

# Secured Framework for Assessment of Chronic Kidney Disease in Diabetic Patients

Sultan Mesfer Aldossary\*

Dean of the College of Arts and Sciences, Prince Sattam Bin Abdulaziz University, Wadi Al-Dawasir, 11990, Saudi Arabia

\*Corresponding Author: Sultan Mesfer Aldossary. Email: s.aldossary@psau.edu.sa

Received: 13 August 2022; Accepted: 04 November 2022

**Abstract:** With the emergence of cloud technologies, the services of healthcare systems have grown. Simultaneously, machine learning systems have become important tools for developing matured and decision-making computer applications. Both cloud computing and machine learning technologies have contributed significantly to the success of healthcare services. However, in some areas, these technologies are needed to provide and decide the next course of action for patients suffering from diabetic kidney disease (DKD) while ensuring privacy preservation of the medical data. To address the cloud data privacy problem, we proposed a DKD prediction module in a framework using cloud computing services and a data control scheme. This framework can provide improved and early treatment before end-stage renal failure. For prediction purposes, we implemented the following machine learning algorithms: support vector machine (SVM), random forest (RF), decision tree (DT), naïve Bayes (NB), deep learning (DL), and k nearest neighbor (KNN). These classification techniques combined with the cloud computing services significantly improved the decision making in the progress of DKD patients. We applied these classifiers to the UCI Machine Learning Repository for chronic kidney disease using various clinical features, which are categorized as single, combination of selected features, and all features. During single clinical feature experiments, machine learning classifiers SVM, RF, and KNN outperformed the remaining classification techniques, whereas in combined clinical feature experiments, the maximum accuracy was achieved for the combination of DL and RF. All the feature experiments presented increased accuracy and increased F-measure metrics from SVM, DL, and RF.

**Keywords:** Cloud computing; diabetic kidney disease; machine learning; prediction system; privacy preservation; integrity of data; secured data transmission; homomorphic authentication

## 1 Introduction

The privacy of patient data has become one of the most pressing concerns. In order to protect sensitive information, there are a number of security strategies and encryption algorithms available [1–3]. Due to the widespread use of machine learning and particularly disease prediction machine learning, it is necessary to



