Delta University Scientific Journal
Journal home page: https://dusj.journals.ekb.eg

Secure Cloud-Integrated IoT Framework for Diabetes Detection

Dalia Ebrahim Hamid [1], Hanan M. Amer [2], Hossam El-Din Salah Moustafa [2], and Hanaa Salem Marie [3]

[1] *Electronics and Communications Department, Faculty of Engineering, Delta University for Science and Technology, Gamsaa, Egypt;*
[2] *Electronics and Communication Engineering Department, Faculty of Engineering, Mansoura University, Mansoura, Egypt;*
[3] *Faculty of Artificial Intelligence, Delta University for Science and Technology, Gamsaa, Egypt;*
Correspondence: [Dalia Ebrahim Hamid]; [Faculty of Engineering, Delta University for Science and Technology, Gamasa, Egypt.]; Tel [+201023534729]; Email: Dalia.Hamid@deltauniv.edu.eg

ABSTRACT
The Internet of Things (IoT) and the smart health devices have enhanced healthcare platforms by enabling the remote monitoring of patients' health. Given the unpredictable rise in the diabetes patients number, it is crucial to regularly assess their health conditions to prevent serious illnesses. However, the transmission of a massive volume of sensitive health data brings significant IoT data security challenges. This paper introduces a secure and remote system for diabetes monitoring that employs the Advanced Encryption Standard (AES) to protect patients' sensitive data on cloud-based IoT platforms. In this model, machine learning (ML) methods analyse health data collected by smart health IoT devices to predict critical situations and determine patients' health statuses. The results show that the AES method provides the fastest encryption and decryption times for data files sent from IoT devices to cloud storage. Additionally, the Support Vector Machine (SVM) classification method demonstrates high performance, with an accuracy of 96%, precision of 92.4%, F-score of 95.3%, and recall of 94.3%. Based on these results, the proposed system successfully establishes a efficient and secure platform for health monitoring.

Keywords: *Internet of Things, healthcare, cloud computing, machine learning, classification, security.*

1. Introduction

Recent advancements in the Internet of Things (IoT) and sensor technologies linked to health wearables have significantly enhanced patient care through intelligent and remote health monitoring systems (Mamdiwar et al. 2021). Integrating IoT with cloud technology offers numerous benefits, such as robust processing capabilities, effective resource allocation, and improved user mobility in monitoring models. In modern cloud-based IoT healthcare systems, patient biological data is transmitted, stored, and shared, allowing for insights to be gathered from any location at any time (El Kafhali and El Mir 2023). However, the transfer and storage of medical data in the cloud raise critical privacy and security concerns. Medical data is in particular sensitive, and any alterations can lead to errors in medical diagnoses (Saha and Debnath 2022). Diabetes is a rapidly increasing metabolic disease and a leading cause of death worldwide. Insufficient insulin production by pancreatic cells leads to elevated blood sugar levels, which can severely affect the eyes, various organs, heart, kidneys, and nerves (ElSayed et al. 2023). Additionally, modern studies show that diabetes currently impacts around half a billion people globally, with projections suggesting an increase of 25% to 51% between 2030 and 2045 (Fitzmaurice et al. 2017)(Saeedi et al. 2019). While there is no permanent cure for diabetes, early diagnosis allows for effective management and control. In such cases, computer-aided technologies are invaluable, enabling precise medical decisions and timely, essential treatments (Hennebelle, Materwala, and Ismail 2023). Consequently, machine learning (ML) advancements have made automated diabetes diagnosis and detection more successful and feasible compared to traditional manual methods. Providers of healthcare can leverage these diagnoses to customize interventions, suggest lifestyle modifications, and start early therapy plans (Dasari, Poonguzhali, and Rayudu 2023). Cryptographic mechanisms are utilized to encrypt collected IoT data before it is stored in the cloud , ensuring that the unauthorized users cannot access the data (Babrahem and Monowar 2021). These techniques ensure the availability, integrity, and confidentiality of data by converting it into an incomprehensible form for unauthorized users. Cryptographic

techniques involve two primary methods: asymmetric and symmetric encryption. Asymmetric encryption utilizes a public key for encryption and a private key for decryption, whereas symmetric encryption employs a single private key for both encryption and decryption (Agarwal, Kaushal, and Chouhan 2020). This study envisions a non-invasive strategy for early-stage diabetes patients, employing ML approaches within a secure remote patients' health monitoring environment that uses both cloud and IoT technologies. The essential contributions of this paper are as follows:

• Propose a novel cryptographic method using Advanced Encryption Standard (AES) to enhance the security of healthcare data during cloud storage transmission.
• Describe a disease prediction methodology utilizing various ML techniques to evaluate classification outcomes for early detection of diabetes mellitus.
• Show that the proposed method exceeds current diabetes monitoring models in terms of privacy and security.

2. Related Work

In recent times, there has been notable progress in the smart technologies within the healthcare sector. This techniques has gained widespread adoption and proven effective across various healthcare applications, with a particular focus on medical cardiology. The substantial increase in medical data has provided researchers with an unparalleled opportunity to create and evaluate novel algorithms in this domain. Princy et al. (Princy et al. 2020) diagnosed diabetes and breast cancer was carried out by integrating adaptability features into SVM. The objective was to provide a swift, automated, and flexible diagnostic approach through the use of adaptive SVM. To enhance performance, modifications were made to the bias value in the conventional SVM. Arumugam et al.(Arumugam et al. 2023) Proposed a refined DT model to achieve optimal performance in predicting the likelihood of heart disease in diabetic patients, as it consistently demonstrated superior performance compared to NB and SVM models. Orabi et al. (Orabi, Kamal, and Rabah 2016) used The DT algorithm to suggest an ML-based diabetes prediction system. Their main priority was to determine if the candidates had diabetes at that age. Sisodia et al. (Sisodia and Sisodia 2018) proposed diabetes prediction utilizing classification techniques such as NB, SVM, and DT. In this study the accuracy was 76.30%. Hasan et al. (Hasan et al. 2020) used multiple ML classifiers, including RF, DT, NB, KNN, and XGBoost. They created a weighted ensemble machine learning model that achieves the highest possible AUC value. Ramesh et al.(Ramesh, Aburukba, and Sagahyroon 2021) developed a comprehensive healthcare monitoring framework to effectively manage diabetes. The accuracy was 83.2%, and the sensitivity was 87.2%. Hrimov et al.(Hrimov et al. 2021) proposed a diabetes classification technique based on backward elimination and SVM. They attained an overall accuracy of 85.71%.

