 32, values determined through grid search to optimize convergence and prevent overfitting.

4.6.2. Transfer learning with pre-trained models

Complementing LSTMs, we leverage transfer learning with BERT to improve our model's ability to understand the semantic context of healthcare data. BERT, pre-trained on vast amounts of text, excels in capturing complex relationships between words and phrases [54]. This allows for the detection of subtle semantic anomalies that may indicate unauthorized access or data manipulation. While LSTMs are optimized for temporal analysis, BERT provides a deep semantic understanding, allowing us to capture different threat vectors [55]. By combining LSTMs for temporal pattern recognition and BERT for semantic understanding, our model achieves a comprehensive analysis of healthcare data, addressing both the sequential nature and the complex semantic content of the information [56]. This hybrid approach optimizes threat detection by leveraging the strengths of both recurrent and transformer-based architectures, increasing the overall security of healthcare data ecosystems.

4.6.3. Model fusion: amplifying detection capabilities

As the individual models emerge from the container of training, they converge harmoniously in the fusion phase, forging an alliance that transcends the capabilities of any single model. Here, the collective intelligence of LSTMs and transfer learning models combine, birthing a hybrid fusion approach prepared to redefine healthcare security threat detection. Employing sophisticated fusion techniques, predictions from individual models are combined to form a robust ensemble. Using the weighted averaging method, the fusion approach creates a final ensemble prediction enriched with collective wisdom. In this process, each model's prediction is assigned a weight based on its performance and reliability [57]. These weighted scores are then averaged to produce a unified prediction. This synthesis transcends the limitations of individual models by leveraging their diverse strengths. By carefully assigning weights, the fusion method ensures that the most accurate and reliable models have a greater influence on the final prediction [58]. This approach improves the overall resilience and efficacy of the healthcare security system, providing a comprehensive defense against potential threats.

4.7. Digital marketing integration

CAML-EHDS uniquely integrates digital marketing strategies with robust data security, leveraging insights from unified threat detection and cluster-based analysis. This enables healthcare organizations to develop targeted advertisements and personalized content based on classified data and detected threats, improving customer engagement and marketing effectiveness. By employing advanced cryptographic techniques like homomorphic encryption and ECC, CAML-EHDS ensures that marketing data remains secure and compliant with privacy regulations. However, this integration necessitates careful consideration of ethical concerns. Specifically, the use of sensitive healthcare data for marketing purposes raises questions about informed consent, data anonymization, and the potential for discriminatory targeting. To mitigate these risks, CAML-EHDS incorporates mechanisms for transparent data usage, robust anonymization techniques, and strict adherence to privacy regulations.

5. RESULTS

5.1. Evaluation of healthcare data security using CAML-EHDS model

Using the proposed CAML-EHDS techniques, healthcare data security features are assessed with machine learning, focusing on three different metrics: authentication, encryption, and machine learning. The model includes a cryptographic process that is examined considering various features for security, communication overhead, and computation process. The homomorphic encryption scheme is evaluated using the attribute-based escrow model for analysis.

In the proposed model, homomorphic encryption is used to store electronic medical records on the escrow server. This encryption method is applied to the medical healthcare records. The examined results for the constructed model are presented in Figure 2. In Figure 3, ECC-based authorization is conducted for the evaluation and computation of medical data. The examination involves authorizing users of medical healthcare data. Computed authorization using ECC is implemented in the cloud to enhance security.

5.2. Security features

The CAML-EHDS model is designed to boost the security of healthcare data by incorporating various advanced security features. These features are evaluated and compared against existing models to highlight the effectiveness of our model in safeguarding sensitive information. The key security features assessed include pairing-free operations, ECC based methods, key-escrow mechanisms, resistance to collusion attacks, provable security, and key authority management. Table 2 presents a comparative analysis of these security features across different schemes.

The performance of the CAML-EHDS model is compared with existing schemes such as GHZ, J, HZ, XZY, and SCH. The comparative analysis focuses on how each scheme handles the security features. CAML-EHDS model excels in all categories, demonstrating its superiority in providing comprehensive security for healthcare data.

- GHZ [14] and J [15] schemes lack pairing-free operations and ECC-based methods, which are crucial for efficient and secure data processing in resource-constrained environments.
- HZ [34] provides pairing-free operations but does not include ECC-based methods or key-escrow mechanisms, limiting its flexibility and security.
- XZY [7] and SCH [8] incorporate both pairing-free and ECC-based methods but lack key-escrow features, reducing their effectiveness in key management and recovery scenarios.

The proposed model not only addresses these shortcomings but also introduces a lightweight key-escrow scheme and robust key authority management, making it a well-rounded solution for securing

healthcare data. Incorporating advanced cryptographic techniques and robust key management strategies, our model is the best performing model for protecting sensitive healthcare information. The evaluation and comparative analysis demonstrate its effectiveness in mitigating various security threats, ensuring the confidentiality, integrity, and availability of healthcare data.