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

collect and transfer data from hospitals via the Internet to train a model. Thus, such sensitive and private data is at risk of being leaked. Therefore, a critical concern for the sharing of intelligence is the feasibility of performing machine learning on private datasets while preventing data leakage. As a result, we have developed a machine learning-based DKD disease prediction system based on cloud computing architecture. Data transactions between entities in the cloud are secured through homomorphic cryptographic algorithms. This paper presents a privacy-protected machine learning algorithm based on homomorphic encryption. A model is trained under the best of optimized hyper parameters under the protection of the Pairing-Based Cryptography (PBC) library, which uses multiplicative cyclic groups of primes. Due to the member inference attack discussed in [4], malicious users could use the plaintext gradient to train a shadow model to compromise the data security of others. Hence, we introduce homomorphic encryption, which allows calculations to be performed on encrypted data without decryption. Moreover, the result of the homomorphic operation after decryption is equivalent to the operation on plaintext data [5]. The framework offers a multi-party privacy protected machine learning framework that combines homomorphic encryption and machine learning to guarantee the protection of data and model security during communications between consultants. Renal function impairment occurring in the advancement of diabetes can predict the future medical condition of kidney disease (KD) [6]. The chronic progression rate of DKD leads to a gradual decrease in the renal function [7], ultimately causing renal failure [8]. The risk factors for DKD include lifestyle, unattended hyperglycemia, dyslipidemia, and genetics. Lack of early recognition and clinical assessments can lead to end-stage renal disease (ESRD) [9]. The major challenges are the detection of DKD in its initial stage as well as its onset, overcoming which can improve DKD treatment outcomes in the kingdom. Although cloud computing and healthcare systems are currently being used, healthcare prediction systems (HCPSs) render services independently [10]. The integration of an HCPS with cloud computing can increase the rate as well as the accuracy of the results provided by the HCPS [11,12]. Diabetes is a complex and chronic illness that demands regular medical attention involving multifactorial risk-reduction strategies along with glycemic control [13]. Globally, diabetes is an epidemic, which has led to a rise in DKD. However, its symptoms are barely perceived until the patient collapses, thereby making the prevention of renal failure a challenge. In most cases, DKD is interpreted clinically based on the presence of albuminuria. The classic signs of DKD include persistent diabetes [13], albuminuria without hematuria [14], retinopathy [15], and steadily progressive KD. Notably, DKD remains a prominent chronic complication of both types of diabetes (Types 1 and 2), which subsequently leads to ESRD, accounting for approximately 50% of the cases globally [16]. DKD is detected clinically by screening for persistent abnormal urine albumin excretion and decrease in the estimate of the glomerular filtration rate (eGFR) [16]. Decreased eGFR and increased albuminuria are the major causes of cardiovascular mortality [17–19]. The major cause of mortality among patients with DKD is the excess load on the cardiovascular system [20]. Patients with diabetes have more probability of dying, primarily from cardiovascular complications, than those with ESRD [21]. Calcium phosphate homeostasis, arterial stiffening, and sudden cardiac death due to vascular disorder and coronary artery disease are other pathological events reported for patients with DKD [22]. The predominance of diabetes is owing to the changing human lifestyle with reduced physical activity and increased consumption of salt- and sugar-rich soft drinks and fast food, leading to obesity. Diabetes is associated with life-threatening complications, such as KD, which can progress to ESRD, requiring dialysis or kidney transplant [23]. The prevalence of diabetes mellitus (DM) is significant, and 90% of diabetes patients have Type II DM [23]. Additionally, uremia is a raised level of blood urea and nitrogenous waste that are not eliminated by kidney during renal disease. Uremic syndrome involves accumulation in plasma of small uremic solute and toxins. These uremic solutes if failed to get excreted out by kidney, they can exert toxic effects or disrupt key signaling pathways leading to life-threatening complications [24]. Moreover, a recent study suggested that the increase in the prevalence of diabetes and Chronic kidney disease (CKD) can increase renal therapy costs [25]. Thus, the present study determines

the prevalence of DKD, emphasizing the need for its timely detection and generation of new knowledge to significantly improve the lives of asymptomatic patients.

Platform as a service (PaaS) is a software development framework, and it is delivered over the internet to personal computers and mobile devices [26,27]. It presents a platform utilizing software components and tools from the other two layers within the cloud computing architecture. PaaS assists developers to modify, assess, and launch their applications. PaaS providers manage the operating system, servers, networking, virtualization, middleware, runtime libraries, and network security patches. Developers focus only on software development and data, ignoring the maintenance of the supported tools. Using such a setup, such as our framework, the development of a decision-making software can be beneficial because of the lack of requirement of the maintenance of the supported software and hardware equipment as mentioned in the service level agreement [28].

The remainder of this paper is organized as follows. Section 2 presents, Section 3 presents the related studies. Section 4 describes the implementation of the proposed framework for predicting DKD and the medical data handling by homomorphic authentication for the privacy preservation of the different schemes of the feature selection of the dataset. Section 4 presents the experimental results and discussion. In the last section, the conclusions are drawn and future studies are suggested.

## 2 Literature Review

This section presents the most recent developments in the prediction services for DKD patients using cloud computing. These innovative services are subsequently compared to the present study. We have summarized the most recent advances in the areas of DKD services and prediction models. Cloud computing and its layered architecture have more potential than the conventional network architecture for significantly enhancing healthcare services. However, cloud computing has problems regarding service management and service level agreements, which are beyond the scope of this study. Herein, we focused on cloud service offerings. Table 1 lists the studies (presented as reference numbers) utilizing cloud-based systems for predicting diseases. In this section, we have listed all the articles related to diseases, particularly CKD and diabetes, as well as general health (GH).

