INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ORIGINAL ARTICLE

Check for updates

OPEN ACCESS

Received: 17/10/2024

Accepted: 04/04/2025

Published: 06/05/2025

Citation: Priyanka K, Vipin S, Shalini C (2025) Prediction of Type-II Diabetic through Blockchain Technology. Indian Journal of Science and Technology 18(15): 1177-1188. https://doi.org/10.17485/IJST/v18i15.3423

* **Corresponding author**.

profvipinsaxena@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Prediction of Type-II Diabetic through Blockchain Technology

Kaushal Priyanka[1], Saxena Vipin[1]*, Chandra Shalini[1]

1 Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, 226025, Uttar Pradesh, India

Abstract

Objectives: Blockchain technology is a recent development for extracting relevant information from the vast amount of data stored over cloud servers. The main objective of the present work is to use blockchain technology to predict diabetic patients at an early stage, and the proposed method improves data security and integrity. **Methods:** The proposed methodology consists of five key stages i.e. data collection, preprocessing, block formation, block evaluation, and finally valuable results are extracted. The source of data is related to diabetic patients which is taken from the Kaggle repository and categorized by age group. A blockchain model is developed and Python programming language is used to secure storage of the data, real-time analysis, and enhanced data security. Additionally, the study also finds association values between age groups and diabetic conditions which provide more meaningful information about affected people from diabetes. **Findings:** In this present work, statistical technique is used to evaluate the results and observe that the prevalence of diabetes increases with age, especially between the age group from 50 to 60. The study shows that blockchain may enhance chronic disease management by providing safe data storage and real-time analysis. The study combines blockchain technology with statistical correlation analysis, making it more suited for prediction rather than monitoring. Additionally, the storage size and retrieval time of the data are also computed in this study. **Novelty:** The primary contribution of the present work is that age-based block segmentation offers an innovative method for organizing and analysing patient data, especially for type-II diabetic, which have not been highlighted in other studies available in the literature.

Keywords: Blockchain Technology; Medical Database; Diabetic Patients; Early Prediction; Data Security

1 Introduction

In the old days, data was only represented in the form of text, but due to the evolution of the Graphical User Interface (GUI), the data may be represented in the form of text, image, audio, and video formats. The approach for storage of data has a centralized storage system in the early days, and the Electronic Health Record (EHR) is based upon said approach [1]. However, when a huge amount of data is stored over the cloud servers in the form of the above-stored formats, extracting the desired record in the minimum time frame is a very challenging task. For efficient retrieval of information, it is used to decentralize the huge amount of database, and one such kind of technique is the blockchain technology which is used in the present work.

The main disadvantage of a centralized database which, is connected across the network is that the whole system may be lost/stolen if something goes wrong with the database [2]. To tackle such kind of problem, the approach of a centralized database is converted into a decentralized database and further, there may be numerous approaches for extracting information from the decentralized database, but for faster access of a decentralized database, one such technology has evolved as a blockchain technology, which has become very popular in the recent years and a better way to manage the database. The definition of blockchain technology is given below:

"Blockchain technology is defined as a sharing of records of patients which facilitates to the specialists for processing of the transaction and tracking patient's information across the medical professionals connected through the high-speed network system".

In 1991, Stuart Haber and W. Scott Stornetta introduced the concept of Blockchain and discussed "How to Time-Stamp a Digital Document" and developed a system for timestamping digital documents, ensuring integrity and immutability [3]. The concepts were expanded upon years later, in 2008, by a group of individuals under the pseudonym Satoshi Nakamoto in innovative whitepaper on "Bitcoin: A Peer-to-Peer Electronic Cash System." Nakamoto proposed blockchain as the foundation of Bitcoin, the first cryptocurrency. It was the first time blockchain technology which has been used in practice, turning it from a theoretical idea into a workable solution for secure, open, and decentralized transactions [4]. The basic key characteristic of blockchain is a distributed ledger, and immutability, which is one of its core features that makes the blockchain incredibly reliable, secure and impossible for unauthorized person to change the data without being discovered. It uses various types of consensus mechanisms to confirm transactions and ensure network stability such as Proof of Stake (PoS), Proof of Work (PoW), and Practical Byzantine Fault Tolerance (PBFT) that work together to monitor and secure blockchain processes [5]. The main advantage of said technology is that when one block of information is not functioning, then the user may use another block available over the network, while the main disadvantage of using this technology is that the establishment of blockchain technology needs high-cost computations. Despite cryptocurrencies, it is frequently used in healthcare to secure patient records and enable the analysis of data. Blockchain has the potential to improve the treatment of chronic diseases like diabetes, where early identification and continued monitoring are important. Diabetes has become one of the world's most prevalent chronic diseases, impacting millions of individuals [6]. Therefore, early detection and intervention are necessary for successful diabetes management. The traditional approaches frequently fail to effectively predict the development of the disease because of factors such as data manipulation, a lack of real-time analysis, and patient privacy concerns. Hence, this study highlights the potential of blockchain technology for safe, real-time diabetes prediction, which addresses the drawbacks of conventional centralized EHR systems. It shows that the blockchain not only improves data security but also provides a scalable and accurate platform for data management, establishing it as an innovative step forward from previous approaches.