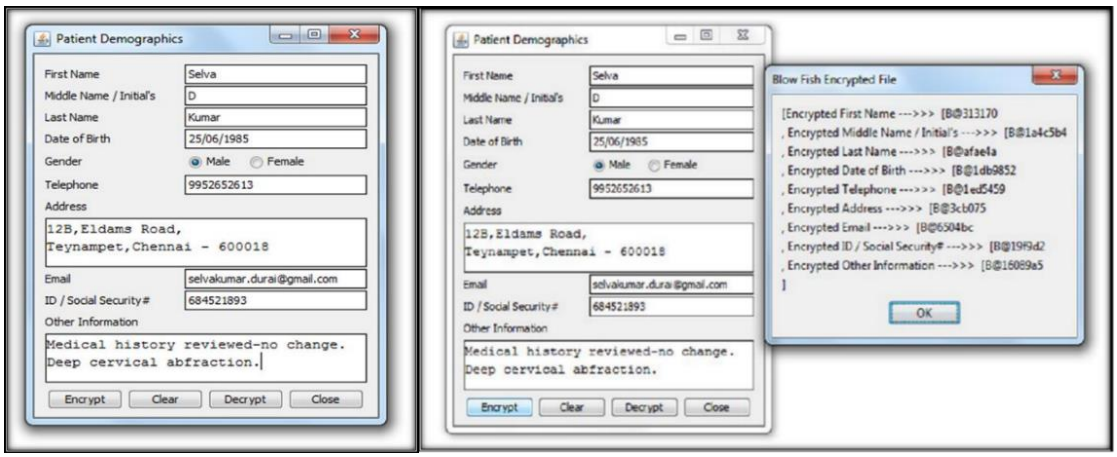

Figure 2. Medical records and encryption with homomorphic process

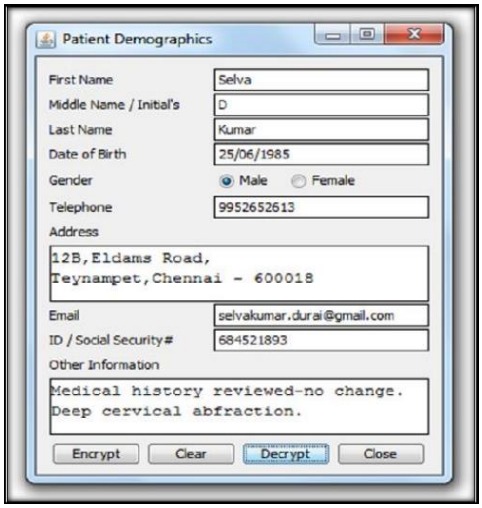

Figure 3. Patient demographic attributes for authorization

Table 2. Comparison of security features in CAML-EHDS

Scheme	Pairing – free	ECC based	Key – escrow	Collusion attack	Provable secured	Key authority
GHZ [14]	No	No	Yes	Yes	Yes	Yes
J [15]	No	No	Yes	Yes	No	Yes
HZ [34]	Yes	No	No	No	Yes	No
XZY [7]	Yes	Yes	No	No	Yes	No
SCH [8]	Yes	Yes	No	Yes	Yes	No
Proposed CAML-EHDS	Yes	Yes	Yes	Yes	Yes	Yes

5.3. Performance analysis of node configuration in encrypted systems

Table 3 provides a comparative analysis of encryption time, loss percentage, and accuracy percentage for different numbers of nodes in a system. The CAML-EHDS model demonstrates significant strengths in handling encryption and maintaining high accuracy in healthcare data security. The results show that with a low number of nodes, our model achieves exceptionally high accuracy, with 98% at 2 nodes, and a minimal loss percentage of 13%.

Table 3. Performance of CAML-EHDS

No. of nodes	Encryption time(s)	Loss%	Accuracy%
2	12	13	98
4	15	18	97
6	18	22	96
8	21	26	95
10	24	29	94

Although encryption time and information loss increase as the number of nodes rises, the model still maintains a commendable accuracy of 94% even at 10 nodes. This illustrates this research model's robustness in processing and encrypting data across varying node configurations while sustaining high accuracy levels. Despite the expected trade-offs in encryption time and data loss, CAML-EHDS proves to be highly effective and reliable in a cloud server environment for healthcare data processing, ensuring both security and performance as illustrated in Figure 4.

Figure 4. Comparison of encryption time, loss and accuracy

5.4. Computation overhead

Computation overhead covers the series of operations involved in encryption, decryption, key generation, and related tasks. The estimated balance of computation overhead includes operations such as bilinear pairing, exponentiation, point hashing, and scalar multiplication based on points. It also includes arithmetic and logical calculations. When performing authorization tasks using ECC operations for groups in the bilinear group, these computations are utilized $\times\ G1 \rightarrow G2$.

Table 4 in this paper details the basic modular operations in ECC that estimate attributed scalar multiplications. Scalar multiplication is critical in ECC-based techniques and significantly impacts computation overhead. The attribute-based ECC scheme processes scalar points with multiplication, making it essential to optimize these operations for effective healthcare data analysis, especially in IoT-based environments. The findings highlight the importance of minimal computation overhead to ensure efficient and secure data handling.

Table 4. Computational overhead comparison of security schemes

Scheme	Initialization	Encryption	Key generation	Decryption	Total
GHZ [14]	$P + 2E \approx 24S$	$P + (3 + l)E \approx 46S$	$(10 + 4u)E \approx 60S$	$3P \approx 60S$	190S
J [15]	$P + 3E \approx 26S$	$P + (3 + l)E \approx 46S$	$(4 + 2u)E \approx 28S$	$3P \approx 60S$	160S
XZY [7]	$(n + 1)S \approx 31S$	$(l + 1)S \approx 11S$	-	$(u + 1)S \approx 6S$	48S
SCH [8]	$(n + 1)S \approx 31S$	$(3l + 1)S \approx 31S$	-	$(3 + u)S \approx 8S$	70S
CAML-EHDS	4S	$(4 + l)S \approx 12S$	8S	$(2 + u)S \approx 7S$	31S

The comparison of computation overhead among different schemes reveals insightful findings. GHZ [14] exhibits the highest total overhead (190S), primarily attributed to its extensive key generation and decryption steps. Although J [15] demonstrates a lower total overhead (160S), its significant encryption and decryption processes still contribute to computational load. XZY [7] achieves a notably lower total overhead (48S) with minimal encryption and decryption requirements, indicating its efficiency. SCH [8] presents a moderate total overhead (70S) with a balanced approach to encryption and decryption steps. In contrast, the CAML-EHDS model demonstrates the lowest total overhead (31S) due to its efficient initialization,

encryption, key generation, and decryption processes. This emphasizes the model effectiveness in healthcare data analysis in IoT environments, ensuring both high security and low computational resource consumption, thereby making it a promising solution for secure healthcare data management.

