# A Novel Hybrid Algorithm for Securing Data Transfer with machine Learning Disease Prediction

Varalakshmi K[1], Vasantha Raja S S[2], Vijaya Narayanan A[3], Saravanan S[4]
[1]*Associate Professor, Computer Science and Engineering, PERI Institute of Technology*
[2,3,4]*Assistant Professor, Computer Science and Engineering, PERI Institute of Technology*

***Abstract*** **- Data security refers to the process of protecting data from unauthorized access and data corruption throughout its life cycle. Data security includes data encryption, hashing, tokenization, and key management practices that protect data across all applications and platforms. The security system used nowadays uses data encryption software to effectively enhance data security by using an algorithm (called a cipher) and an encryption key to turn normal text into encrypted cipher text. To an unauthorized person, the cipher data will be unreadable. That data can then be decrypted only by a user with an authorized key. Whereas with the improving data insecurity nowadays leads to loss of confidential data as the key is easily hack able because of a single algorithm usage. Diabetes-related complications include damage to large and small blood vessels, which can lead to heart attack and stroke, and problems with the kidneys, eyes, feet and nerves. The risk of most diabetes-related complications can be reduced if diagnosed early. To overcome this problem this project presents a medical application that accepts and analyses a patient's medical data to give a diagnosis to check if he/she is diabetic with an efficient data security system where two security algorithms will be merged to secure the patient's medical data stored and accessed in cloud. In addition, the emerging block chain technologies with the wireless data transfer system, which makes easy the interaction between the data and cloud. The medical data is analyzed and the result is stored securely in a cloud database such as mongoDB as an highly secure encrypted key. The result is received from the database and decrypted in the frontend to view the test results. Thus, this project presents an effective end to end security for medical applications.**

***Index Terms*** **- Data Security, Encryption, Hashing, Hackable, Token, Diabetes.**

## I.INTRODUCTION

Data security refers to the process of protecting data from unauthorized access and data corruption throughout its lifecycle. Data security includes data encryption, hashing, tokenization, and key management practices that protect data across all applications and platforms.

Organizations around the globe are investing heavily in information technology (IT) cyber security capabilities to protect their critical assets. Whether an enterprise needs to protect a brand, intellectual capital, and customer information or provide controls for critical infrastructure, the means for incident detection and response to protecting organizational interests have three common elements: people, processes and technology.

Diabetes is fast gaining the status of a potential epidemic in India with more than 62 million diabetic individuals currently diagnosed with the disease. In 2000, India (31.7 million) topped the world with the highest number of people with diabetes mellitus followed by China (20.8 million) with the United States (17.7 million) in second and third place respectively. According to Wild et al. the prevalence of diabetes is predicted to double globally from 171 million in 2000 to 366 million in 2030 with a maximum increase in India. It is predicted that by 2030 diabetes mellitus may afflict up to 79.4 million individuals in India, while China (42.3 million) and the United States (30.3 million) will also see significant increases in those affected by the disease. India currently faces an uncertain future in relation to the potential burden that diabetes may impose upon the country. Many influences affect the prevalence of disease throughout a country, and identification of those factors is necessary to facilitate change when facing health challenges. So what are the factors currently affecting diabetes in India that are making this problem so extreme?

The aetiology of diabetes in India is multifactorial and includes genetic factors coupled with environmental

influences such as obesity associated with rising living standards, steady urban migration, and lifestyle changes. Yet despite the incidence of diabetes within India, there are no nationwide and few multi-centric studies conducted on the prevalence of diabetes and its complications. The studies that have been undertaken are also prone to potential error as the heterogeneity of the Indian population with respect to culture, ethnicity, socio- economic conditions, mean that the extrapolation of regional results may give inaccurate estimates for the whole country

## II.LITERATURE REVIEW AND PREVIOUSWORK

Defending against malicious attacks has become increasingly important in various cyber- physical systems. This paper presents an encryption-based countermeasure against stealthy attacks on remote state estimation. Smart sensors transmit data to a remote estimator through a wireless communication network, in which data packets can be intercepted and compromised by attackers. The remote end is equipped with a false data detector that monitors the system. To avoid being detected, the attack should follow the stealthiness constraint. A linear encryption scheme is proposed to reduce the influence of potential stealthy attacks. For arbitrary linear encryption, the worst-case linear attack that yields the largest estimation error is derived. Accordingly, the optimal linear encryption, which minimizes the worst-case estimation error, is designed based on the Stackelberg game analysis. The above optimal strategies are considered in both the complete and partial measurement information scenarios for the attacker. Moreover, the generalization to nonlinear encryption strategies is also discussed. Comparisons of attack and encryption strategies through numerical examples are provided to illustrate the theoretical results.