A brief review is conducted on the prediction of information about diabetic patients which is necessary for formulation of the research problem. The following table provides brief information on the collected research work related to the last five years:

The above literature mostly focuses on monitoring and storing the data. Still, it lacks predictive abilities based on age group distribution, essential for conducting focused assessments of diabetes patterns. To address the concerns, the study presents an innovative method using blockchain technology for the early prediction of diabetes, focusing on age-based data segmentation inside the blockchain to analyze diabetes patterns across age groups. Statistical analysis is also used to figure out the correlation between age and diabetes conditions. It also evaluates storage size, retrieval time, and data consistency to guarantee integrity and accuracy. This complete architecture makes the blockchain concept secure, efficient, and trustworthy for large-scale healthcare applications.

2 Methodology

In the present work, blockchain technology is used for predicting diabetes at an early stage via a newly designed system model that contains five different stages as shown in Figure 1.

The procedure begins with collection of datasets on diabetes patients from the EHR data center and then analyzed data followed by a data cleaning operation. Further, the data is divided according to age groups with a gap of 10 years each like 0 to 10 years is one group, 10 to 20 years is the second group, and so on. Furthermore, a blockchain is created in which group-wise data is stored like, the first group is stored in the first block, the second group is stored in the second block, and so on. The system verifies each block of data to ensure that it is correct, consistent, and has not been tampered with. In the last stage, prediction and analysis are done. The detailed proposed system model is depicted in Figure 2.

Table 1. Summary of review of research work

Reference	Methodology	Major contribution	Research gap (if any)
(7)	HealthMudra: a blockchain-based recommender system	The approach manages diabetic patient data and provides personalized health advice	Scalability issues for managing diabetes on a large scale
(8)	Rule-based clustering and feature selection-based adaptive neuro-fuzzy inference system (FS-ANFIS)	A proposed approach for forecasting diabetes and cardiac disease using blockchain and fog computing	Requires security method, and high prediction accuracy
(9)	Combination of blockchain and machine learning classifier	Handle medical data and diagnose diabetes	The security, storage size, or data retrieval time assessment is not provided
(10)	AI-based method, and dataset	Reviews 178 studies on artificial intelligence for diabetic retinopathy diagnosis	Low evaluation of practical implementation
(11)	IoT devices, Random Forest, and blockchain	Predict diabetes and integrate IoT and blockchain for storage	Limited analysis on real-time monitoring
(12)	Combining IoT sensors with Hyperledger Fabric	Presented a scalable blockchain system with privacy controls for exchanging patient data	Lack of experimental assessment and implementation details
(13)	Integration of digital twins, blockchain, and AI	Proposed approach to enhance public health and data management	Lacks practical implementation and privacy solutions
(14)	Analysis of blockchain's possibilities for payment, credit systems, and secure transactions	Improve data security, and transaction efficiency, and eliminate intermediary dependency	Imitated discussion on privacy laws, and real-world implementation
(15)	Integration of lockchain, and federated learning	A privacy-preserving system for personalized diabetes prediction was created	Scalability issues, lack of standard data models
(16)	Integration of IoT, blockchain, and machine learning (AdaBoost)	Developed system for diabetes categorization and achieved 92.64% accuracy	Limited analysis into advanced machine learning models
(17)	IoT, edge computing, blockchain, and machine learning (RF, LR, SVM) are all combined	A safe and effective diabetes prediction system was created that improved feature selection	Execution time was not calculated and Scalability issues
(18)	Deep learning and TensorFlow	AI-based early diagnosis and evaluation for diabetic retinopathy	Limited practical evaluation, need for optimized models
(19)	AE-based diagnostics, SIMON encryption, and GTOA key generation are all combined	A reliable, and effective structure was created for medical data and diagnosis	The VAE model's optimization and learning rate schedules were not properly examined
(20)	Combining neural networks, KNN, fuzzy logic, and MLR	A hybrid approach was developed for accurate diabetes dosage and obtained 94% accuracy	Real-world implementations were not fully examined
(21)	Combining machine learning models (RF, DT, KNN, Naïve Bayes, ANN) with blockchain	Blockchain improves disease prediction and provides accurate diagnostic results	Lack of real-time implementation

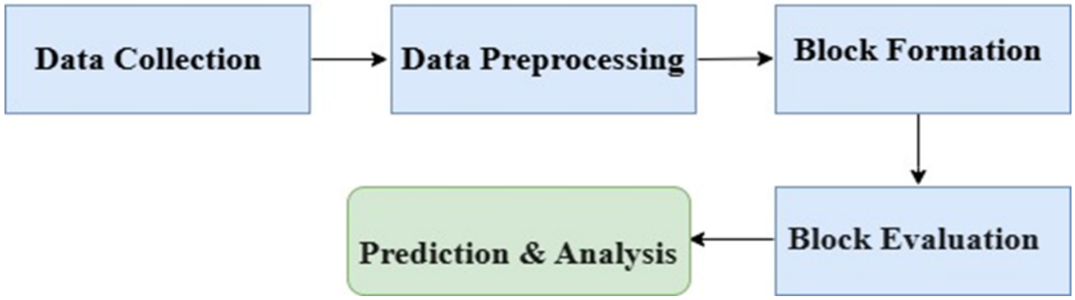

Fig 1. Basic block diagram of the proposed model

2.1 Data collection

A dataset containing information on diabetic patients is collected from the Kaggle repository named as "Diabetes Prediction Dataset"[(22)], and publicly available at link: https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset. This includes the patient's medical record, demographical information, and diabetes status (positive or negative). The dataset has several attributes, including age, gender, heart condition, Body Mass Index (BMI), hypertension, smoking history, HbA1c level, and blood glucose level, displayed in Table 2. There are 100,000 records in the collection and may be used to create a machine-learning model that predicts diabetes in individuals based on the medical history of patients and demographics, however, in the present study, a concept of blockchain technology is used to predict diabetes at an early stage of the life of the patient.