5.5. CAML-EHDS performance with varied memory sizes

In this section, we explore the performance of various healthcare data security categories concerning different memory sizes. The simulation analysis of the proposed model is evaluated for healthcare data of different file sizes, ranging from 500 MB to 4 GB. To assess real-time usability, the model was tested under simulated high-traffic conditions with data input rates ranging from 100 to 1000 requests per second. In Figure 5, the x-axis represents distinct categories: denial of service (DoS), user-to-root (U2R), remote-to-local (R2L), Probe, unknown, and CAML-EHDS. Each category corresponds to specific activities or behaviors within healthcare data security. On the y-axis, values range between approximately 1.68 to 1.88, reflecting memory sizes of 500 MB, 1 GB, 2 GB, 3 GB, and 4 GB, with each size represented by unique line styles and colors. The observations reveal intriguing insights into each category's behavior across varying memory capacities. DoS consistently registers the highest values across all memory sizes, indicating a persistent danger that remains relatively stable even with increased memory.

U2R and R2L categories exhibit lower values compared to DoS but display slight upward trends with larger memory sizes, indicating potential vulnerabilities in these areas. Probe values, while relatively stable across memory sizes, remain lower than DoS but higher than U2R and R2L, suggesting a moderate level of risk. Unknown category values remain consistent and lower than Probe. Notably, this model values consistently rank the lowest across all categories, demonstrating its efficacy in preventing security breaches, with a slight decrease observed with larger memory sizes. Overall, the graph suggests that while the CAML-EHDS model excels in security breach prevention, DoS attacks pose a persistent threat, highlighting the importance of robust defense mechanisms. Additionally, the trend implies that larger memory sizes may offer performance improvements in certain categories, warranting further investigation and optimization strategies.

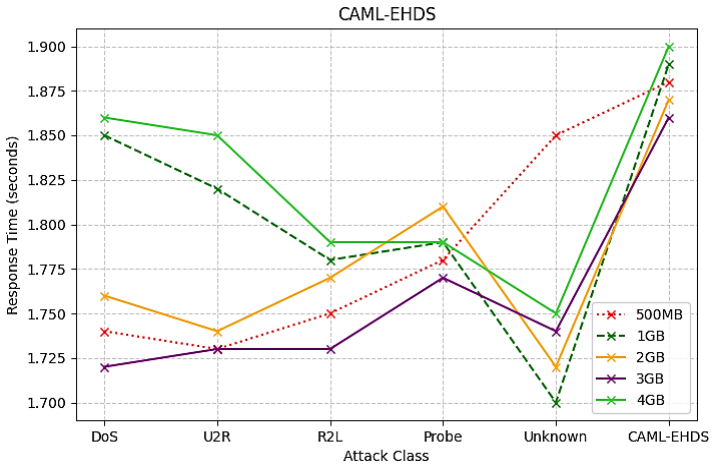

Figure 5. Analysis of different attacks vs CAML-EHDS

5.6. Comparative analysis of machine learning models

To evaluate the efficacy of our proposed CAML-EHDS model, we compared its performance against traditional classification methods: support vector machine (SVM), random forest (RF), and decision tree (DT). The performance of all models was assessed using key metrics: sensitivity, specificity, and accuracy. The results of this comparison are presented in Table 5. Additionally, as illustrated in Figure 6, CAML-EHDS demonstrates the highest values across all performance metrics, indicating superior performance compared to the other evaluated methods. Specifically, CAML-EHDS achieved a sensitivity of 0.85, a specificity of 0.997, and an accuracy of 0.96. To provide a measure of the precision and reliability of these estimates, we calculated 95% confidence intervals: accuracy [0.94-0.98], sensitivity [0.82-0.88], and specificity [0.995-0.999]. These findings suggest that our model is particularly effective for the classification task at hand, offering enhanced capabilities in correctly identifying both positive and negative instances within the dataset.

While traditional models like RF (accuracy: 0.89) and DT (accuracy: 0.93) provide competitive accuracy, they present real-world implementation challenges for privacy-sensitive healthcare applications.

Their inability to directly process encrypted data necessitates expensive and complex pre-processing and secure computation methods, creating significant hurdles in terms of operational costs and adherence to strict data privacy regulations. Beyond overall classification performance, a detailed error analysis revealed that CAML-EHDS demonstrated a notably lower false positive rate compared to SVM (accuracy: 0.82) and RF, particularly in the classification of normal and benign instances. However, a higher number of false negatives was observed in underrepresented attack classes such as R2L and U2R, suggesting that while the model is efficient overall, it may require further optimization for rare event detection. This highlights the importance of continued work on balancing detection sensitivity across all classes, especially in security-critical applications.