Comparisons of attack and encryption strategies through numerical examples are provided to illustrate the theoretical results. But it is not suitable for encryption and decryption based application.

A control system based on Blockchain and electronic devices is adequate in this environment, thanks to the ability of Blockchain to provide trust. The results show that the water control system is not very secure. This method is extendable to other Mendelian-based and genetically influenced diseases. The SVM algorithm is not suitable for large datasets. The AES algorithm uses a too simple algebraic structure, making it vulnerable to hacking.

The results demonstrate that credit based PoW mechanism and data access control are secure and efficient in IoT. In this system DAG is used, no node in the graph can reference back to itself. This framework may likewise ensure the protection of the patients and moreover keeps up the security and trustworthiness of the health care system. The SHA algorithms used in here can only be used to store low entropy data.

## III.PROPOSEDSYSTEM

Data security refers to the process of protecting data from unauthorized access and data corruption throughout its lifecycle. Data security includes data encryption, hashing, tokenization, and key management practices that protect data across all applications and platforms. The security system used nowadays uses Data encryption software to effectively enhance data security by using an algorithm (called a cipher) and an encryption key to turn normal text into encrypted ciphertext. To an unauthorized person, the cipher data will be unreadable. That data can then be decrypted only by a user with an authorized key. Whereas with the improving data insecurity nowadays leads to loss of confidential data as the key is easily hackable because of a single algorithm usage. Diabetes-related complications include damage to large and small blood vessels, which can lead to heart attack and stroke, and problems with the kidneys, eyes, feet and nerves. The risk of most diabetes-related complications can be reduced if diagnosed early. To overcome this problem this project presents a medical application that accepts and analyses a patient‘s medical data to give a diagnosis to check if he/she is diabetic with an efficient data security system where two security algorithms will be merged to secure the patient‘s medical data stored and accessed in cloud. In this project logistic regression algorithm, a regression machine learning algorithm is used to process the entered patient‘s medical data and to predict if the patient is diabetic or normal. A web application using reactJS is developed from which the patient‘s medical data which will be encrypted using the hybrid secure algorithms before being uploaded to the database. In addition, the emerging block chain technology with the wireless data transfer system,

makes easy the interaction between the data and cloud. The medical data is analysed and the result is stored securely in a cloud database such as mongoDB as an highly secure encrypted key. The result is received from the database and decrypted in the frontend to view the test results. Thus, this project presents an effective end to end security for medical applications.

## IV.ENCRYPTIONALGORITHMS

JSON Web Token (JWT)

JWT is an open standard (RFC 7519) that defines a compact and self-contained way for securely transmitting information between parties as a JSON object. This information can be verified and trusted because it is digitally signed. JWTs can be signed using a secret (with the HMAC algorithm) or a public/private key pair using RSA orECDSA.

Although JWTs can be encrypted to also provide secrecy between parties, we will focus on signed tokens. Signed tokens can verify the integrity of the claims contained within it, while encrypted tokens hide those claims from other parties. When tokens are signed using public/private key pairs, the signature also certifies that only the party holding the private key is the one that signed it.

Cipher AES algorithm

The more popular and widely adopted symmetric encryption algorithm likely to be encountered nowadays is the Advanced Encryption Standard (AES). It is found at least six times faster than triple DES.

A replacement for DES was needed as its key size was too small. With increasing computing power, it was considered vulnerable against exhaustive key search attack. Triple DES was designed to overcome this drawback but it was found slow.

The simplest type of artificial neural network. With this type of architecture, information flows in only one direction, forward. It means, the information's flows start at the input layer, goes to the "hidden" layers, and end at the output layer. The network does not have a loop. Information stops at the output layers.

SHA-256 cryptographic hash algorithm

A cryptographic hash (sometimes called ‗digest') is a kind of ‗signature'for a text or a data file. SHA-256 generates an almost-unique 256-bit (32-byte) signature for a text.

A hash is not ‗encryption' – it cannot be decrypted back to the original text (it is a ‗one-way' cryptographic function, and is a fixed size for any size of source text). This makes it suitable when it is appropriate to compare hashed 'version softexts, as opposed to decrypting the text to obtain the original version.