Table 2. Description of dataset

Feature	Description	Index
Gender	Total ratio of Male (41%) and Female (59%)	1
Age	The person's age in years	2
Hypertension	Shows whether the person has hypertension (1 = yes, 0 = no)	3
Heart_disease	Indicates whether the person has a history of heart disease (1 = yes, 0 = no)	4
Smoking history	The person's smoking history (never = 35%, no info = 36%, others = 29%)	5
BMI	Body fat is a measured by body mass index based on height and weight	6
HbA1c_level	HbA1c (Hemoglobin A1c), indicates average blood sugar levels during the previous 2-3 months	7
Blood_glucose_level	The blood glucose level of the person was measured at the time of the test	8
Diabetes	Determines whether a person has been diagnosed with diabetes (1 if yes, 0 if no)	9
Total collected record = 1,00,000		

2.2 Data preprocessing

Data preparation is an essential process that involves cleaning and processing raw data to make it suitable for analysis. Therefore, before utilizing the collected data, it goes through a cleaning procedure to eliminate any duplicate or redundant information. To minimize the size of datasets, duplicate records are eliminated. Next, irrelevant columns and missing or null values are deleted to get the best results. Hence, the dataset was pre-processed using the following steps:

2.2.1. Duplicate removal

The dataset contains some redundant records which may cause overfitting and impact the model performance. In this study, 3,854 duplicate entries were detected and eliminated from the initial dataset of 1,000,000 records, resulting in a cleaned dataset of 96,146 records. This phase guarantees that each record is unique and contributes significantly to the study.