**Table 1:** List of studies discussing cloud computing for health care

| Reference | AI model              | Diseases | Cloud computing | Privacy preservation | Cryptography |
|-----------|-----------------------|----------|-----------------|----------------------|--------------|
| [29]      | LR and NN             | CKD      | Yes             | None                 | None         |
| [30]      | Fuzzy NN              | GH       | Yes             | None                 | None         |
| [31]      | Stochastic prediction | GH       | None            | None                 | None         |
| [32]      | None                  | GH       | Yes             | Yes                  | Yes          |
| [33]      | LR and NN             | CKD      | Yes             | None                 | None         |
| [34]      | SVM                   | DKD      | None            | None                 | None         |
| [35]      | PCA                   | DM       | None            | None                 | None         |
| [36]      | Stochastic prediction | DM       | Yes             | None                 | None         |
| [37]      | LR                    | DM       | None            | None                 | None         |
| [38]      | None                  | GH       | Yes             | Yes                  | None         |
| [39]      | None                  | GH       | Yes             | Yes                  | Yes          |
| [40]      | None                  | GH       | Yes             | Yes                  | None         |
| [41]      | None                  | GH       | Yes             | Yes                  | None         |

(Continued)

**Table 1 (continued)**

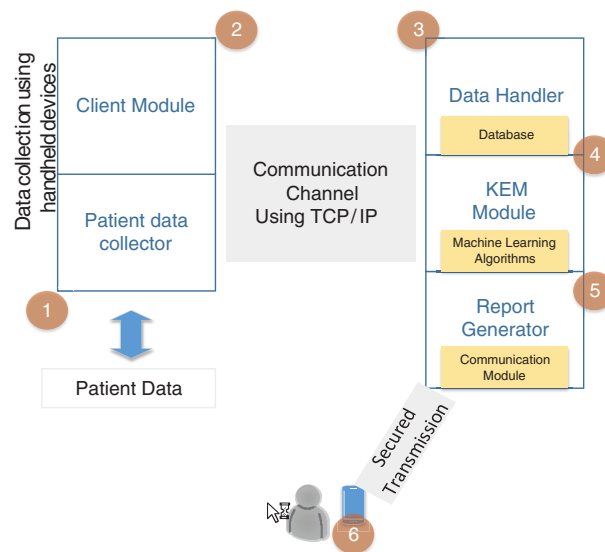
| Reference | AI model | Diseases | Cloud computing | Privacy preservation | Cryptography |
|-----------|----------|----------|-----------------|----------------------|--------------|
| [42]      | None     | GH       | Yes             | Yes                  | None         |
| [43]      | None     | GH       | Yes             | Yes                  | None         |