Table 1. Comparing factors in the proposed model with the previous works

Reference	Applied technology	Healthcare System	Security	AES Encryption
(Arumugam et al. 2023)	Cloud-based IoT	Heart disease detection	x	x
(Kumar et al. 2018)	Cloud-based IoT	Diabetes Detection	x	x
(Ahmed et al. 2018)	Cloud	Heart disease detection	x	x
(Khanna et al. 2023)	IoT	Heart disease detection	x	x
(Deepika et al. 2021)	Cloud-based IoT	Medical image diagnosis	√	x
(Akhbarifar et al. 2020)	Cloud-based IoT	Diabetes and heart prediction	√	x
(Hosseinzadeh et al. 2021)	Cloud-based IoT	chronic kidney disease	x	x
(Nigar et al. 2023)	Cloud-based IoT	Identification of six key chronic diseases	x	x
(Siddiqui et al. 2023)	IoT	Covid-19 detection	x	x
(Stergiou et al. 2023)	Cloud-based IoT	Identify harmful forms of viruses	√	x
(Asghari et al. 2019)	Cloud-based IoT	Predicting a combination of diseases	√	x
proposed system	Cloud-based IoT	Diabetes detection	√	√

In the reviewed literature, the security issues were considered in (Deepika et al. 2021), (Akhbarifar et al. 2020),(Stergiou et al. 2023),(Asghari et al. 2019) while in others were not as focused. Compared to the reviewed studies, we aim to provide a secure remote health monitoring model in a IoT-based cloud framework utilizing ML methods for early ddiagnosis of diseases, which employs an AES encryption method and is a appropriate solution for constrained health IoT resources, which has not been used in previous works. As a result, our suggested model's key addition in comparison to earlier research is that it considers confidentiality and security problems in an operational way while taking into account IoT resource restrictions. To demonstrate this benefit, certain comparative elements are evaluated in Table 1, including the following: presenting applicable technologies, healthcare systems, security challenges, and the use of encryption methods in the researched studies vs our suggested framework. As indicated in Table 1, our contribution extends beyond the introduction of a cloud-based IoT health monitoring model; it includes the incorporation of a hybrid data encryption method, a feature that was not taken into account in other studies.

3. Methodology

3.1. Layers of the IoT environment

In the IoT context, architecture refers to the framework that outlines both the hardware and software components of the system, the standard organization and configuration of the network, as well as the operational methods and data formats to be employed. The architecture of IoT systems varies depending on the specific application, as each has its own unique requirements and implementation needs(Balaji, Nathani, and Santhakumar 2019). Figure 1 illustrates the four-layer architecture that utilized for the proposed system.

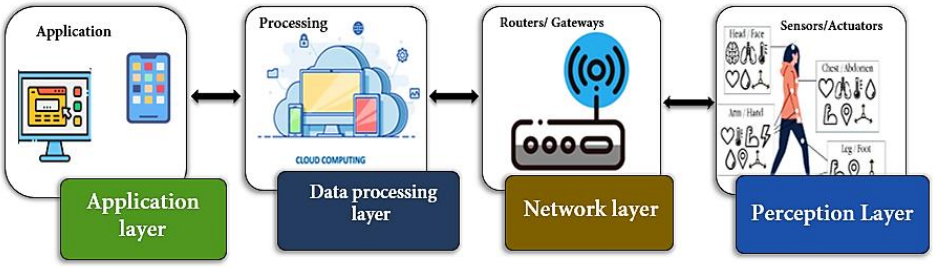

Figure 1: Four-layer system architecture

The system's four-layer architecture includes:

A. Application Layer: This is the user interface layer, responsible for executing user commands and providing information and notifications.
B. Data Processing Layer: This layer handles the processing and analysis of data collected by the perception layer, using advanced analytics to make informed decisions.
C. Network Layer: This layer manages the data transmission from sensors and directs commands to actuators. It also links the system with other devices such as smartphones.
D. Perception Layer: This physical layer involves sensors collecting data and actuators interacting with the environment.

3.2. Proposed system architecture

Healthcare monitoring systems present significant opportunities to revolutionize traditional patient management. These systems help medical centers enhance patient treatment, reduce healthcare costs, and facilitate remote health monitoring. The architecture of these systems leverages advanced technologies, including IoT, embedded systems, smartphone applications, and programmable development boards. Figure 2 demonstrates the structure and interconnections of the system components.

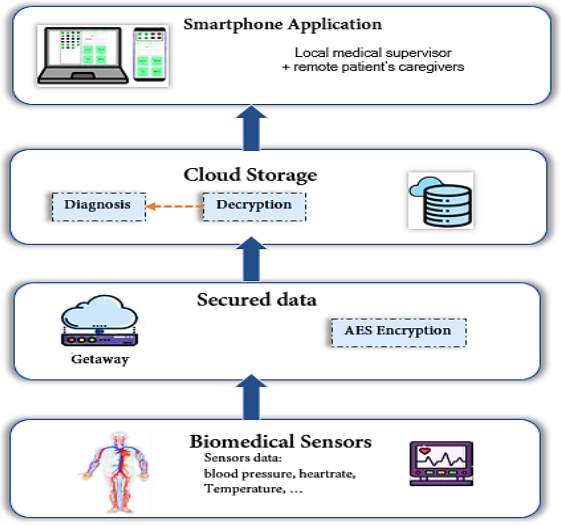

Figure 2: The Proposed Health Condition Monitoring in a cloud-based IoT platform

3.2.1. Biomedical sensors

The application of medical sensors is rapidly increasing. These sensors play a crucial role in medical diagnosis by collecting data related to patients' medical conditions, thereby enhancing their quality of life. There are four main types of sensors: off-body, environmental, implantable, and wearable. In the proposed system, the sensing section is tasked with measuring various physiological indicators of patients, such as heart rate, body temperature, and blood oxygen levels. Additionally, it includes several sensors that gather data about the patient's environment.