SHA-256 is one of the successor hash functions to SHA-1 (collectively referred to as SHA- 2), and is one of the strongest hash functions available. SHA-256 is not much more complex to code than SHA-1, and has not yet been compromised in any way. The 256-bit key makes it a good partner-function for AES. It is defined in the NIST (National Institute of Standards and Technology) standard ‗'. NIST also provide a number of test vectorsto verifycorrectness ofimplementation.

## V.METHODOLOGY DISEASE PREDICTION USING MACHINELEARNING

In this project logistic regression a machine learning algorithm is used to predict if the patient is normal or diabetic based on the input medical data.

Diabetes mellitus, commonly known as diabetes, is a metabolic disease that causes high blood sugar. The hormone insulin moves sugar from the blood into your cells to be stored or used for energy.

Logistic Regression statistical method is used for analyzing the dataset and produces a binary outcome. One or more autonomous variables may have consisted of the dataset. The result is determined by these variables that are dichotomous in nature. Which means only two results are possible. It is a specific category of regression and it is used in the best way to predict the binary and categorical output. Logistical Regression method is used to regulate the impact of numerous autonomous variables which are conferred at the same time. This method also predicts any one of the two independent categories of variables. Logistic regression designs the best-fitting function with the help of the maximum likelihood method in order to maximize the probability of classifying the recognized data into the proper division. Various appliances of logistic regression are forecast market trends, to find the success and failure rates in results, the true or false category in recruiting employees based on their performance in need of employment in a company, image categorization, health care and analyze a group of people affected by Myocardial Infarction.

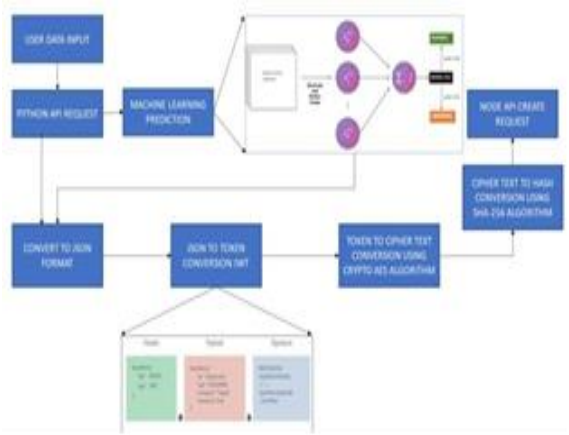

Figure 1.SystemArchitecture

Figure 2. Machine Learning Architecture

JWT GENERATION:

JSON Web Token is an Internet standard for creating data with optional signature and/or optional encryption whose payload holds JSON that asserts some number of claims. The tokens are signed either using a private secret or a public/private key. The tokens can be signed by one party's private key (usually the server's) so that party can subsequently verify the token is legitimate. If the other party, by some suitable and trustworthy means, is in possession of the corresponding public key, they too are able to verify the token's legitimacy. The tokens are designed to be compact, URL-safe, and usable especially in a web- browser single-sign-on (SSO)context.

JWT claims can typically be used to pass identity of authenticated users between an identity provider and a service provider, or any other type of claims as required by business processes. JWT relies on other JSON-based standards: JSON Web Signature and JSON Web Encryption.

JWT structure consists of three phases:

HEADER:

The header typically consists of two parts: the type of token, which is JWT, and the hashing algorithm that is used, such as HMAC SHA256 or RSA. Then, this JSON is Base64Url-encoded to form the first part of the JWT.

PAYLOAD:

The second part of the token is the payload, which contains the claims. Claims are statements about an entity.

Registered claims: These are a set of predefined claims which are not mandatory but recommended, to provide a set of useful, interoperable claims.

Public claims: These can be defined at will by those using JWTs. But to avoid collisions they should be defined

Private claims: These are the custom claims created to share information between parties that agree on using them and are neither registered or public claims.

SIGNATURE:

To create the signature part, you have to take the encoded header, the encoded payload, a secret, the algorithm specified in the header, and sign that.

Then, you have to put it all together. The following shows a JWT that has the previous header and payload encoded, and it is signed with asecret.