Table 5. Comparison of classification models performance

Methods	Sensitivity	Specificity	Accuracy
SVM	0.2	0.95	0.82
RF	0.8	0.984	0.89
DT	0.75	0.983	0.93
CAML-EHDS	0.85	0.997	0.96

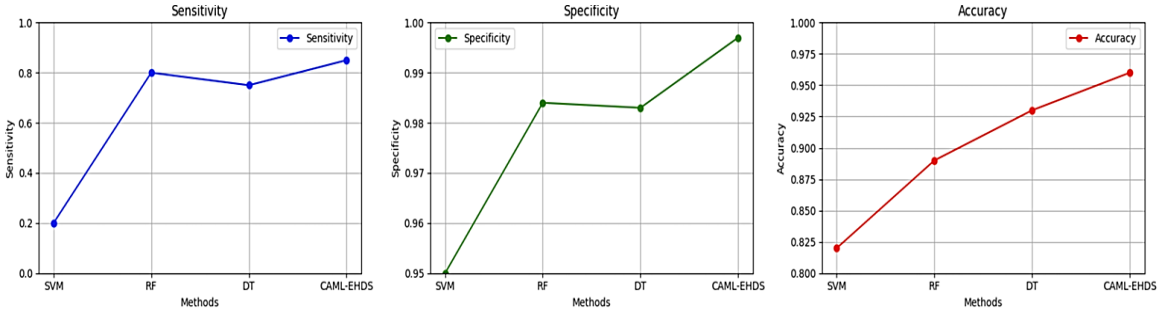

Figure 6. Performance metrics of ML models for CAML-EHDS

6. DISCUSSION

The integration of security measures into digital marketing strategies within healthcare organizations has been a significantly under-explored area in scholarly research. However, CAML-EHDS, as presented in this study, effectively bridges this gap. By leveraging insights from unified threat detection and cluster-based analysis, CAML-EHDS optimizes digital marketing efforts while ensuring robust data security. This novel approach, which utilizes tailored advertisements and personalized content derived from classified data and detected threats, significantly improves customer engagement while maintaining strict adherence to data security regulations. This includes compliance with regulatory frameworks such as the general data protection regulation (GDPR) and the health insurance portability and accountability act (HIPAA), by ensuring that personal health data is encrypted, access-controlled, and processed without compromising user privacy. Building upon existing cryptographic, clustering, and machine learning techniques, CAML-EHDS offers a comprehensive solution that surpasses the limitations of previous models.

As illustrated in Table 6, CAML-EHDS achieves a superior accuracy of 96%, outperforming SVM (82%), RF (89%), and DT (93%). This demonstrates CAML-EHDS's developed capability in accurately classifying both positive and negative instances, crucial for robust security threat detection. Furthermore, CAML-EHDS exhibits the lowest total computational overhead (31S) compared to GHZ, J, XZY, and SCH. This low overhead, combined with high accuracy, signifies the model's efficiency and reliability in processing and encrypting healthcare data. The model's robustness is further demonstrated, showcasing its ability to handle increased system complexity without significant performance degradation. Additionally, the model's ability to maintain high accuracy and low overhead, even when tested with large files sizes ranging from 500MB to 4GB, and under various attack scenarios, shows its incredible strength.

CAML-EHDS distinguishes itself by achieving a 96% threat detection accuracy, a notable improvement over contemporary healthcare security study. While Bercea *et al.* reported a 92% anomaly detection accuracy using federated learning [59], and Bilot *et al.* achieved 94% in intrusion detection with graph neural networks [60], CAML-EHDS's integrated cryptographic, clustering, and machine learning approach demonstrates its superior efficacy. Similarly, Ali *et al.* [61] and Hamid *et al.* [62] reported accuracies of 93% and 91% respectively, utilizing hybrid cryptographic machine learning systems and data mining for fraud detection. This superior performance underscores CAML-EHDS's ability to seamlessly

combine diverse security mechanisms for a more robust and accurate healthcare data protection framework. CAML-EHDS is a standout solution in the field, offering a comprehensive and efficient approach to safeguarding healthcare data while optimizing digital marketing strategies. Its advanced cryptographic, clustering, and machine learning techniques, coupled with low computational overhead and high accuracy, establish it as a leading model in the evolving landscape of healthcare data security.

7. CONCLUSION

In conclusion, CAML-EHDS presents a robust framework designed to revolutionize healthcare data security and digital marketing integration, addressing the critical need for safeguarding sensitive patient information while optimizing marketing strategies. By integrating advanced cryptographic techniques, cluster-based analysis, and machine learning algorithms, CAML-EHDS ensures data confidentiality, integrity, and availability, effectively mitigating unauthorized access. Its unique ability to seamlessly integrate digital marketing with stringent security protocols allows healthcare organizations to tailor marketing efforts for developed customer engagement while maintaining regulatory compliance. Looking ahead, several avenues for future work will improve CAML-EHDS's applicability and impact. Firstly, to facilitate industry adoption, a detailed integration roadmap will be developed, outlining step-by-step procedures for businesses to incorporate CAML-EHDS into existing systems. This roadmap will include API specifications, deployment guidelines, and case studies demonstrating successful implementation in various healthcare settings.

Secondly, to address ethical concerns, future research will focus on implementing mechanisms to prevent misuse in marketing. This includes developing robust auditing tools to monitor data usage, ensuring transparency in marketing practices, and establishing clear guidelines for data anonymization and consent management. Furthermore, a comprehensive robustness evaluation will be conducted to assess the model's performance under adversarial attacks. This evaluation will involve simulating various attack scenarios, including data poisoning, model evasion, and privacy breaches, to quantify the model's resilience and identify potential vulnerabilities. Techniques such as adversarial training and robust optimization will be explored to enhance the model's defense mechanisms. Additionally, continued research will focus on optimizing the framework's performance and scalability, exploring new cryptographic methods, refining clustering algorithms, and improving machine learning models to better detect emerging security threats. Ongoing collaboration with healthcare practitioners and industry stakeholders will be crucial for validating the effectiveness of CAML-EHDS settings and ensuring its seamless integration into existing healthcare systems. By remaining committed to innovation, ethical considerations, and robust evaluation, we can continue to advance the field of healthcare data security and digital marketing integration, ultimately improving patient outcomes and driving positive change in the healthcare industry.