Algorithm 1. outlines the procedures for gathering the IoT data for the disease prediction process.

Algorithm 1: Data Acquiring

Input: IoT medical device sensor data

Output: Required medical data

1. Enter the IoT device data including the identification data and clinical data of the patient.
2. Collect the IoT medical device sensor data by sensors.
3. Transfer all the Acquired medical data to the algorithm for encryption process.

3.2.2. Providing the data security in the proposed model

- *Advanced Encryption Standard Algorithm (AES)*

AES is a commonly used symmetric encryption method. It is a symmetric key algorithm, which means it uses the same key for encryption and decryption. The U.S. National Institute of Standards and Technology (NIST) adopted AES as a standard in 2001, replacing the previous Data Encryption Standard (Joan and Vincent 2002).
Key Sizes: AES supports key sizes of 256, 192, and 128 bits. larger size of key, stronger the encryption, but it also increases the computational complexity.
Block Size: AES operates on fixed-size blocks of data. The block size for AES is 128 bits.
Rounds: AES uses a fixed number of rounds for processing data, with the rounds number dependent on the key size. For example, 14 rounds for 256-bit keys, 12 rounds for 192-bit keys, and 10 rounds for 128-bit keys.
Symmetric Cryptography: As a symmetric algorithm, AES employs the same key for both decryption and encryption. This contrasts with asymmetric (or public-key) cryptography, where separate keys are utilized for encryption and decryption.
Security: AES is considered highly secure and is widely adopted for securing sensitive data. The security strength of AES is directly related to the key size, with longer keys providing a higher level of security.

AES algorithm consists of several rounds, each involving Shift Rows, Sub Bytes, Add Round Key, and Mix Columns functions, except for the final round. In the Sub Bytes step, a linear substitution is performed for each byte using an 8-bit S-box derived from the multiplicative inverse over the finite Galois Field GF (2^8), This S-box ensures nonlinearity in the cipher system, preventing fixed-point and opposite-fixed-point attacks (Abdullah 2017). The ShiftRows step cyclically shifts the bytes in every row, with the first row remaining unaffected. The second row shifts one byte to the left, and the third and fourth rows shift by offsets of two and three bytes, respectively. This pattern is consistent for 192-bit or 128-bit blocks. In 256-bit blocks, the first row remains unchanged, and the second, third, and fourth rows shift by 1, 3, and 4 bytes, respectively. The MixColumns step involves merging the four bytes of each column using a linear invertible transformation. This function treats each column as a polynomial over GF (2^8), and multiplies it with a fixed polynomial, providing diffusion in the cipher system. The MixColumns function, combined with ShiftRows, contributes to the overall security of the AES algorithm. The MixColumns step can alternatively be viewed as the multiplication by a specific matrix, as illustrated in Figure 5.

$$c(x) = 3x^3 + x^2 + x + 2 \quad (1)$$

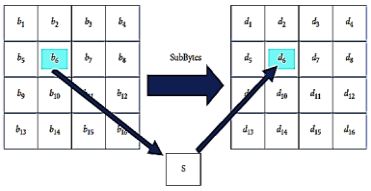

Figure 3: SubBytes step

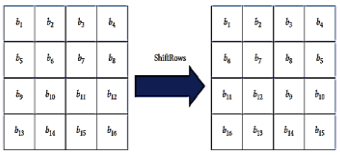

Figure 4: ShiftRows step

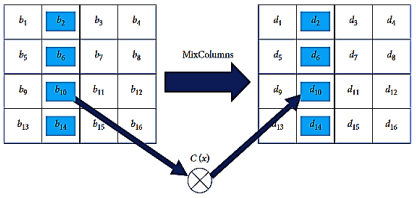

Figure 5: MixColumns step (Siam et al. 2021)

The depicted process illustrates the secure handling of a user's data request within the server infrastructure. Figure 6 illustrates the user and data stored on the cloud server, depicting how a user's data request is processed by the server and how the data is accessed securely. By meticulously following these steps and incorporating robust security measures, the depicted process ensures that user data requests are processed securely, maintaining the confidentiality, integrity, and availability of sensitive information within the healthcare system.

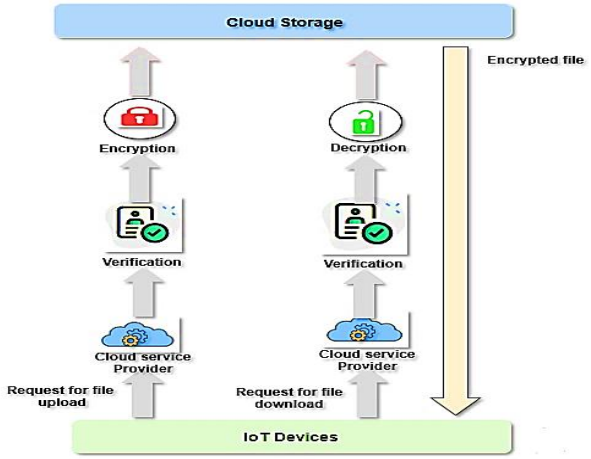

Figure 6: Secure IoT data stored on cloud server

Applying AES encryption for securing communication between IoT devices and cloud computing involves several key steps. AES is a symmetric encryption algorithm, meaning the same key is used for both encryption and decryption. By following these steps, AES encryption can be effectively applied to secure communication between IoT devices and cloud computing platforms, protecting sensitive data from unauthorized access and ensuring the integrity of the transmitted information.