Figure 3. JWT Format

JWT TO CIPHER GENERATION

The Advanced Encryption Standard (AES) is a symmetric block cipher chosen by the U.S. Government to protect classified information. AES is implemented in software and hardware throughout the world to encrypt sensitive data. It is essential for government computer security, cyber security and electronic data protection. AES includes three block ciphers: AES-128, AES-192 and AES-256.AES-128 uses a 128-bitkey length to encrypt and decrypt a block

of messages, while AES-192 uses a 192-bit key length and AES- 256 a 256-bit key length to encrypt and decrypt messages. Each cipher encrypts and decrypts data in blocks of 128 bits using cryptographic keys of 128, 192 and 256 bits, respectively.

AES is used widely for protecting data at rest. Applications for AES include self- encrypting disk drives, database encryption and storage encryption. On the other hand, the RSA(Rivest-Shamir-Adleman) algorithm is often used in web browsers to connect to websites, in virtual private network (VPN) connections and in many other applications.

Symmetric, also known as secret key, ciphers use the same key for encrypting and decrypting, so the sender and the receiver must both know -- and use -- the same secret key. The government classifies information in three categories: Confidential, Secret or Top Secret. All key lengths can be used to protect the Confidential and Secret level. Top Secret information requires either 192- or 256-bit key lengths.

There are 10 rounds for 128-bit keys, 12 rounds for 192-bit keys and 14 rounds for 256- bit keys. A round consists of several processing steps that include substitution, transposition and mixing of the input plaintext to transform it into the final output of cipher text.

MERGING OF CIPHER AND SHA 256

The SHA-256 algorithm is one flavor of SHA-2 (Secure Hash Algorithm 2), which was created by the National Security Agency in 2001 as a successor to SHA-1. SHA-256 is a patented cryptographic hash function that outputs a value that is 256 bits long.

What is hashing? In encryption, data is transformed into a secure format that is unreadable unless the recipient has a key. In its encrypted form, the data may be of unlimited size, often just as long as when unencrypted. In hashing, by contrast, data of arbitrary size is mapped to data of fixed size. For example, a 512-bit string of data would be transformed into a 256-bit string through SHA-256 hashing.

In cryptographic hashing, the hashed data is modified in a way that makes it completely unreadable. It would be virtually impossible to convert the 256-bit hash mentioned above back to its original 512-bit form. So why would you want to create a scrambled message that can‘t be recovered? The most common reason is to verify the content of data that must be kept secret. For example, hashing is used to verify the integrity of secure messages and files. The hash code of a secure file can be posted publicly so users who download the file can confirm they have an authentic version without the contents of the file being revealed. Hashes are similarly used to verify digital signatures.

Password verification is a particularly important application for cryptographic hashing. Storing users‘ passwords in a plain-text document is a recipe for disaster; any hacker that manages to access the document would discover a treasure trove of unprotected passwords. That‘s why it‘s more secure to store the hash values of passwords instead. When a user enters a password, the hash value is calculated and then compared with the table. If it matches one of the saved hashes, it‘s a valid password and the user can be permitted access.

What role does SHA-256 hashing play in cyber security? SHA-256 is used in some of the most popular authentication and encryption protocols, including SSL, TLS, IPsec, SSH, and PGP. In Unix and Linux, SHA-256 is used for secure password hashing. Crypto currencies such as Bitcoin use SHA-256 for verifying transactions.

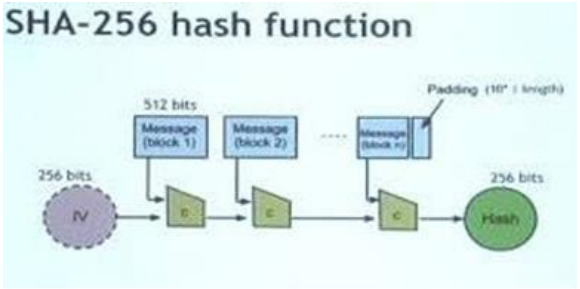


Figure 4. SHA-256 Format

NODE API GENERATION

In this project nodeJS is used for api generation.

An application programming interface (API) is a computing interface which defines interactions between multiple software intermediaries. It defines the kinds of calls or requests that can be made, how to make them, the data formats that should be used, the conventions to follow, etc. An application programming interface (API) is a computing interface which defines interactions between multiple software intermediaries. It defines the kinds of calls or requests that can be made, how to make them, the data formats that should be used, the conventions to follow, etc.