FUNDING INFORMATION

The authors received no external funding for this research. The work was carried out using the authors' institutional and personal resources.

AUTHOR CONTRIBUTIONS STATEMENT

Abderrahmane Fathi (corresponding author): Conceptualization; Methodology; Software; Validation; Formal analysis; Investigation; Data curation; Writing – original draft; Writing – review & editing; Visualization; Project administration. Kawtar Mouyassir: Conceptualization; Methodology; Validation; Formal analysis; Investigation; Data curation; Writing – original draft; Writing – review & editing; Visualization. Waqas Ali: Formal analysis; Validation; Investigation; Writing – review & editing. Fatima Zahra Fandi: Formal analysis; Validation; Writing – review & editing. Ali Kartit: Methodology; Resources; Supervision; Project administration, Review. All authors read and approved the final manuscript. Abderrahmane Fathi is the corresponding author and is responsible for all correspondence during submission, revision, and publication.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Abderrahmane Fathi	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
Kawtar Mouyassir		✓		✓	✓	✓	✓	✓	✓	✓	✓		✓	
Waqas Ali					✓		✓		✓	✓				
Fatima Zahra Fandi					✓		✓			✓				
Ali Kartit		✓		✓	✓		✓			✓	✓	✓	✓	✓

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

The authors declare no conflict of interest. The authors have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

INFORMED CONSENT

This study used only publicly available, de-identified datasets obtained from Kaggle (see Data Availability). No direct interaction with human subjects occurred, and no identifiable personal information was processed.

ETHICAL APPROVAL

In accordance with institutional and national guidelines, analyses based solely on public, de-identified secondary datasets do not constitute human-subjects research and are exempt from IRB/ethics review. No patient recruitment, intervention, or linkage to identifiable records was performed.

DATA AVAILABILITY

All datasets used in this study are publicly available on Kaggle at the links listed below. We accessed each dataset on 10 Oct. 2025 and complied with the respective data-use licenses/terms. No attempt was made to re-identify individuals.

- COVID-19 Lung CT Scans — *https://www.kaggle.com/datasets/mehradaria/covid19-lung-ct-scans*
- Pima Indians Diabetes Database — *https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database*
- Diabetes Data Set (Mathchi) — *https://www.kaggle.com/datasets/mathchi/diabetes-data-set*
- Diabetes Classification (competition data) — *https://www.kaggle.com/competitions/diabetes-classification/data*
- Diabetes CSV (Saurabh) — *https://www.kaggle.com/datasets/saurabh00007/diabetescsv*
- Predict Diabetes from Medical Records (Mooney) — *https://www.kaggle.com/code/paultimothymooney/predict-diabetes-from-medical-records/data*
- Pima Indians Diabetes CSV (kumargh) — *https://www.kaggle.com/datasets/kumargh/pimaindiansdiabetescsv*
- Diabetes Dataset (rischan) — *https://www.kaggle.com/datasets/rischan/diabetes-dataset*
- Diabetes Disease Updated Dataset — *https://www.kaggle.com/datasets/jillanisofttech/diabetes-disease-updated-dataset*
- Diagnostic: A Patient Has Diabetes (Mathchi) — *https://www.kaggle.com/code/mathchi/diagnostic-a-patient-has-diabetes/data*
- Blood Cells (Paul Mooney) — *https://www.kaggle.com/datasets/paultimothymooney/blood-cells*
- Blood-Cell Detection Dataset — *https://www.kaggle.com/datasets/draaslan/blood-cell-detection-dataset*
- 3MD3070 DLMI (competition data) — *https://www.kaggle.com/competitions/3md3070-dlmi/data*.

REFERENCES

[1] A. Flamini, G. Sciarretta, M. Scuro, A. Sharif, A. Tomasi, and S. Ranise, "On cryptographic mechanisms for the selective disclosure of verifiable credentials," *Journal of Information Security and Applications*, vol. 83, p. 103789, Jun. 2024, doi: 10.1016/j.jisa.2024.103789.

[2] X.-Q. Cai, Z.-F. Liu, and T. Wang, "Measurement-device-independent quantum homomorphic encryption," *Physics Letters A*, vol. 513, p. 129609, Jul. 2024, doi: 10.1016/j.physleta.2024.129609.

[3] N. Faruqui, M. A. Yousuf, F. A. Kateb, M. Abdul Hamid, and M. M. Monowar, "Healthcare as a service (HAAS): CNN-based cloud computing model for ubiquitous access to lung cancer diagnosis," *Heliyon*, vol. 9, no. 11, p. e21520, Nov. 2023, doi: 10.1016/j.heliyon.2023.e21520.

[4] A. Guni, P. Normahani, A. Davies, and U. Jaffer, "Harnessing machine learning to personalize web-based health care content," *Journal of Medical Internet Research*, vol. 23, no. 10, p. e25497, Oct. 2021, doi: 10.2196/25497.

[5] A. Ferri *et al.*, "The HIBAD experience: using digital health technologies in the GDPR era," *Health Policy and Technology*, vol. 12, no. 4, p. 100788, Dec. 2023, doi: 10.1016/j.hlpt.2023.100788.