1. Key Generation and Management

Generate a strong, random cryptographic key. The key length can be 128, 192, or 256 bits, with 256 bits providing the highest level of security. The key should be generated using a secure random number generator to ensure that it is unpredictable. Store the encryption key securely in both the IoT device and the cloud. This might involve using a secure hardware module, such as a Trusted Platform Module (TPM) or a Hardware Security Module (HSM), to protect the key from unauthorized access. Implement key rotation policies to periodically change the encryption key, reducing the risk of key compromise over time. Use a key management service in the cloud to handle key generation, storage, and rotation securely.

2. Data Encryption on the IoT Device

The IoT device collects data from sensors or other inputs that need to be securely transmitted to the cloud. Before transmitting the collected data, the IoT device encrypts it using the AES encryption algorithm.

3. Secure Data Transmission to the Cloud

Establish a secure communication channel between the IoT device and the cloud using protocols like TLS (Transport Layer Security). TLS provides an additional layer of security, ensuring that the encrypted data is not tampered with or intercepted during transmission. Ensure that mutual authentication is in place so that both the IoT device and the cloud can verify each other's identity before transmitting data. The encrypted data, along with any necessary metadata, is transmitted securely to the cloud over the established secure communication channel.

4. Data Decryption in the Cloud

The cloud server receives the encrypted data from IoT device. The server of cloud uses the same AES key and the corresponding block cipher mode (along with the metadata like IV or nonce) to decrypt the data. Ensure that the decryption process is performed in a secure environment, such as within an HSM, to protect the decryption key and the plaintext data.

5. Secure Data Processing and Storage in the Cloud

Once decrypted, the data can be processed as needed by cloud-based applications or services. However, to maintain security, consider processing sensitive data in memory or within a secure environment that limits exposure to unauthorized entities. If the decrypted data needs to be stored, apply additional encryption for data at rest. This could involve encrypting the data again using a different key, ensuring that even if storage systems are compromised, the data remains secure. Use access control mechanisms to restrict who can access the decrypted data, ensuring that only authorized users or services have access.

6. Key Rotation and Revocation

Periodically rotate encryption keys to minimize the risk of key compromise. This involves generating new keys and securely distributing them to the IoT devices and cloud systems. Ensure that both old and new keys are handled correctly during the rotation process to avoid data access issues. Implement mechanisms for key revocation in case an encryption key is compromised. Once a key is *revoked*, any data encrypted with that key should be re-encrypted with a new key.

3.2.3. The architecture of the medical data classification component in the proposed healthcare system

ML methods are employed to classify patients as sick or healthy based on their vital signs. Figure 7 depicts the proposed block diagram for creating the ML model for diabetes prediction, outlining the approach used to update or train the model and forecast the occurrence of diabetes. Below is an explanation of the workflow for training, updating, and predicting in detail.

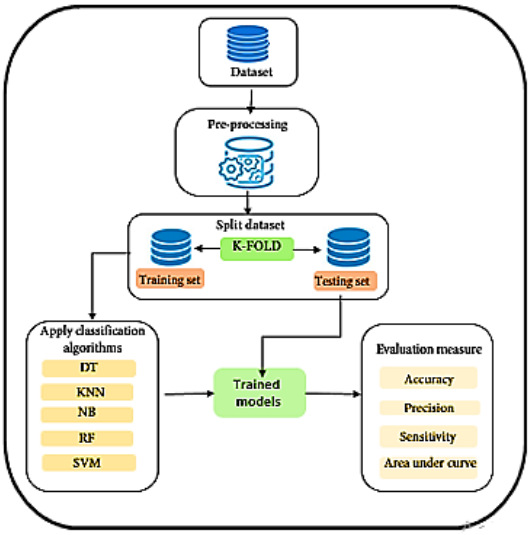

Figure 7: Framework for ML model

a) Dataset Pre-processing

The pre-processing stage of the proposed algorithm involves outlier rejection (n), and imputing missing values (Q). An outlier is a noticing that significantly deviates from other data points in a dataset (Hamid et al. 2024). Since algorithms of classification are sensitive to data distribution, it is essential to exclude such deviations. We used the Interquartile Range (IQR) method to remove outliers, with the mathematical formula for this process provided in equation (1).

$$\mathrm{p(a)} = \begin{cases} a & if\ Q1 - 1{,}5 * IQR \leq a \leq Q3 + 1{,}5 * IQR \\ reject & otherwise \end{cases} \quad (1)$$

Here, a represents the dimensional space of m feature vector instances, where $a \in \boldsymbol{R^m}$. Given that $Q1$, $Q3$, and the IQR belong to $\boldsymbol{R^m}$, Q1 represents the first quartile, $Q3$ corresponds to the third quartile, and the IQR denotes the interquartile range of the features used.

After outliers were removed, the attributes underwent further processing to address any missing or null values (Fan et al. 2021). Null or missing values can negatively affect the accuracy of classifier predictions. In the proposed framework, rather than discarding instances with missing values, imputation was performed using the mean values of the features, as shown in equation (2). This approach is advantageous because it enables the imputation of continuous data without creating additional outliers.

$$Q(f) = \begin{cases} mean(f)\ . & if\ f = null/missed \\ f\ . & otherwise \end{cases} \quad (2)$$

where f represents the frequency of the feature vector's occurrences in n-dimensional space, $f \in R^m$.

b) K-Fold Cross-Validation (KCV)

KCV is a commonly employed method for selecting models and estimating the error of classifiers (Arlot and Celisse 2010). Figure 8 illustrates the visual representation of the data splitting technique utilized in this study, which follows a 5-fold cross-validation approach. The dataset was separated into K folds, and the models were trained using the K-1 folds. The optimal hyperparameters and unreleased testing data were used (K times) To evaluate the models' performance through the outermost loop. Furthermore, to maintain a consistent percentage of samples in each class, KCV has been employed as the dataset consists of both positive and negative samples. Equation (3) was used to generate the final evaluation metrics (Hasan et al. 2020).

$$\mathrm{M} = \frac{1}{\mathrm{k}} \times \sum_{\mathrm{n=1}}^{\mathrm{k}} \mathrm{P_n} \pm \sqrt{\frac{\sum_{\mathrm{n=1}}^{\mathrm{k}} (\mathrm{p_n} - \bar{\mathrm{p}})^2}{\mathrm{k} - 1}} \quad (3)$$

In the above equation, M denotes the final performance metric of the classifiers, K represents the number of folds utilized in the evaluation, and $\mathrm{P_n} \in \mathrm{R}$, where n = 1, 2, . . ., K, signifies the efficiency metric for each fold.