NodeJS is an open-source, cross-platform, back-end, JavaScript runtime environment that executes

JavaScript code outside a web browser. NodeJS lets developers use JavaScript to write command line tools and for server-side scripting—running scripts server-side to produce dynamic web page content before the page is sent to the user's web browser. Consequently, NodeJS represents a "JavaScript everywhere" paradigm, unifying web-application development around a single programming language, rather than different languages for server- and client-side scripts.
NodeJS allows the creation of Web servers and networking tools using JavaScript and a collection of "modules" that handle various core functionalities. Modules are provided for file system I/O, networking (DNS, HTTP, TCP, TLS/SSL,orUDP), binary data (buffers), cryptography functions, data streams, and other core functions. NodeJS modules use an API designed to reduce the complexity of writing server applications.


Figure 5. Node API Architecture

DATABASE INTEGRATION

In this project mongoDB is used for database integration.
MongoDB is a cross-platform, document-oriented database that provides, high performance, high availability, and easy scalability. MongoDB works on concept of collection and document.

Database

Database is a physical container for collections. Each database gets its own set of files on the file system. A single MongoDB server typically has multiple databases.

Collection

Collection is a group of MongoDB documents. It is the equivalent of an RDBMS table. A collection exists within a single database. Collections do not enforce a schema. Documents within a collection can have different fields. Typically, all documents in a collection are of similar or related purpose.

Document

A document is a set of key-value pairs. Documents have dynamic schema. Dynamic schema means that documents in the same collection do not need to have the same set of fields or structure, and common fields in a collection's documents may hold different types of data.

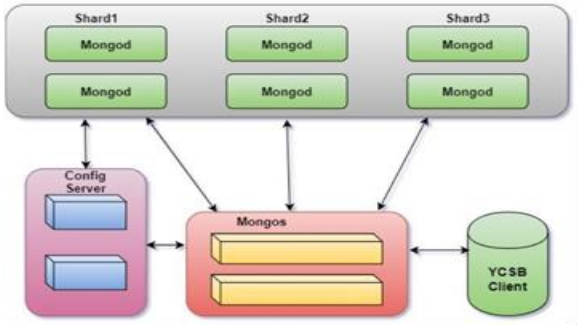


Figure 6. Database Architecture

WEB DEVELOPMENT

ReactJS is JavaScript library used for building reusable UI components. According to React official documentation, following is the definition −React is a library for building composable user interfaces. It encourages the creation of reusable UI components, which present data that changes over time. Lots of people use React as the V in MVC. React abstracts away the DOM from you, offering a simpler programming model and better performance. React can also render on the server using Node, and it can power native apps using React Native. React implements one-way reactive data flow, which reduces the boilerplate and is easier to reason about than traditional data binding.

JSX

JSX stands for JavaScript XML. JSX allows us to write HTML in React. JSX makes it easier to write and add HTML in React.
It is faster than normal JavaScript as it performs optimizations while translating to regular JavaScript. It makes easier for us to create templates. Instead of separating the markup and logic in separated files, React uses components for this purpose. We will learn about components in details in further articles.

Components

Components are independent and reusable bits of code. They serve the same purpose as JavaScript

functions but work in isolation and returns HTML via a render function.
Components come in two types, Class components and Function components, in this tutorial we will concentrate on Class components.

Unidirectional data flow and Flux
React implements one-way data flow which makes it easy to reason about your app. Flux is a pattern that helps keeping your data unidirectional.

## VI.RESULTS ANDDISCUSSIONS

A web application using a javascript framework reactJS was successfully developed, to enter the patient details to perform diagnosis for diabetes.
The below screenshot shows the web application for the project

USER SIGN-IN PAGE

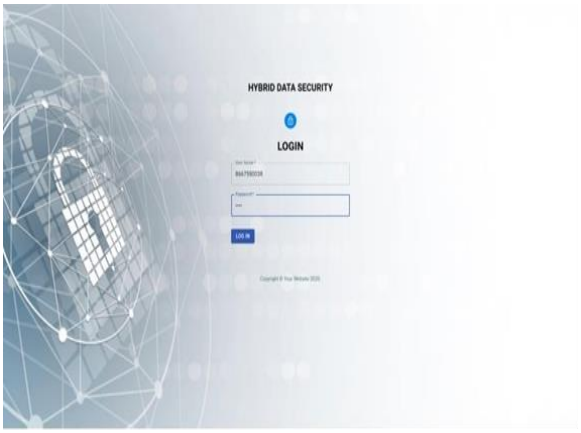


Figure 7. User Sign-In page

The user enters his/her login credentials to enter into the webpage. The below screenshot shows the user entering his/her login credentials:

HOMEPAGE
Once the user has successfully logged in, the home page is displayed to the user.
The patient enters his/her medical details to be used to perform diagnosis for diabetes. On clicking the save location button, the location is saved.
The below screenshot shows the form onto which the user uploads the medical data for diagnosis of diabetes.