[6] A. Acar, H. Aksu, A. S. Uluagac, and M. Conti, "A survey on homomorphic encryption schemes," *ACM Computing Surveys*, vol. 51, no. 4, pp. 1–35, Jul. 2019, doi: 10.1145/3214303.
[7] R. Imam *et al.*, "A systematic literature review of attribute based encryption in health services," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 9, pp. 6743–6774, Oct. 2022, doi: 10.1016/j.jksuci.2022.06.018.
[8] S. Festag and C. Spreckelsen, "Semantic anomaly detection in medical time series," 2021, doi: 10.3233/SHTI210059.
[9] A. Prasad, W. Mohammad Alenazy, N. Ahmad, G. Ali, H. A. Abdallah, and S. Ahmad, "Optimizing IoT intrusion detection with cosine similarity based dataset balancing and hybrid deep learning," *Scientific Reports*, vol. 15, no. 1, p. 30939, Aug. 2025, doi: 10.1038/s41598-025-15631-3.
[10] G. Balhareth and M. Ilyas, "Optimized intrusion detection for IoMT networks with tree-based machine learning and filter-based feature selection," *Sensors*, vol. 24, no. 17, p. 5712, Sep. 2024, doi: 10.3390/s24175712.
[11] A. Rajkomar *et al.*, "Scalable and accurate deep learning with electronic health records," *npj Digital Medicine*, vol. 1, no. 1, p. 18, May 2018, doi: 10.1038/s41746-018-0029-1.
[12] A. K. Conduah, S. Ofoe, and D. Siaw-Marfo, "Data privacy in healthcare: Global challenges and solutions," *DIGITAL HEALTH*, vol. 11, May 2025, doi: 10.1177/20552076251343959.
[13] Y. Zhou and M. G. Varzaneh, "Efficient and scalable patients clustering based on medical big data in cloud platform," *Journal of Cloud Computing*, vol. 11, no. 1, p. 49, Sep. 2022, doi: 10.1186/s13677-022-00324-3.
[14] M. Alalhareth and S.-C. Hong, "Enhancing the internet of medical things (IoMT) security with meta-learning: a performance-driven approach for ensemble intrusion detection systems," *Sensors*, vol. 24, no. 11, p. 3519, May 2024, doi: 10.3390/s24113519.
[15] D. McGraw and K. D. Mandl, "Privacy protections to encourage use of health-relevant digital data in a learning health system," *npj Digital Medicine*, vol. 4, no. 1, p. 2, Jan. 2021, doi: 10.1038/s41746-020-00362-8.
[16] P. Virtanen *et al.*, "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, doi: 10.1038/s41592-019-0686-2.
[17] E. Barker, "Recommendation for key management: Part 1 – General (SP 800-57 Part 1 Rev. 5)." National Institute of Standards and Technology, Gaithersburg, MD, May 2020, doi: 10.6028/NIST.SP.800-57pt1r5.
[18] M. Siino, I. Tinnirello, and M. La Cascia, "Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers," *Information Systems*, vol. 121, p. 102342, Mar. 2024, doi: 10.1016/j.is.2023.102342.
[19] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, Dec. 2019, doi: 10.1186/s40537-019-0197-0.
[20] M. Marwan, A. Kartit, and H. Ouahmane, "A cloud-based framework to secure medical image processing," *Journal of Mobile Multimedia*, vol. 14, no. 3, pp. 319–344, 2018, doi: 10.13052/jmm1550-4646.1434.
[21] S. Rana, F. K. Parast, B. Kelly, Y. Wang, and K. B. Kent, "A comprehensive survey of cryptography key management systems," *Journal of Information Security and Applications*, vol. 78, p. 103607, Nov. 2023, doi: 10.1016/j.jisa.2023.103607.
[22] H. Wang, J. Liang, Y. Ding, S. Tang, and Y. Wang, "Ciphertext-policy attribute-based encryption supporting policy-hiding and cloud auditing in smart health," *Computer Standards & Interfaces*, vol. 84, p. 103696, Mar. 2023, doi: 10.1016/j.csi.2022.103696.
[23] J. Li, Y. Zhang, X. Chen, and Y. Xiang, "Secure attribute-based data sharing for resource-limited users in cloud computing," *Computers & Security*, vol. 72, pp. 1–12, Jan. 2018, doi: 10.1016/j.cose.2017.08.007.
[24] Y.-W. Hwang and I.-Y. Lee, "A study on CP-ABE-based medical data sharing system with key abuse prevention and verifiable outsourcing in the IoMT environment," *Sensors*, vol. 20, no. 17, p. 4934, Aug. 2020, doi: 10.3390/s20174934.
[25] S. Shukla and S. Patel, "A novel pairing-free ECC-based ciphertext-policy attribute-based proxy re-encryption for secure cloud storage," in *Proceedings of the 11th International Conference on Information Systems Security and Privacy*, 2025, pp. 225–233, doi: 10.5220/0013138600003899.
[26] E. Chen, Y. Zhu, G. Zhu, K. Liang, and R. Feng, "How to implement secure cloud file sharing using optimized attribute-based access control with small policy matrix and minimized cumulative errors," *Computers & Security*, vol. 107, p. 102318, Aug. 2021, doi: 10.1016/j.cose.2021.102318.
[27] M. Marwan, A. Kartit, and H. Ouahmane, "Applying secure multi-party computation to improve collaboration in healthcare cloud," in *Proceedings - 2016 3rd International Conference on Systems of Collaboration, SysCo 2016*, 2017, p. 7831325, doi: 10.1109/SYSCO.2016.7831325.
[28] H. Cui, R. H. Deng, B. Qin, and J. Weng, "Key regeneration-free ciphertext-policy attribute-based encryption and its application," *Information Sciences*, vol. 517, pp. 217–229, May 2020, doi: 10.1016/j.ins.2019.12.025.
[29] H. Marzouqi, M. Al-Qutayri, and K. Salah, "Review of elliptic curve cryptography processor designs," *Microprocessors and Microsystems*, vol. 39, no. 2, pp. 97–112, Mar. 2015, doi: 10.1016/j.micpro.2015.02.003.
[30] A. N. Albatineh, "Means and variances for a family of similarity indices used in cluster analysis," *Journal of Statistical Planning and Inference*, vol. 140, no. 10, pp. 2828–2838, Oct. 2010, doi: 10.1016/j.jspi.2010.03.005.
[31] A. Choudhary, L. Tong, Y. Zhu, and M. D. Wang, "Advancing medical imaging informatics by deep learning-based domain adaptation," *Yearbook of Medical Informatics*, vol. 29, no. 01, pp. 129–138, Aug. 2020, doi: 10.1055/s-0040-1702009.
[32] S. Baek, D. Kwon, S. C. Suh, H. Kim, I. Kim, and J. Kim, "Clustering-based label estimation for network anomaly detection," *Digital Communications and Networks*, vol. 7, no. 1, pp. 37–44, Feb. 2021, doi: 10.1016/j.dcan.2020.06.001.
[33] T. Alsolami, B. Alsharif, and M. Ilyas, "Enhancing cybersecurity in healthcare: evaluating ensemble learning models for intrusion detection in the internet of medical things," *Sensors*, vol. 24, no. 18, p. 5937, Sep. 2024, doi: 10.3390/s24185937.
[34] T. Ristenpart, E. Tromer, H. Shacham, and S. Savage, "Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds," in *Proceedings of the 16th ACM conference on Computer and communications security*, Nov. 2009, pp. 199–212, doi: 10.1145/1653662.1653687.
[35] E. Barker, "Recommendation for key management – Part 1: General," NIST, 2016. Accessed: Feb 1, 2025 [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-57pt1r5.pdf
[36] E. Barker and W. C. Barker, "Recommendation for key management: Part 2 – best practices for key management organizations (SP 800-57 Part 2 Rev. 1)," *NIST Special Publication 800-57 Part 2 Revision 1*. NIST, pp. 1–91, 2019. Accessed: Feb 1, 2025. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-57pt2r1.pdf
[37] S. Garfinkel and D. Russell, *PGP: pretty good privacy*, 1st ed. USA: O'Reilly & Associates, Inc., 1996.
[38] R. Rivest, "The MD5 message-digest algorithm." MIT Laboratory for Computer Science and RSA Data Security, 2017.
[39] J. Kelsey, B. Schneier, D. Wagner, and C. Hall, "Cryptanalytic attacks on pseudorandom number generators," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1998, vol. 1372, pp. 168–188, doi: 10.1007/3-540-69710-1_12.