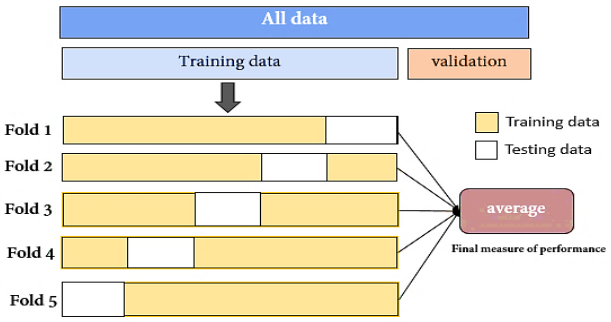

Figure 8: 5-cross-fold validation of the applied dataset

4. Performance Evaluation

The effectiveness of the proposed security-enhanced IoT-based cloud system is validated through performance evaluation metrics, including encryption time, decryption time, accuracy, specificity, sensitivity, precision, recall, and error rate. The efficiency of this secured method is also compared with several traditional approaches.

4.1 Dataset Description

The primary dataset used in this study was the Pima Indian Diabetes Dataset (PIDD), obtained from the standard dataset repository at the University of California, Irvine (UCI). The PIDD includes archives from 768 diabetic patients and contains 8 unique features. For analysis, the dataset is split into two groups: 500 non-diabetic individuals labeled as (0) and 268 diabetic patients labeled as (1). Table 2 outlines the PIDD's features and offers a brief statistics overview. Furthermore, it depicts a sample of the obtained occurrences in the PIDD.

Table 2 A PIDD statistics summary

No.	Features	Descriptions	Min.	Max.	Mean ± Std
F1	Pregnancies	Number of times pregnant	0	17	3.85 ± 3.37
F2	Glucose	Plasma glucose concentration 2 hours in an oral glucose tolerance test	0	199	120.9 ± 31.97
F3	Blood Pressure	Diastolic blood pressure	0	122	69.11 ± 19.36
F4	Skin Thickness	Triceps skin fold thickness	0	99	20.54 ± 15.95
F5	Insulin	2-h serum insulin	0	846	79.81 ± 115.24
F6	BMI	Body mass index (weight in kg/(height in m)^2)	0	67.1	32.00 ± 7.88
F7	Diabetes Pedigree Function	Diabetes pedigree function	0.08	2.42	0.47 ± 0.33
F8	Age	Age of person	21	81	33.24 ± 11.76

4.2 Encryption Time

The time required to encrypt plaintext into ciphertext using the proposed AES is measured and compared with the existing techniques of DES, Blowfish, and ECC, as illustrated in Figure 9.

4.3 Decryption Time

As a result, the time required to decrypt the ciphered text back into plain text is measured for the proposed AES method and compared with existing techniques such as DES, Blowfish, and ECC, as illustrated in Figure 10.

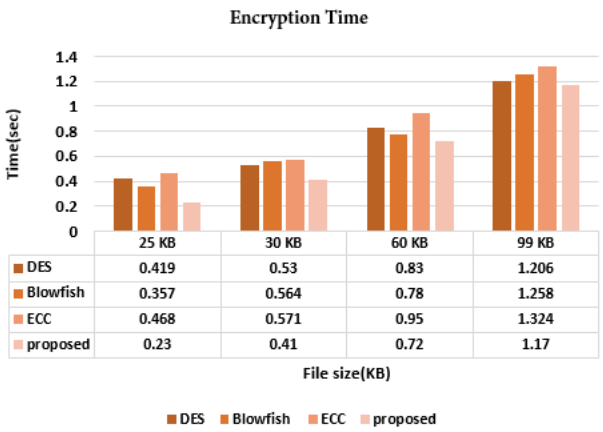

Figure 9: Encryption time comparison.

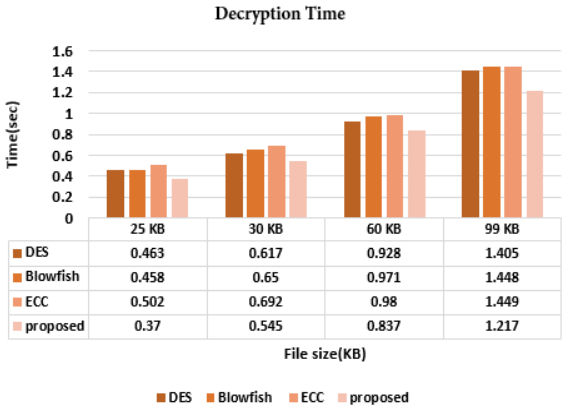

Figure 10: Decryption time comparison.

4.4 Comparisons of the proposed Cryptography algorithm with different approaches

Various techniques were assessed for their space optimization and functionality within cloud-IoT environments. Table 3 presents a various algorithms comparison based on different parameters. The performance of the cryptographic algorithms was evaluated by examining rounds, the number of keys used, block size, and key length.

Table 3. Cryptography Algorithm Comparison.

	Blowfish	**DES**	**ECC**	**RSA**	**AES**
Cipher type	Symmetric	Symmetric	Asymmetric	Asymmetric	Symmetric
No. of key	1	1	2	2	1
Key length	32 to 448bits	56bits	160 bits	1024 bits	128,192,256 bits
Rounds	16	16	16	1	10,12,14
Block size	64bits	64bits	64bits	Min 512 bits	128bits

4.5 classification results

The metrics utilized for comparison include accuracy, precision, f-score, and recall. Before delving into the performance measures, Table 4 presents the concept of the confusion matrix.