A study emphasized the prediction of DKD in a cloud computing environment using hybrid technologies such as neural networks and linear regression [37]. However, it was limited to only two hybrid neural network strategies. A review article presented the potential of cloud service as a healthcare and prediction system [38]. A mobile health system was built using the cloud and the Internet of Things (IoT) for monitoring and diagnosing serious diseases; however, the diseases were not specified. A study proposed a data collection strategy based on a probabilistic mechanism and performed correlation analysis on the collected data. Moreover, it proposed a stochastic prediction model to predict the health of patients based on their current health status [38]. Another study proposed a predictive analysis algorithm using MapReduce techniques to identify DKD and its associated complications and to determine the probability of its diagnosis [39]. The algorithms used in the above studies were subsequently used to predict disease severity; however, they were not implemented for predicting any particular disease. The most recent contributions in security and privacy [39–41] provide frameworks for storing and policies for accessing medical records by key control schemes utilizing block chains and conventional encryptions using the IoT and the cloud architecture. However, the block chain technique is CPU-intense and requires a high bandwidth, and therefore, it is inappropriate for most resource-demanding IoT devices for remote medical services [42]. Health-based cloud computing provides highly flexible and efficient services, such as easy and ubiquitous access to patient data, and also has the potential for establishing new business models. In addition to the major benefits, cloud computing and its extended IoT architecture can be used to solve security issues, which are the main challenges for healthcare systems because of the daily data transactions between their entities when serving diagnostic results. Hence, to secure patient data and maintain privacy, numerous approaches have been proposed, and although these protocols are efficient, they are also CPU-intense. Hence, such methods cannot be implemented for a healthcare system because of the small size of the patient data [42,43]. Furthermore, these are conventional cryptographic methods, and access to the cloud data while ensuring security and privacy is needed in an insecure cloud computing architecture. Cloud computing architecture can provide quality care facility for patients using data ML and deep learning techniques for making medical decisions and reducing the costs by efficiently allocating resources in terms of personnel and equipment. Personal health records are provided by giant companies, such as Microsoft, Google, and Life Sensors. These systems allow patients to store their health records, which can also be retrieved by data protection techniques. Furthermore, patients can create access control for their data records so that their respective doctors can also access them. However, such access control policies can attract unwanted security breaches and also lead to the loss of privacy. These disadvantages are mitigated by introducing advanced E-health cloud infrastructures, which are specialized infrastructure for accessing medical services using well-defined software and hardware services. Under the protection of privacy, the aforementioned machine learning models have been trained to yield the most accurate results according to their respective hyper-parameters. The following are the advantages based on which the ML algorithms were chosen as the primary algorithms in our proposed methodology:

- The suggested algorithms are extremely effective with a strong margin of separation of the datasets.
- They are most operative in high-dimensional spaces.
- They use a subset of the training dataset in the decision function; therefore, they are memory efficient.

### 3 Proposed Framework

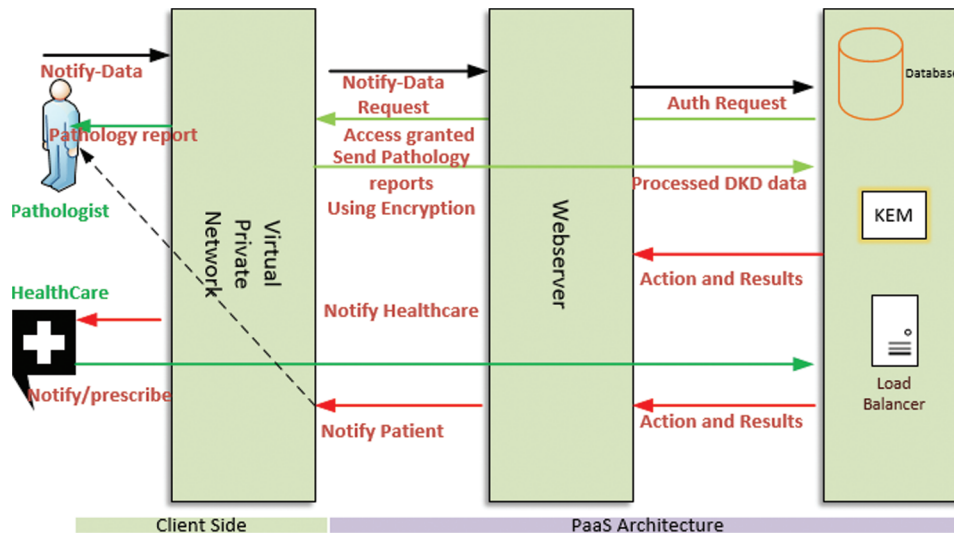
Our proposed framework was implemented for the early detection of the end stage of DKD using the ML classifiers, and the data transactions (viz patient data transmission) between the cloud server and the client were privacy preserved. We used six classifiers for appropriately classifying DKD and its various features; these classifiers were extracted from the UCI dataset and its subsets. Our approach used symptoms and clinical data readings as the features to determine and predict the end stage of DKD. Fig. 1 shows our proposed flow diagram. The client module contains the client system (typically operated by doctors), and the server module includes a kidney event manager (KEM), data handler module, and report generator module (these modules are managed by the cloud service provider). The client module collects the data from the end users, such as doctors, pathology departments, or any smart device(s) that can communicate using the transmission control protocol/Internet protocol. The KEM acts on the patient data and executes ML scripts to predict the level of disease severity. Finally, the report generator module generates reports based on the KEM module instructions.