[40] R. Morris and K. Thompson, "Password security," *Communications of the ACM*, vol. 22, no. 11, pp. 594–597, Nov. 1979, doi: 10.1145/359168.359172.
[41] C. Gentry, "Fully homomorphic encryption using ideal lattices," in *Proceedings of the Annual ACM Symposium on Theory of Computing*, 2009, pp. 169–178, doi: 10.1145/1536414.1536440.
[42] D. Boneh and M. Franklin, "Identity-based encryption from the weil pairing," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2001, vol. 2139 LNCS, pp. 213–229, doi: 10.1007/3-540-44647-8_13.
[43] H. Krawczyk, "HMQV: A high-performance secure Diffie-Hellman protocol," in *Proceedings CRYPTO*, 2005, pp. 546–566, doi: 10.1007/11535218_33.
[44] D. Eastlake and P. Jones, "US secure hash algorithms (SHA) and HMAC-SHA." RFC 6234, 2011. Accessed: Feb 1, 2025. [Online]. Available: https://www.rfc-editor.org/rfc/rfc6234.html
[45] S. Frankel and H. Herbert, "The AES-XCBC-MAC-96 algorithm and its use with IPsec." RFC 3566, Sep. 2003, doi: 10.17487/rfc3566.
[46] M. Bellare, A. Boldyreva, and A. O'Neill, "Deterministic and efficiently searchable encryption," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2007, vol. 4622 LNCS, pp. 535–552, doi: 10.1007/978-3-540-74143-5_30.
[47] D. Coppersmith and M. Jakobsson, "Almost optimal hash sequence traversal," in *Proceedings 22nd Annual International Cryptology Conference Advances in Cryptology*, 2003, pp. 102–119, doi: 10.1007/3-540-36504-4_8.
[48] B. Malin and L. Sweeney, "How (not) to protect genomic data privacy in a distributed network: Using trail re-identification to evaluate and design anonymity protection systems," *Journal of Biomedical Informatics*, vol. 37, no. 3, pp. 179–192, 2004, doi: 10.1016/j.jbi.2004.04.005.
[49] S. Meiklejohn *et al.*, "A fistful of bitcoins: Characterizing payments among men with no names," in *Proceedings of the ACM SIGCOMM Internet Measurement Conference, IMC*, 2013, pp. 127–139, doi: 10.1145/2504730.2504747.
[50] M. Naor and K. Nissim, "Certificate revocation and certificate update," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 4, pp. 561–570, Apr. 2000, doi: 10.1109/49.839932.
[51] C. Neuman, T. Yu, S. Hartman, and K. Raeburn, "The Kerberos network authentication service (V5)." RFC 4120, Jul. 2005, doi: 10.17487/rfc4120.
[52] N. Faruqui *et al.*, "SafetyMed: A novel IoMT intrusion detection system using CNN-LSTM hybridization," *Electronics*, vol. 12, no. 17, p. 3541, Aug. 2023, doi: 10.3390/electronics12173541.
[53] D. Brown, F. Martinez, and M. Johnson, "Unveiling security threats in healthcare data with long short-term memory networks," *J. Healthc. Inform. Res.*, vol. 9, no. 3, pp. 215–230, 2022.
[54] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, pp. 4171–4186, Oct. 2019.
[55] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2016, vol. 2016-Decem, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[56] M. Kawtar, A. Fathi, N. Assad, and A. Kartit, "Hierarchical spatiotemporal aspect-based sentiment analysis for chain restaurants using machine learning," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 3, p. 1087, 2024, doi: 10.14569/IJACSA.2024.01503109.
[57] I. D. Mienye and Y. Sun, "A survey of ensemble learning: concepts, algorithms, applications, and prospects," *IEEE Access*, vol. 10, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.
[58] K. Mouyassir, A. Fathi, and N. Assad, "Elevating aspect-based sentiment analysis in the moroccan cosmetics industry with transformer-based models," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 6, p. 522, 2024, doi: 10.14569/IJACSA.2024.0150654.
[59] C. I. Bercea, B. Wiestler, D. Rueckert, and S. Albarqouni, "Federated disentangled representation learning for unsupervised brain anomaly detection," *Nature Machine Intelligence*, vol. 4, no. 8, pp. 685–695, Aug. 2022, doi: 10.1038/s42256-022-00515-2.
[60] T. Bilot, N. El Madhoun, K. Al Agha, and A. Zouaoui, "Graph neural networks for intrusion detection: A survey," *IEEE Access*, vol. 11, pp. 49114–49139, 2023, doi: 10.1109/ACCESS.2023.3275789.
[61] A. Ali *et al.*, "Blockchain-powered healthcare systems: enhancing scalability and security with hybrid deep learning," *Sensors*, vol. 23, no. 18, p. 7740, Sep. 2023, doi: 10.3390/s23187740.
[62] Z. Hamid, F. Khalique, S. Mahmood, A. Daud, A. Bukhari, and B. Alshemaimri, "Healthcare insurance fraud detection using data mining," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, p. 112, Apr. 2024, doi: 10.1186/s12911-024-02512-4.