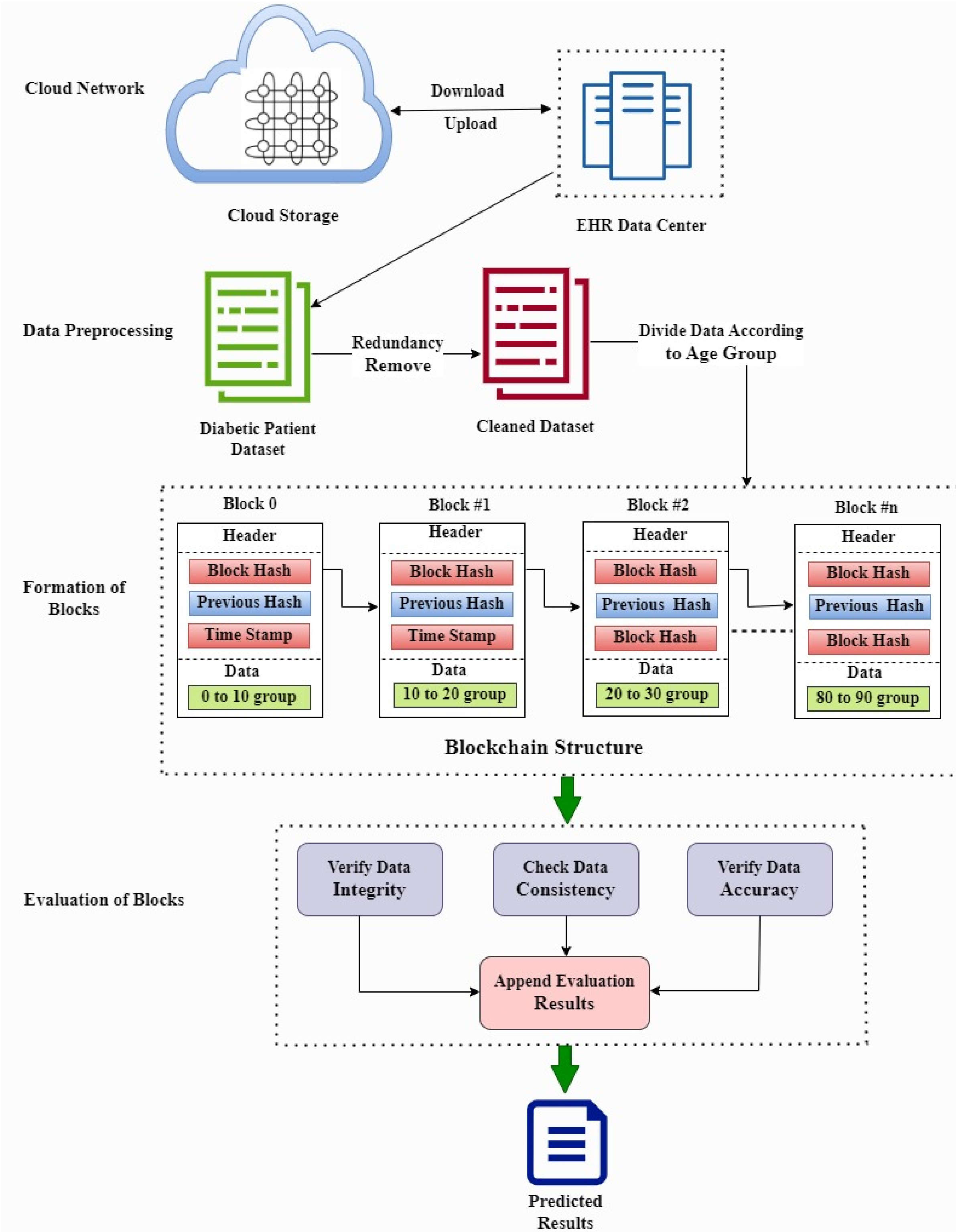

Fig 2. Proposed system model

2.2.2. Managing null or missing values

Missing and null data are common problems in datasets, that frequently appear as blank spaces. If not addressed properly, it creates mistakes or reduce the model accuracy. The difficulties may be addressed using approaches such as imputation, deletion, or substitution. In the present study, the dataset was thoroughly analysed, and no missing or null values were discovered. As a result, the process moved to the next stage.

2.2.3. Remove irrelevant columns

The columns with no variation, with a single value that did not improve performance are found and eliminate such columns to make the dataset smaller and more manageable.

The above-mentioned preprocessing methods were selected because of getting solution of common issues in data while boosting overall system performance. Data integrity is ensured by eliminating irrelevant columns, addressing missing values, and eliminating duplicates, leading to more precise and effective models. When taken as a whole, the steps strengthen the blockchain-based diabetes prediction model. The pseudocode for data collection and data preprocessing is given below:

Pseudocode 1: dcdp() /*data collection and data preprocessing*/

```
begin
function collect_and_preprocess_data():
diabetic_dataset = download_data_from_cloud()
# Initial size of dataset
total_records = 100000
# Remove redundant entries
duplicates_removed = 3854
cleaned_dataset_size = total_records - duplicates_removed
cleaned_dataset = remove_redundancies(diabetic_dataset)
# Verify the cleaned dataset size is accurate
 assert size(cleaned_dataset) == cleaned_dataset_size
# Divide the cleaned dataset into age groups
grouped_dataset = divide_by_age_group(cleaned_dataset)
return grouped_dataset
end
```

2.3 Formation of blocks

After the above steps, the cleaned data is divided into groups according to age like 0-10 years, 10-20 years, and so on to make it more particular and useful for assessment.

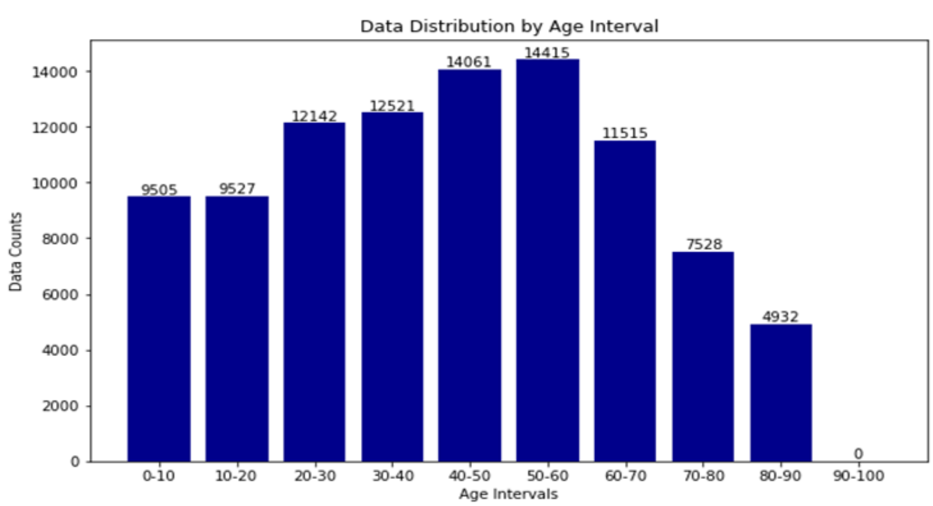

Fig 3. Categorization of the cleaned dataset as per age group

As shown in Figure 3 the data is distributed by age group. The first block provides data for the first age group, the second block has data for the next 10 to 20 years, and so on. Hence, the blockchain structure created in Figure 2 as shown above consists of the following components:

2.3.1. Creating blocks

The blockchain structure is made by connecting individual blocks one by one, which makes a chain, that is why it is called a blockchain. Each block is divided into two components: the first one is the block header, and the second is the block body. The list of transactions or data is recorded in the block body, and other things, such as its hash, timestamp, and the previous block hash are all stored in the header. The first block in the chain is known as the "Genesis" block. This structure makes it extremely difficult to modify any information because changing one block would need changing all the subsequent blocks, which is nearly impossible. After creating blocks, each age group data is stored in a block on the blockchain.