Table 4. Confusion Matrix

	Actual positive	Actual negative
Predicted positive	True Positive (TP)	False Positive(FP)
Predictive negative	False Negative (FN)	True Negative(TN)

Accuracy represents the percentage of instances that are correctly classified. It is one of the most widely used metrics for evaluating classification performance, with higher values (closer to 100) indicating better performance.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Precision measures the proportion of true positive results among all the positive results predicted by the model. It indicates how many of the predicted positive instances are actually correct. High precision means that the model has a low false positive rate.

$$\text{Precision} = \frac{TN}{TN+FP} \tag{5}$$

Recall, also known as sensitivity or true positive rate, is mathematically defined as:

$$\text{recal} = \frac{TP}{TP+FN} \tag{6}$$

The F-score evaluates the accuracy of the testing process by averaging both precision and recall. It is represented in Equation (7).

$$\text{F} - \text{score} = \frac{2TP}{2TP+FP+TN} \tag{7}$$

To classify the instances of the disease, multiple experiments were conducted using different classification algorithms including DT, SVM, KNN, and RF. A performance comparison was carried out using various metrics such as accuracy, precision, recall, and F-score. The performance evaluation based on these metrics can be found in Table 5, verifying the performance examination. The outcomes for selecting the optimal pre-processing and ML model are presented in Table 5. This table displays the accuracy, precision, recall, and F-score values, allowing for comparison among the different approaches. The summary outlines each model's ability to achieve the highest accuracy when following the proposed pipeline. It has been observed that all classifiers achieve their optimal performance when outlier rejection (P), filling missing values (Q), and correlation-based feature selection techniques (CRB) are applied to the PIDD. Each classifier demonstrates its highest performance when these pre-processing steps are employed. The two experiments, as shown in Table 5, show that The Accuracy of the SVM, DT, KNN, and RF has improved by 29.3%, 12.9%,17%, 20.2%, and 21.4% respectively when P, Q, and CRB.

Table 5. The performance of the proposed approach is compared to existing approaches, and it is observed that the folding technique yields the best result among all the approaches.

	Classifier	Accuracy	Precision	F-score	Recall
Without pre-processing	DT	0.747	0.625	0.730	0.727
	RF	0.721	0.607	0.659	0.618
	KNN	0.695	0.583	0.759	0.509
	SVM	0.667	0.655	0.735	0.76
P+Q+ CRB (n-attribute=6) proposed	DT	0.876	0.853	0.830	0.891
	RF	0.935	0.915	0.926	0.930
	KNN	0.897	0.886	0.896	0.902
	SVM	0.960	0.924	0.953	0.943

Figure 11, Figure 12, Figure 13, Figure 14 demonstrate The performance of multiple graphical representation was assessed to identify the optimal ML models for proposed approach with the highest accuracy, precision, recall, and f-score. The corresponding best-performing models are presented in Table 5.