SAVED PATIENT DETAILS

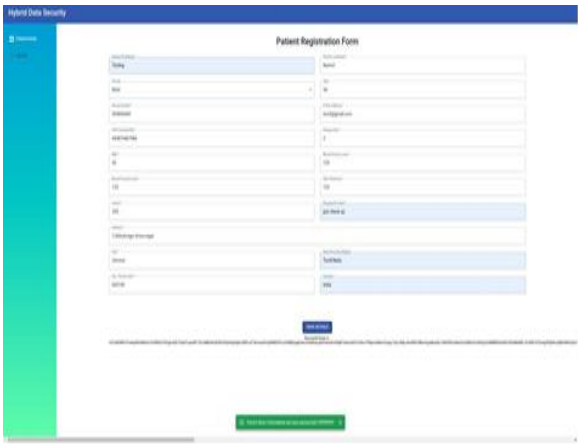


Figure 8. Patient Registration Form

The saved location can be viewed on the patient details tab. The below screenshot shows the user saved patient details:

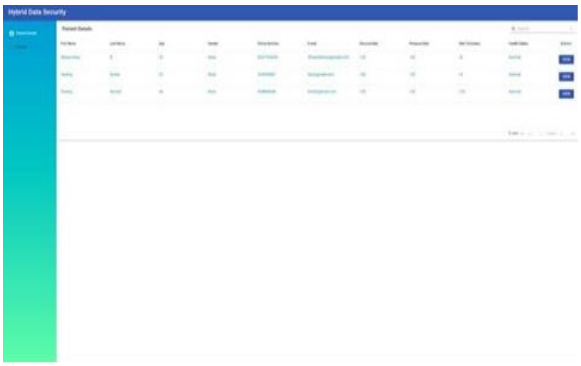

Figure 9. Saved Patient Details

9.6 SPECIFIC PATIENT DETAILS
On clicking the view button in patient details tab of a specific patient, the specific medical data of the patient is displayed.

GRAPH

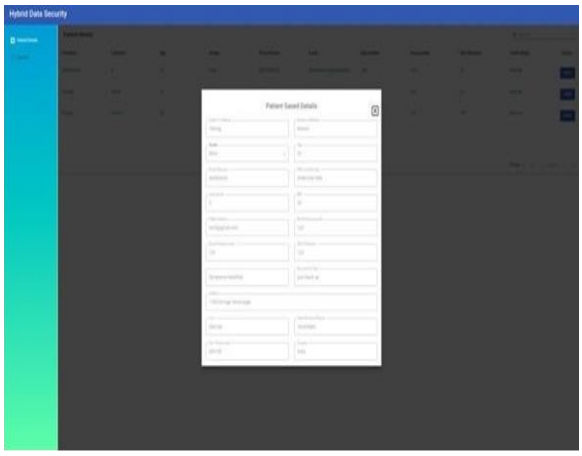

Figure 10. Specific Patient Details

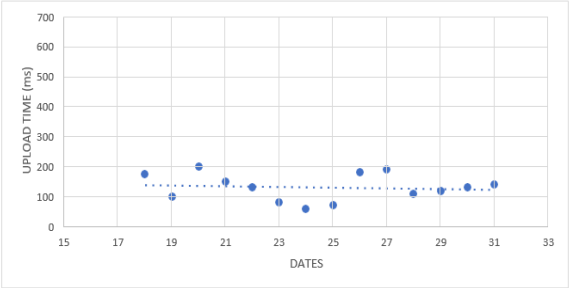


Figure 11. Upload Time – Dates Graph

Thus, from the above results and discussion, it is clear that we have efficiently made a project for successfully securing the data to be transmitted through wireless communications with a web application developed using a JavaScript framework ReactJS to encrypt and send medical data of a patient through wireless communications. Thus, we have successfully implemented the scope of the project.