**Figure 1:** Flow graph to illustrate the need for the proposed approach

The data handler module allows doctors and pathologists to upload data on the server. The data must be stored with access control of the database. Hence, we utilized some access-controlled database upload and retrieval methods. To deal with a large database, we employed a secured virtual private network for storing the dataset in the database. Smart devices that can communicate using the transmission control protocol/Internet protocol or those that can feed medical records to our system were not used. However, we utilized the UCI Machine Learning Repository datasets in the present study. The data pertaining to patients contain sensitive and personal information which may be of interest to malicious users. Consequently, it is imperative to ensure both the security of the proposed system as well as the security of the underlying systems and components that constitute the proposed framework; additionally, healthcare data must be protected both from external threats and from unauthorized access attempts from within the network. In order to regulate and limit access to patient data, access control models using cryptographic primitives, such as those based on public key infrastructure and public clouds, are employed. These models are especially effective against both external and internal threats.

PaaS, a cloud computing model in which the cloud service provider distributes the hardware and software tools generally required for application development by the users on the Internet. We used this PaaS model to develop a DKD prediction system using ML algorithms. Fig. 2 illustrates the proposed PaaS-based cloud-assisted DKD framework comprising two major modules: client and server.



**Figure 2:** Proposed PaaS-based prediction framework

The DKD clinical data were uploaded to the remote PaaS database for further use by local and remote consultants or healthcare professionals. Patients can access the framework by their mobile devices to determine their next steps. The KEM conducts the complete supervision of the proposed framework and its services. Moreover, in the PaaS container, it can allocate the required resources, report the performance of the patient, and control administrative services, depending on the profile of each patient and doctor. The PaaS platform assigns various virtual machine resources to execute the classifiers and their sessions on the web server operating environment. For convenience, the abbreviations and their corresponding definitions used in the paper are listed in Table 2.

**Table 2:** Privacy symbols and description

| Signs and symbols                          | Description                      |
|--------------------------------------------|----------------------------------|
| $\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T$ | Cyclic group of order $r$ & $e$  |
| PBC                                        | Pairing-based cryptography       |
| $F$                                        | Medical file or patient file     |
| $m_1, m_2, m_3, \dots, m_n$                | Logical division of medical data |
| $P_k$                                      | Public key                       |
| $S_k$                                      | Secret key                       |
| $Sign_k$                                   | Applying HA algorithm            |
| HA                                         | homomorphic authenticator        |
| SecuredKeyGen                              | secured key generation algorithm |

(Continued)

**Table 2 (continued)**

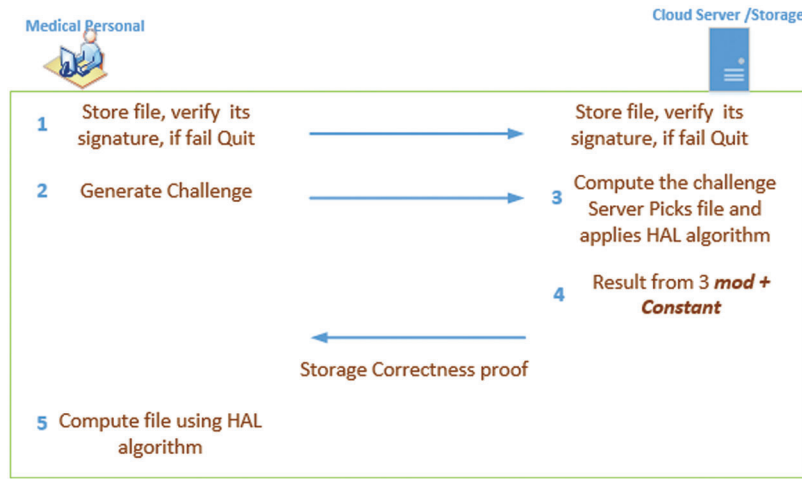
| Signs and symbols              | Description                                |
|--------------------------------|--------------------------------------------|
| GenerateProof                  | proof of data storage correctness          |
| <i>VerifyProof</i>             | It verifies the received data from server  |
| <i>fileTags</i>                | Medical data tags for the security purpose |
| <i>metadata<sub>file</sub></i> | Medical file metadata                      |
| KEM                            | kidney event manager                       |
| CSP                            | cloud service provider                     |
| class                          | Nominal                                    |
| eGFR                           | Estimated Glomerular filtration rate       |