2.3.2. Block hash

When it comes to protecting data integrity from tampering or illegal access, cryptographic hash functions like SHA-256 are used that provide block identification. Each block can be identified by its hash. Even a small change in data will produce an entirely different hash value. Hence, block hash is calculated using the formula:

$$H_i = \text{SHA256}\left(index_i \parallel previous\_hash_{i-1} \parallel data_i\right) \tag{1}$$

2.3.3. Previous block hash

It provides a reference to the hash of the preceding block, ensuring the immutability of the blockchain. The previous block hash guarantees that blocks always appear in the right sequence. If someone attempts to change the blocks, the hashes will not match, and the chain will break. This maintains the blockchain organized and reliable.

2.3.4. Timestamp

It assigns a time or date of creation to a block. It is a string of characters that runs continuously and updates every second so that we can know how much time any block took to be created.

2.3.5. Data

Each block under the blockchain contains data on diabetic patients for a certain age group (e.g., 0-10 years, 10-20 years, etc). The Figure 3 displays that the first block has 9505 data, the second block contains 9527 data, and so on.

Pseudocode 2:bf_bc() /*blocks formation under blockchain */

```
begin
function create_blockchain(grouped_dataset):
 blockchain = initialize_blockchain()
for age_group in grouped_dataset:
  block = create_block()
# Assign age group data to block
block.data = age_group.data
block.timestamp = get_current_timestamp()
if block is first:
  set block.previous_hash = null
else:
  set block.previous_hash = hash_of_previous_block
block.hash = calculate_block_hash(block)
add_block_to_blockchain(blockchain, block)
return blockchain
end
```

2.4 Evaluation of blocks

At this stage, the system verifies the integrity of the data and guarantees consistency among the blocks. Furthermore, the system model ensures that the data correctly represents the real information. The pseudo-code for the security checks is given below:

Pseudocode 3:sc_bica() /* Security checks of blocks: Integrity, consistency, and accuracy*/

```
begin
function evaluate_blockchain(blockchain):
 for every block in blockchain_structure:
  verify data_integrity(block)
if not verify_block_integrity(block):
log_error("Data Integrity Check Failed", block)
 check data_consistency(block)
if not check_data_consistency(block):
log_error("Data Consistency Check Failed", block)
 verify data_accuracy(block)
 if not verify_data_accuracy(block):
 log_error("Data Accuracy Check Failed", block)
 # Store evaluation results
 append_evaluation_results(results, block)
 return results
end
```

2.5 Analysis and prediction

After the evaluation phase, the system examines the blockchain data to predict outcomes such as the number of normal, prediabetic, and diabetic patients in each group. For this purpose, the pseudo-code is given below:

Pseudocode 4: pr() /*Generate Predicted Results*/

```
begin
function generate_predictions(results):
 predictions = analyze_evaluation_results(results)
 return predictions
end
```

Diabetes is identified with an HbA1c test, and Table 3 shows the range of values that are used to show whether a person has normal blood sugar, diabetic, or pre-diabetic [23]. The HbA1c test, also known as the glycated hemoglobin test, determines the average blood sugar level over the previous two or three months by evaluating the percentage of hemoglobin coated with sugar (glycated). When glucose levels are high, more hemoglobin becomes glycated. This test is essential for diagnosing diabetes and prediabetes because it provides a longer-term view of blood sugar control than regular glucose tests. Higher HbA1c values indicate more severe blood sugar levels and an increased risk of diabetic complications, making it an important tool for detecting and treating the condition of diabetes. Therefore, the prediction over the collected dataset shall be considered according to the following Table 3.

Table 3. Test range of HbA1c for prediction of diabetic patient

Category of diabetic patient	Range of HbA1c
Normal	Below 42 mmol/mol (6.0%)
Pre-diabetic	42 to 47 mmol/mol (6.0 to 6.4%)
Diabetic	48 mmol/mol (6.5% or over)

2.6 Correlation coefficient

A correlation coefficient is a statistical tool used for determining the direction and strength of the association between two variables. Various types of correlation coefficients are present such as Pearson, Spearman, and Kendall rank correlation coefficients. The present work uses the Pearson correlation coefficient to determine a linear correlation between two variables. The range of correlation is -1 to 1 and is described below briefly:

- **-1:** Shows a complete negative connection. As one variable increases, the other decreases automatically;
- **0:** Shows the absence of correlation between the variables;
- **1:** Indicates a complete positive relationship. As one variable increases, the other also increase accordingly.

The formula for Pearson's correlation coefficient *r* is:

$$r = \frac{n\left(\sum xy\right) - \sum(x)\sum(y)}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\left[n\sum y^2 - \left(\sum y\right)^2\right]}} \tag{2}$$

where *r* is the correlation coefficient, a and y are two variables and n are the number of data points.

3 Results and Discussion

The proposed system model examined performance metrics on a device with specified characteristics such as Intel(R) Core(TM) i5-11260H, 2.60GHz with 8 GB RAM, and a 64-bit Windows 11 operating system. Python 3.10.9 using Jupyter Notebook and Spyder IDE is used for programming purpose. The details of the experimental results, security analysis, performance evaluations, and statistical correlation between three categories (normal, prediabetic, and diabetic) patients are explained below in brief:

3.1 Experimental results

The proposed model is focused on early-stage diabetes prediction with the help of blockchain technology and the findings are contained in Table 4. Further, the number of persons of various ages fits into three categories depending on blood sugar levels which are categorized as normal, prediabetes, and diabetes, as explained below. The results are based on HbA1c test ranges, which makes more trustworthy. The ranges give important information for predicting and controlling diabetes, allowing both patients and doctors to take preventative steps. Unlike standard blood sugar tests, which indicate glucose levels at a single moment in time, the HbA1c test provides a more accurate picture of how effectively blood sugar is managed over time. Table 3 displays the ranges.