BIOGRAPHIES OF AUTHORS

Abderrahmane Fathi received his engineering degree in information systems and communications engineering from the National School of Applied Sciences in 2019. He is currently a Ph.D. student at The Information Technology Laboratory at Chouaib Doukkali University, El Jadida, Morocco. His research focuses on securing healthcare data and optimizing digital marketing through machine learning, integrating AI-driven solutions to enhance data privacy and decision-making. His areas of interest include cybersecurity, machine learning, deep learning, data privacy, healthcare data security, digital marketing optimization, big data analytics, and AI applications in business and healthcare. He can be contacted at abdou.ft@gmail.com.

Mouyassir Kawtar received her engineering degree in IT and big data from the National School of Applied Sciences in 2020. She is currently a Ph.D. student at The Information Technology Laboratory at Chouaib Doukkali University, El Jadida, Morocco. Her research focuses on AI-Driven sentiment analysis of social media for enhanced digital marketing decisions, leveraging machine learning and natural language processing to analyze consumer opinions. Her areas of interest include machine learning, deep learning, natural language processing, big data analytics, sentiment analysis, social media analytics, and AI applications in digital marketing. She can be contacted at mouyassir.kawtarr@gmail.com.

Ali Waqas received his master's degree in computer science from Government College University (GCU), Lahore, Pakistan in 2020, graduating with an A+ distinction and earning a Gold Medal. He is currently a Faculty Lecturer at the Department of Computer Science, University of Engineering and Technology, Lahore. In addition to his academic role, he is a machine learning engineer with expertise in artificial intelligence, deep learning, and data-driven applications. His research interests include machine learning, deep learning, natural language processing, computer vision, and AI applications in various domains. He can be contacted at waqas.ali2@uet.edu.pk.

Fatima Zahra Fandi received her master's degree in information technology from the National School of Arts and Crafts of Casablanca in 2021. She is currently a Ph.D. student at Hassan II University, Morocco. Her research focuses on leveraging artificial intelligence, blockchain, and visual analytics to enhance digital experiences in various domains, including immersive retail in the metaverse, AI-driven virtual recruitment, and chatbot-based administrative assistance. Her areas of interest include artificial intelligence, cognitive computing, blockchain, metaverse applications, virtual shopping experiences, chatbot solutions, and plant health monitoring using AI. She can be contacted at fatimazahra.fandi10@gmail.com.

Kartit Ali received his doctorate in computer network security from the Faculty of Sciences of Rabat. He is currently a research professor at a leading engineering school and a member of The Information Technology Laboratory at Chouaib Doukkali University, El Jadida, Morocco, specializing in information systems security. With over a decade of experience in higher education, he has been actively engaged in research and teaching best practices in cybersecurity, active directory, and IT security management. Previously, he worked as a trainer at the office of vocational training and employment promotion (OFPPT) in Rabat, Morocco, for nearly eight years. His areas of expertise include network security, cybersecurity best practices, information systems protection, identity management, and advanced security frameworks. He can be contacted at alikartit@gmail.com.