### 3.1 Medical Data Handling Using Homomorphic Authentication for Privacy Preservation

The privacy-preserving file scheme developed using a pairing-based cryptography (PBC) library file is implemented by C programming and produces different-sized security bits, such as 80 or 160 bits. These bits are used to generate the security parameter keys. The PBC library uses multiplicative cyclic groups of prime numbers of  $r$  order. It also employs a hash function to generate random numbers within a cyclic group. The cyclic groups are typically represented as  $\mathbb{G}_1$ ,  $\mathbb{G}_2$ , and  $\mathbb{G}_T$  of prime orders  $r$  and  $e$ . We have built a privacy-preserving data transmission scheme that uses a homomorphic authenticator (HA). HA schemes are established using two key concepts known to provide malleability and block-less verification. Block-less verification enables a cloud server to verify remote data in the absence of data and metadata. To execute this process, we need to generate tags for all the blocks of the data. The cloud server uses these uniquely numbered tag files to generate the proof of file storage. Any third party wanting to retrieve the medical data from the cloud should possess a file tag. Initially, the client implements the setup and verifies the phase. The setup and verify phases initialize the parameters, called as public and secret keys, to build this process (preliminary process). Our proposed scheme consists of four algorithms: **SecuredKeyGen**, **SigningKey**, **GenerateProof**, and **Verifyproof**. **SecuredKeyGen** is a secure key generation algorithm executed by the client machine to setup the process (privacy-preserving scheme). Signing Key is employed by the client to assign the proof-of-file tags to store on the cloud server. The proof includes the HA, signatures, or other medical-related evidence (metadata) used for uploading medical data to the server. Generate Proof is implemented by the cloud server to produce the proof of data storage correctness, whereas Verify Proof is run by the medical personnel to assess the proof from the cloud server. A client (medical personnel) can request verification of any medical data by issuing a challenge message to the cloud server to ensure that the server has stored the medical data file. The server subsequently responds with a message from a function of the stored medical data file, and it authenticates metadata by executing Generate Proof. The client then validates the response using Verify Proof.

Each medical data file  $F = (m_1, m_2, m_3, \dots, m_n)$ , as shown in Fig. 3, uses an algorithm to create metadata and Signing Key for authenticator. Subsequently, the server ensures the integrity of the unique identifier for each file name, generating a signature using a private key. The public and secret can be obtained from the Signing Key algorithm.

$$(P_k, S_k) \leftarrow (Sign_k). \quad (1)$$



**Figure 3:** Privacy preserving medical data storage

The client sends a file along with the metadata; subsequently, the server initially stores the file using public and secret parameters, and quits if its verification fails. Hence, only authorized medical personnel can view, store, and edit the medical data. The client executes the following equation to encode the file blocks to store on the cloud server:

$$(metadata_{file}, file_{Tags}, Sign_k) \leftarrow (SecuredKeyGen_k) \quad (2)$$

To verify a storage without compromising the privacy of the medical data, the client issues a challenge by selecting the file with a random value. The challenge message specifies the part of the file that needs to be checked. After receiving the challenge from the client, the server generates the proof using Generate Proof. The server executes the HAL algorithm with a random number, the public key, and the secret key, and along with metadata of the file, the server generates the proof. With this information, the client generates the verification using the response generated by the server.