Table 4. Categorization of the diabetic patients according to age group

Group	Normal	Prediabetes	Diabetes	Total record
0-10	5481	2330	1694	9505
10 -20	5453	2373	1701	9527
20 -30	7101	2950	2091	12142
30 -40	7065	3106	2350	12521
40 -50	7800	3486	2775	14061
50 -60	7585	3542	3288	14415
60 -70	5787	2750	2978	11515
70 -80	3640	1898	1990	7528
80 -90	2469	1227	1236	4932

3.1.1. Normal

The second column of Table 4 represents the number of people in each age group with normal blood sugar level. Hence, the table shows that the percentage of people with normal blood sugar level gradually decreases with age.

Furthermore, to identify a correlation between age data and normal patients. Authors first give numerical values to age groups, as the groups consist of ranges and are represented by midpoints. For example: from 0 to 10 the midpoint is 5, from 10 to 20 the midpoint is 15, from 20 to 30 the midpoint is 25, and so on. After that, the above **equation (2)** is used to determine the linear relationship between the two variables i.e. the age group midpoints are treated as x and the values in the Normal column are treated as y.

Therefore, the computed correlation is -0.492, which indicates a moderate negative correlation between age and normal blood sugar levels. This proves that as age increases, the number of individuals with normal blood sugar decreases.

3.1.2. Pre-diabetic

The third column of Table 4 represents the number of persons in each age group who have higher blood sugar levels but are not classified as diabetic. Pre-diabetic also rises with age, especially between 30 and 60 years, indicating an age-related risk of developing diabetes.

Furthermore, to identify a correlation between age and pre-diabetic categories, a similar process is used as described earlier. The correlation between age and the number of pre-diabetics is around -0.3504, showing a weak negative association. This shows that the number of pre-diabetic people decreases as age increases, the association is not as strong as the one between age and normal blood sugar levels.

3.1.3. Diabetic

The last column of Table 4 shows the number of persons diagnosed with diabetes by age group. The probability of diabetes often increases with older age, as seen by data, with a major increase in diabetes cases between the ages of 40 to 70, particularly highest in the 50-60 age categories, as represented in Table 4.

Moreover, to compute the correlation between age and diabetic condition, the relationship between age and diabetes patients is computed through similar process. The correlation between age and the number of diabetics is 0.1185, showing a very weak positive relationship. It indicates that the number of diabetic patients increases with age, although the association is not strong.

3.2 Security analysis and performance evaluation

Blockchain offers a secure and reliable foundation for data storage and management. Some crucial safety measures and performance assessment criteria are applied to guarantee the dependability, efficiency, and security inside a blockchain system. Additionally, it covers critical factors such as storage size, retrieval time, data integrity, consistency, and accuracy which often overlooked in other studies.

*3.2.1. **3.2.1 Storage size***

Storage capacity of each block is evaluated in megabytes (MB) which is used to evaluate the scalability of the blockchain system. This metric is essential to determine the system capability to handle growing data. The storage size is calculated using the following formula:

$$\text{Storage Size (MB)} = \frac{\text{Total Size (bytes)}}{(1024 \times 1024)} \tag{3}$$

Table 5 shows that the storage size varied among blocks, from 0.852530 MB to 2.489556 MB, depending on the data volume in each block.

Table 5. Performance metrics evaluation for different blocks

Block Index	Storage Size (MB)	Retrieval Time (μs)
1	1.655451	176.900066
2	1.646116	112.799928
3	2.097990	134.400092
4	2.163375	55.800192
5	2.428307	178.300310
6	2.489556	66.400040
7	1.989020	116.600189
8	1.301023	57.499856
9	0.852530	112.200156

3.2.2. Retrieval time

Retrieval time is also evaluated in microseconds (s) to determine how well the system performed when accessing and retrieving stored data. The metric gives important information about the overall speed of the system. The retrieval time was determined using this formula.

$$\text{Retrieval Time } (\mu s) = (\text{End Time (s)} - \text{Start Time (s)}) \times 1,000,000 \tag{4}$$

The retrieval time for blocks ranged between 55.800192 μs and 178.300310 μs, indicating quick access as more data was uploaded to the blockchain, as shown in Table 5.

Additionally, in order to highlight the differences, Table 6 provides a comparison between previous studies and the present work.