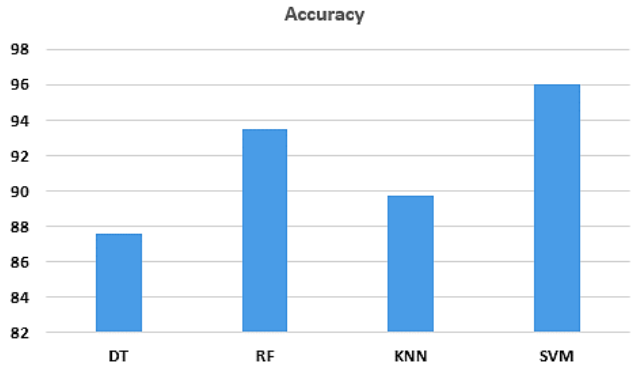

Figure 11 Accuracy representation for proposed ML models

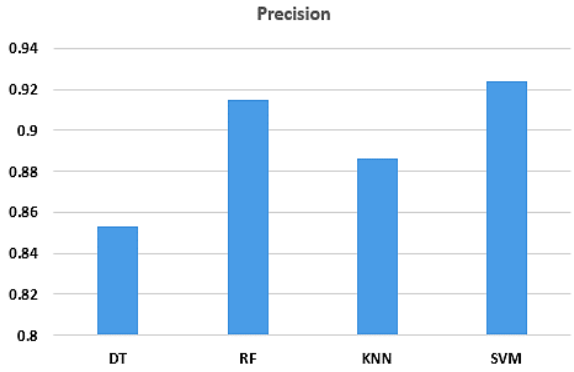

Figure 12 Precision representation for proposed ML models

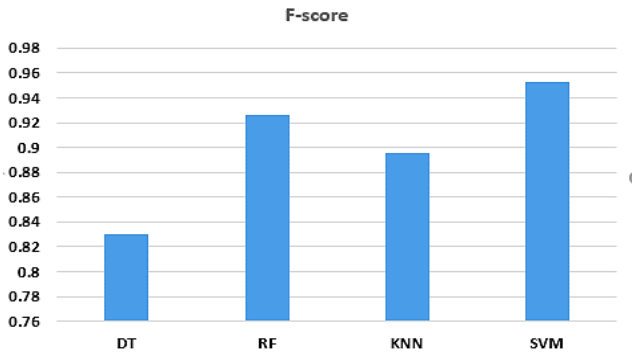

Figure 13 F-score representation for proposed ML models

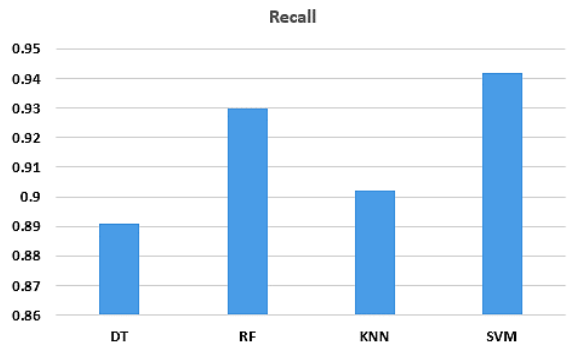

Figure 14 Recall representation for proposed ML models

5. Conclusion

Given the significant rise in the number of diabetics, there is an increasing need for IoT-based cloud platforms for disease prediction systems and healthcare monitoring. Ensuring patient privacy and securing sensitive healthcare data are major challenges. This study introduces an efficient approach that prioritizes patients' data privacy while utilizing medical data for disease detection within the recent healthcare environment. The system employs the AES technique for encryption. To assess the system's effectiveness, analyses were performed on both classification performance and secure transmission of data. The AES algorithm's performance was compared with conventional encryption algorithms like ECC, Blowfish, and DES, focusing on decryption and encryption times. Additionally, the SVM classifier's performance was assessed against classifiers such as KNN, DT, and RF. The SVM classifier demonstrated superior performance with 96% accuracy, 92.4% precision, 94.3% recall, and a 95.3% F-score. The results indicate that The suggested method surpasses current systems in predicting diseases with greater accuracy while also improving security measures. Future research could focus on integrating more generalized approaches to manage a wider range of datasets, extending beyond those gathered by IoT devices. Additionally, investigating deep learning methods could enhance disease prediction accuracy while ensuring strong privacy and security protections.

References

Abdullah, Ako Muhamad. 2017. "Advanced Encryption Standard (AES) Algorithm to Encrypt and Decrypt Data." *Cryptography and Network Security* 16(1):11.

Agarwal, Vidushi, Ashish K. Kaushal, and Lokesh Chouhan. 2020. "A Survey on Cloud Computing Security Issues and Cryptographic Techniques." Pp. 119–34 in *Social Networking and Computational Intelligence: Proceedings of SCI-2018*. Springer.

Ahmed, Md Razu, S. M. Hasan Mahmud, Md Altab Hossin, Hosney Jahan, and Sheak Rashed Haider Noori. 2018. "A Cloud Based Four-Tier Architecture for Early Detection of Heart Disease with Machine Learning Algorithms." Pp. 1951–55 in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*. IEEE.

Akhbarifar, Samira, Hamid Haj Seyyed Javadi, Amir Masoud Rahmani, and Mehdi Hosseinzadeh. 2020. "A Secure Remote Health Monitoring Model for Early Disease Diagnosis in Cloud-Based IoT Environment." *Personal and Ubiquitous Computing* 1–17.

Arlot, Sylvain, and Alain Celisse. 2010. "A Survey of Cross-Validation Procedures for Model Selection."

Arumugam, K., Mohd Naved, Priyanka P. Shinde, Orlando Leiva-Chauca, Antonio Huaman-Osorio, and Tatiana Gonzales-Yanac. 2023. "Multiple Disease Prediction Using Machine Learning Algorithms." *Materials Today: Proceedings* 80:3682–85.

Asghari, Parvaneh, Amir Masoud Rahmani, and Hamid Haj Seyyed Javadi. 2019. "A Medical Monitoring Scheme and Health-medical Service Composition Model in Cloud-based IoT Platform." *Transactions on Emerging Telecommunications Technologies* 30(6):e3637. doi: 10.1002/ett.3637.

Babrahem, Afnan Salem, and Muhammad Mostafa Monowar. 2021. "Preserving Confidentiality and Privacy of the Patient's EHR Using the OrBAC and AES in Cloud Environment." *International Journal of Computers and Applications* 43(1):50–61. doi: 10.1080/1206212X.2018.1505025.

Balaji, Subramanian, Karan Nathani, and Rathnasamy Santhakumar. 2019. "IoT Technology, Applications and Challenges: A Contemporary Survey." *Wireless Personal Communications* 108:363–88.

Dasari, Srilaxmi, Boo Poonguzhali, and Manjula Sri Rayudu. 2023. "An Efficient Machine Learning Approach for Classification of Diabetic Retinopathy Stages." *Indonesian Journal of Electrical Engineering and Computer Science* 30(1):81–88. doi: 10.11591/ijeecs.v30.i1.pp81-88.

Deepika, J., C. Rajan, and T. Senthil. 2021. "Security and Privacy of Cloud-and IoT-Based Medical Image Diagnosis Using Fuzzy Convolutional Neural Network." *Computational Intelligence and Neuroscience* 2021:1–17. doi: 10.1155/2021/6615411.

ElSayed, Nuha A., Grazia Aleppo, Vanita R. Aroda, Raveendhara R. Bannuru, Florence M. Brown, Dennis Bruemmer, Billy S. Collins, Jason L. Gaglia, Marisa E. Hilliard, and Diana Isaacs. 2023. "2. Classification and Diagnosis of Diabetes: Standards of Care in Diabetes—2023." *Diabetes Care* 46(Supplement_1):S19–40. doi: 10.2337/dc23-S002.

Fan, Cheng, Meiling Chen, Xinghua Wang, Jiayuan Wang, and Bufu Huang. 2021. "A Review on Data Preprocessing Techniques toward Efficient and Reliable Knowledge Discovery from Building Operational Data." *Frontiers in Energy Research* 9:652801. doi: 10.3389/fenrg.2021.652801.

Fitzmaurice, Christina, Christine Allen, Ryan M. Barber, Lars Barregard, Zulfiqar A. Bhutta, Hermann Brenner, Daniel J. Dicker, Odgerel Chimed-Orchir, Rakhi Dandona, and Lalit Dandona. 2017. "Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived with Disability, and Disability-Adjusted Life-Years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study." *JAMA Oncology* 3(4):524–48. doi: 10.1001/jamaoncol.2016.5688.

Hamid, Dalia Ebrahim, Hanan M. Amer, Hossam El-Din Salah Moustafa, and Hanaa Salem Marie. 2024. "Empowering Health Data Protection: Machine Learning-Enabled Diabetes Classification in a Secure Cloud-Based IoT Framework." *Indonesian Journal of Electrical Engineering and Computer Science* 34(2):1100–1111.