## VII.CONCLUSION

This project is used to provide a solution to perform diagnosis for diabetics and securely transmit data through a wireless communication medium using a web application developed using a JavaScript framework reactJS. By this project, we can protect data from unauthorized access. Thus, this project provides an affordable and efficient means to protect and preserve data.

## REFERENCES

[1] Borja BORDEL, Diego MARTIN, Ramón ALCARRIA and Tomás ROBLES ,‖ A Blockchain -basedWaterControlSystemfortheAutomaticManagementofIrrigationCommunities‖,2019

[2] CaixueZhou,‖ Comments on ―Light-weight and Robust Security-Aware D2D-assist Data Transmission Protocol for Mobile-HealthSystems‖‖,[Vol:8937,2017]

[3] CaioDavi, André Pastor, Thiego Oliveira, Fernando B. de Lima Neto, Ulisses Braga-Neto, AbigailW.Bigham,MichaelBamshad,ErnestoT.A. Marques,BartolomeuAcioli-Santos,‖DengueE proteindetectionusinggrapheneoxideintegratedtaperedopticalfibersensor‖,[Vol:6781,2018]

[4] CaioDavi, Andre Pastor, Thiego Oliveira, Fernando B. de Lima Neto, Ulisses Braga-Neto, Abigail W. Bigham, Michael Bamshad, Ernesto T. A. Marques, Bartolomeu Acioli-Santos,‖SevereDengue Prognosis Using Human Genome Data and MachineLearning‖[2020]

[5] Mauro Mangia, Member, IEEE,, Alex Marchioni, Student Member, IEEE,, Fabio Pareschi, Member,IEEE,RiccardoRovatti,Fellow,IEEE,,GianlucaSetti,Fellow,IEEE,―ChainedCompressed Sensing: A Block-Chain-inspired Approach for Low-cost Security in IoTSensing‖,[2019]

[6] Mary Subaja Christo, AnigoMerjora A, ParthaSarathy G, Priyanka C and Raj Kumari M,‖An EfficientDataSecurityinMedicalReportusingBlockChainTechnology‖,April4-6,2019

[7] Y. Mustapha Kamil, Member, IEEE, M. H. Abu Bakar, Member, IEEE, M. H. Yaacob, Senior Member, IEEE, A. Syahir, H. N. Lim, M. A. Mahdi, Senior Member, IEEE,‖ A New Intelligence- Based Approach for Computer-Aided Diagnosis of DengueFever‖[Vol:2341-6781,2019]

[8] Shuai Wang ,Liwei Ouyang, Yong Yuan , Senior Member, IEEE, Xiaochun Ni, Xuan Han,and Fei-YueWang,Fellow,IEEE,‖Blockchain-EnabledSmartContracts:Architecture,Applications,and FutureTrends‖,[Vol:26,2019]

[9] VadrevuSree Hari Rao, Senior Member, IEEE, and Mallenahalli Naresh Kumar,‖Real-Time Dengue Prediction using Naive Bayes Predicator in theIoT‖,[Vol:6785,2018]

[10] Sundararajan Bhavani, Sengottaiyan Shanmugan, Venkatesan Chithambaram, Fadl Abdelmonem Elsayed Essa, Abd-Elnaby Kabeel, Periyasami Selvaraju‖ Simulation study on thermal performance of a Solar box Cooker using nanocomposite for natural Food invention, Environmental Science and Pollution Research, 28, pages 50649–50667 (2021)

[11] Junqin Huang, Linghe Kong, Senior Member, IEEE, GuihaiChen,‖ Novel Hybrid CMOS/Memristor Implementation of the AES Algorithm Robust against Differential Power Analysis Attack‖,[Vol:54804,2018]

[12] Jun Shang, Maoyin Chen, Member, IEEE, and Tongwen Chen, Fellow, IEEE,―Optimal Linear Encryption Against Stealthy Attacks on Remote StateEstimation‖,[2020]

[13] C.Suresh, S.Shanmugan, M.V.Bharath, B.Naveen, V.Chithambaram, Experimental analysis of Energy and Environment redeemable in solar Nano-basin still to improve Sullage Water Natural Treatment of Fuzzy Application, Materials today proceedings, 18, (2019) 1263-1271

[14] Yong Yu, Yannan Li, Jianbing Ni, Guomin Yang, Yi Mu and Willy Susil,‖ Teaching Network Security With IP DarkspaceData‖,[Vol:4738, 2017]