Table 6. Comparison of various approaches with present work

Crucial factors	References					
	(17)	(18)	(19)	(20)	(21)	Present
Age-specific blockchain structure	No	No	No	No	No	Yes
Statistical correlation	Yes	No	No	No	Yes	Yes
Storage size	No	No	No	No	No	Yes
Retrieval time	No	No	No	No	No	Yes
Integrity	Yes	No	No	No	No	Yes
Consistency	No	No	No	No	No	Yes
Accuracy	Yes	Yes	Yes	Yes	No	Yes

3.2.3. Data integrity, consistency, and accuracy

Maintaining integrity, consistency, and correctness is critical for ensuring data security and dependability in a blockchain system. Data integrity ensures that the information saved on the blockchain is unchanged and safe. This is verified by recalculating each block's hash and comparing it to the stored hash, with identical results indicating that no tampering happened. The check of data integrity check performed using **equation (1)**. Cryptographic hash algorithms, consensus procedures, and immutability are all critical techniques for establishing data integrity. Similarly, data consistency guarantees that information across all blocks is uniform and error-free. The transaction validation is performed with the help of consensus mechanisms which play an important part in ensuring consistency of data. Any variations in the chain compromise its dependability and show invalidity. Furthermore, data accuracy guarantees that the information saved on the blockchain exactly matches the original input data. Successful verification ensures that the obtained information matches with the original data exactly. Figure 4 illustrates verification of data integrity, consistency, and accuracy across all blocks.

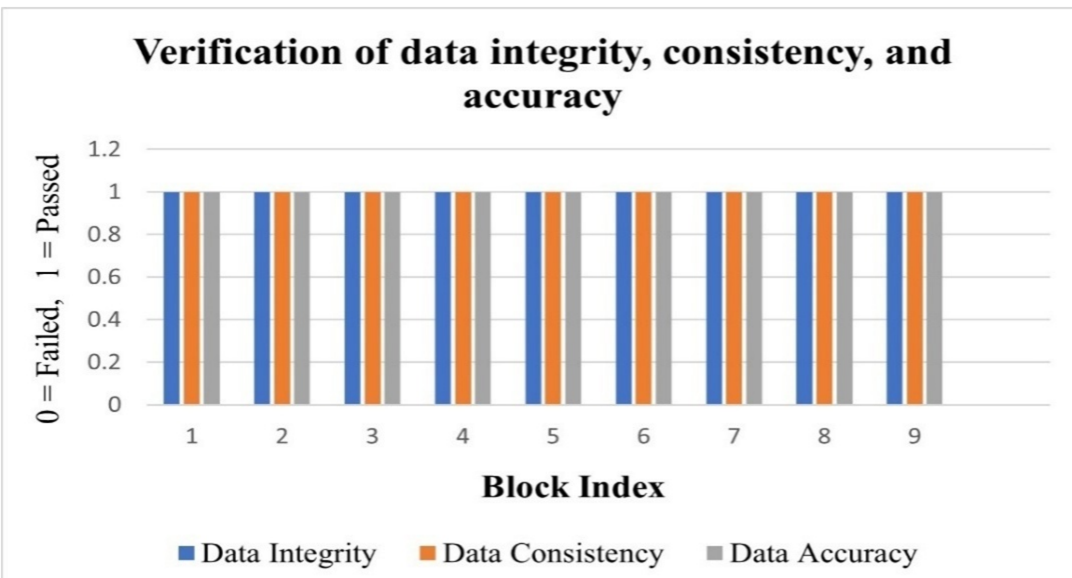

Fig 4. Verification of all security checks for different blocks

4 Conclusions and Future Scope

From the above work it is concluded that the present work is related to the application of blockchain technology in revolutionizing diabetes prediction. Firstly, patients are divided according to age groups, thereafter, three categories of patients with the help of the HbA1c test range are identified and categorised as normal, prediabetes, and diabetes. Further, the results of the present work are based on age-specific groups which represent that diabetes is becoming a more common disease among human beings with increasing age, especially in the age of 50 to 60 years. Additionally, the system model enhances data security, integrity, and accessibility and computes the storage size and retrieval time of each block. Block six has the highest 2.489556 mb storage size because it has the maximum number of diabetic patients, and the minimum retrieving time is 55.800192μs. The proposed method addresses the constraints of traditional healthcare data systems, highlighting the significance of early identification. Future research should look into combining powerful machine learning algorithms with blockchain technology to improve prediction models and increase the scalability of such systems in diverse healthcare contexts. Random Forest (RF) may help with feature selection, whereas Support Vector Machine (SVM) may accurately classify patients based on medical information. Artificial Neural Networks (ANNs) may enhance deep learning-based diabetes prediction, while K-Nearest Neighbors (KNNs) may be used to diagnose patients based on similarity. Additionally, Gradient Boosting (XGBoost) may improve prediction accuracy via ensemble learning. There are some additional machine learning methods that may be implemented in the future.