Hasan, Md Kamrul, Md Ashraful Alam, Dola Das, Eklas Hossain, and Mahmudul Hasan. 2020. "Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers." *IEEE Access* 8:76516–31.

Hennebelle, Alain, Huned Materwala, and Leila Ismail. 2023. "HealthEdge: A Machine Learning-Based Smart Healthcare Framework for Prediction of Type 2 Diabetes in an Integrated IoT, Edge, and Cloud Computing System." *Procedia Computer Science* 220:331–38. doi: 10.1016/j.procs.2023.03.043.

Hosseinzadeh, Mehdi, Jalil Koohpayehzadeh, Ahmed Omar Bali, Parvaneh Asghari, Alireza Souri, Ali

Mazaherinezhad, Mahdi Bohlouli, and Reza Rawassizadeh. 2021. "A Diagnostic Prediction Model for Chronic Kidney Disease in Internet of Things Platform." *Multimedia Tools and Applications* 80:16933–50. doi: 10.1007/s11042-020-09049-4.

Hrimov, Andrew, Ievgen Meniailov, Dmytro Chumachenko, Kseniia Bazilevych, and Tetyana Chumachenko. 2021. "Classification of Diabetes Disease Using Logistic Regression Method." Pp. 147–57 in *Integrated Computer Technologies in Mechanical Engineering-2020: Synergetic Engineering*. Springer.

Joan, Daemen, and Rijmen Vincent. 2002. "The Design of Rijndael: AES-the Advanced Encryption Standard." *Information Security and Cryptography* 196.

El Kafhali, Said, and Iman El Mir. 2023. "Exploring the Effectiveness of Cloud, Internet of Things and Fog Computing for Healthcare Monitoring Systems." Pp. 77–91 in *Computational Intelligence for Medical Internet of Things (MIoT) Applications*. Elsevier.

Khanna, Ashish, Pandiaraj Selvaraj, Deepak Gupta, Tariq Hussain Sheikh, Piyush Kumar Pareek, and Vishnu Shankar. 2023. "Internet of Things and Deep Learning Enabled Healthcare Disease Diagnosis Using Biomedical Electrocardiogram Signals." *Expert Systems* 40(4):e12864.

Kumar, Priyan Malarvizhi, S. Lokesh, R. Varatharajan, Gokulnath Chandra Babu, and P. Parthasarathy. 2018. "Cloud and IoT Based Disease Prediction and Diagnosis System for Healthcare Using Fuzzy Neural Classifier." *Future Generation Computer Systems* 86:527–34.

Mamdiwar, Shwetank Dattatraya, Zainab Shakruwala, Utkarsh Chadha, Kathiravan Srinivasan, and Chuan-Yu Chang. 2021. "Recent Advances on IoT-Assisted Wearable Sensor Systems for Healthcare Monitoring." *Biosensors* 11(10):372. doi: 10.3390/bios11100372.

Nigar, Natasha, Abdul Jaleel, Shahid Islam, Muhammad Kashif Shahzad, and Emmanuel Ampoma Affum. 2023. "IoMT Meets Machine Learning: From Edge to Cloud Chronic Diseases Diagnosis System." *Journal of Healthcare Engineering* 2023. doi: 10.1155/2023/9995292.

Orabi, Karim M., Yasser M. Kamal, and Thanaa M. Rabah. 2016. "Early Predictive System for Diabetes Mellitus Disease." Pp. 420–27 in *Advances in Data Mining. Applications and Theoretical Aspects: 16th Industrial Conference, ICDM 2016, New York, NY, USA, July 13-17, 2016. Proceedings 16*. Springer.

Princy, R. Jane Preetha, Saravanan Parthasarathy, P. Subha Hency Jose, Arun Raj Lakshminarayanan, and Selvaprabu Jeganathan. 2020. "Prediction of Cardiac Disease Using Supervised Machine Learning Algorithms." Pp. 570–75 in *2020 4th international conference on intelligent computing and control systems (ICICCS)*. IEEE.

Ramesh, Jayroop, Raafat Aburukba, and Assim Sagahyroon. 2021. "A Remote Healthcare Monitoring Framework for Diabetes Prediction Using Machine Learning." *Healthcare Technology Letters* 8(3):45–57. doi: 10.1049/htl2.12010.

Saeedi, Pouya, Inga Petersohn, Paraskevi Salpea, Belma Malanda, Suvi Karuranga, Nigel Unwin, Stephen Colagiuri, Leonor Guariguata, Ayesha A. Motala, and Katherine Ogurtsova. 2019. "Global and Regional Diabetes Prevalence Estimates for 2019 and Projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas." *Diabetes Research and Clinical Practice* 157:107843. doi: 10.1016/j.diabres.2019.107843.

Saha, Himadri Nath, and Subhradip Debnath. 2022. "Security and Privacy of IoT Devices in Healthcare Systems." *Smart Healthcare System Design: Security and Privacy Aspects* 143–65. doi: 10.1002/9781119792253.ch7.

Siam, Ali I., Mohammed Amin Almaiah, Ali Al-Zahrani, Atef Abou Elazm, Ghada M. El Banby, Walid El-Shafai, Fathi E. Abd El-Samie, and Nirmeen A. El-Bahnasawy. 2021. "Secure Health Monitoring Communication Systems Based on IoT and Cloud Computing for Medical Emergency Applications." *Computational Intelligence and Neuroscience* 2021.

Siddiqui, Salman Ahmad, Anwar Ahmad, and Neda Fatima. 2023. "IoT-Based Disease Prediction Using Machine Learning." *Computers and Electrical Engineering* 108:108675.

Sisodia, Deepti, and Dilip Singh Sisodia. 2018. "Prediction of Diabetes Using Classification Algorithms." *Procedia Computer Science* 132:1578–85.

Stergiou, Christos L., Andreas P. Plageras, Vasileios A. Memos, Maria P. Koidou, and Konstantinos E. Psannis. 2023. "Secure Monitoring System for IoT Healthcare Data in the Cloud." *Applied Sciences* 14(1):120. doi: 10.3390/app14010120.