References

1) Cichosz SL, Stausholm MN, Kronborg T, Vestergaard P, Hejlesen O. How to use blockchain for diabetes health care data and access management: an operational concept. *Journal of diabetes science and technology*. 2019 Mar;13(2):248–253. https://doi.org/10.1177/1932296818790281.
2) Niranjanamurthy M, Nithya BN, Jagannatha SJ. Analysis of Blockchain technology: pros, cons and SWOT. *Cluster Computing*. 2019 Nov;22:14743–14757. https://doi.org/10.1007/s10586-018-2387-5.
3) Haber S, Stornetta WS. How to time-stamp a digital document. Springer Berlin Heidelberg. 1991. https://doi.org/10.1007/BF00196791.
4) Yaqoob S, Khan MM, Talib R, Butt AD, Saleem S, Arif F, et al. Use of blockchain in healthcare: a systematic literature review. *International journal of advanced computer science and applications*. 2019;10(5). https://doi.org/10.14569/IJACSA.2019.0100581.
5) Uddin MA, Stranieri A, Gondal I, Balasubramanian V. Block chain leveraged decentralized IoT e Health framework. *Internet of Things*. 2020 Mar;9:100159. https://doi.org/10.1016/j.iot.2020.100159.
6) Mohsen F, Al-Absi HR, Yousri NA, Hajj NE, Shah Z. A scoping review of artificial intelligence-based methods for diabetes risk prediction. *NPJ Digital Medicine*. 2023 Oct;6(1):197. https://doi.org/10.1038/s41746-023-00933-5.
7) Bhardwaj R, Datta D. Development of a recommender system HealthMudra using blockchain for prevention of diabetes. *Recommender system with machine learning and artificial intelligence: practical tools and applications in medical, agricultural and other industries*. 2020 Jun;p. 313–327. https://doi.org/10.1002/9781119711582.ch16.
8) Shynu PG, Menon VG, Kumar RL, Kadry S, Nam Y. Blockchain-based secure healthcare application for diabetic-cardio disease prediction in fog computing. *IEEE Access*. 2021 Mar;9:45706–45720. https://doi.org/10.1109/ACCESS.2021.3065440.
9) Chen M, Malook T, Rehman AU, Muhammad Y, Alshehri MD, Akbar A, et al. Blockchain-Enabled healthcare system for detection of diabetes. *Journal of Information Security and Applications*. 2021 May;58:102771. https://doi.org/10.1016/j.jisa.2021.102771.
10) Bidwai P, Gite S, Pahuja K, Kotecha K. A Systematic Literature Review on Diabetic Retinopathy Using an Artificial Intelligence Approach. *Big Data Cogn Comput*. 2022;6:152. https://doi.org/10.3390/bdcc6040152.
11) Azbeg K, Boudhane M, Ouchetto O, Andaloussi SJ. Diabetes emergency cases identification based on a statistical predictive model. *Journal of Big Data*. 2022 Mar;9(1):31. https://doi.org/10.1186/s40537-022-00582-7.
12) Khan D, Jung LT, Hashmani MA, Cheong MK. Blockchain Enabled Diabetic Patients' Data Sharing and Real Time Monitoring. *CS&IT Conference Proceedings*. 2022 Mar;12(6). https://dx.doi.org/10.5121/csit.2022.120620.
13) Novi GD, Sofia N, Vasiliu-Feltes I, Zang CY, Ricotta F. Blockchain Technology Predictions 2024: Transformations in Healthcare, Patient Identity, and Public Health. *Blockchain in Healthcare Today*. 2023;6. https://doi.org/10.30953/bhty.v6.287.
14) Choudhary B, Saxena V. Blockchain Implementation for Faster Accessing of Database of Banking Industry. *International Journal on Recent and Innovation Trends in Computing and Communication*. 2023 Sep;7:504–512. Available from: https://www.researchgate.net/publication/379986243_Blockchain_Implementation_for_Faster_Accessing_of_Database_of_Banking_Industry.
15) Hasan MR, Li Q, Saha U, Li J. Decentralized and Secure Collaborative Framework for Personalized Diabetes Prediction. *Biomedicines*. 2024;12:1916. https://doi.org/10.3390/biomedicines12081916.
16) Ratta P, Sharma S. A blockchain-machine learning ecosystem for IoT-Based remote health monitoring of diabetic patients. *Healthcare Analytics*. 2024 Jun;5:100338. https://doi.org/10.1016/j.health.2024.100338.
17) Hennebelle A, Ismail L, Materwala H, Kaabi JA, Ranjan P, Janardhanan R. Secure and privacy-preserving automated machine learning operations into end-to-end integrated IoT-edge-artificial intelligence-blockchain monitoring system for diabetes mellitus prediction. *Computational and Structural Biotechnology Journal*. 2024 Dec;23:212–233. https://doi.org/10.1016/j.csbj.2023.11.038.
18) Ghouali S, Onyema EM, Guellil MS, Wajid MA, Clare O, Cherifi W, et al. Artificial intelligence-based teleopthalmology application for diagnosis of diabetics retinopathy. *IEEE Open Journal of Engineering in Medicine and Biology*. 2022 Jul;3:124–133. https://doi.org/10.1109/OJEMB.2022.3192780.
19) Sammeta N, Parthiban L. Hyperledger blockchain enabled secure medical record management with deep learning-based diagnosis model. *Complex & Intelligent Systems*. 2022 Feb;8(1):625–640. https://doi.org/10.1007/s40747-021-00549-w.
20) Laxmi N, Mohana J. Blockchain-based hybrid method for diabetic management on daily basis through insulin dosage prediction. *Soft Computing*. 2023 Jul;p. 1–8. https://doi.org/10.1007/s00500-023-08932-0.
21) Imran A, Malik KR, Khan AH, Sajid M, Arslan M. Methodology for Ensuring Secure Disease Prediction using Machine Learning Techniques. *Journal of Computing & Biomedical Informatics*. 2024 Jun;7(1):15–25. https://doi.org/10.56979/701/2024.
22) Mustafa I. *Diabetes prediction dataset*. 2023.
23) Blood sugar level ranges. *Diabetescouk*.