The challenge can be issued using Eq. (3), where  $C$  is the challenge with a public key and file tags.

$$\mu = Chal(P_k, File_{Tags}, C) \quad (3)$$

Subsequently, the server computes the challenge by executing Eq. (4).

$$\mu = compute(\mu + r \bmod p) \quad (4)$$

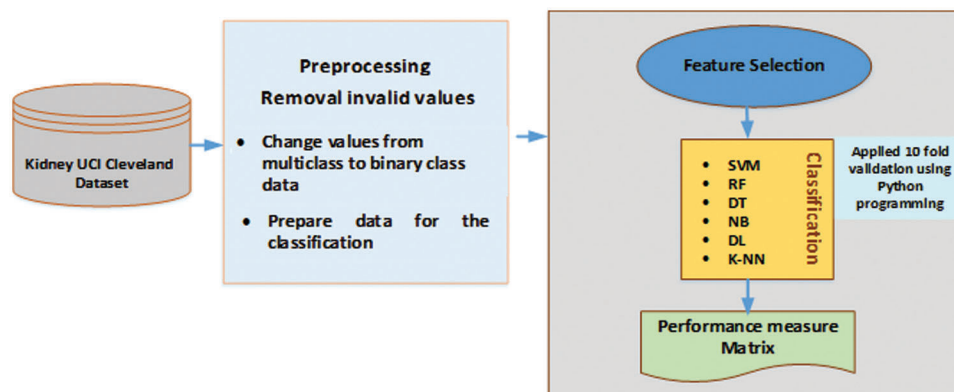
The client verifies the proof generated by the server by executing Eq. (5).

$$VerifyProof(P_k, metadata_k, c, \mu) \quad (5)$$

### 3.2 Dataset Preprocessing

The dataset formatting for this study is performed using the pre-processing technique presented in Fig. 4. We used Python programming to build scripts for the support vector machine (SVM), decision tree (DT), KNN, naïve Bayes (NB), random forest (RF), and deep learning (DL) algorithms. The UCI dataset for DKD was downloaded from UCI website. The dataset can be ported directly into our framework, it was preprocessed to suit our implementation. Subsequently, we performed a feature-based execution. The above process was executed for each ML classifier technique. The results of all the ML techniques were recorded based on the features marked for the DKD dataset. The DKD dataset was downloaded from the

UCI Machine Learning Repository. This dataset comprises 400 instances of data of various numerical parameters and includes 24 attributes, of which 13 were employed to predict the end stage of DKD. Due to the fact that these 13 factors are recommended by many medical research studies as being crucial in identifying DKD in any given patient. Table 3 lists the factors that are included in our framework. The 400 instances of data in the dataset, obtained from various hospitals in Chennai, India comprise data from 250 DKD patients and 150 non-DKD patients. Table 3 summarizes the description of the types of attributes. The UCI dataset contains the numerical values for most attributes; however, some are marked as present or absent, e.g., CKD when present or absent can be marked as 1 or 0, respectively. Similarly, hypertension, DM, coronary artery disease, pedal edema, and anemia are denoted as Boolean values. Specific gravity is represented using a range of values, such as {1.005, 1.010, 1.015, 1.020, 1.025}. Albumin is represented as {0, 1, 2, 3, 4} for its normal and abnormal ranges, whereas pus cells and red blood cells are represented as normal and abnormal ranges.



**Figure 4:** Pre-processing phase

**Table 3:** Description of attributes of the DKD dataset from the UCI machine learning repository

| Attribute | Employed in our framework | Description          | Type of value |
|-----------|---------------------------|----------------------|---------------|
| Age       | Yes                       | Age                  | Numeric       |
| BP        | Yes                       | Blood pressure       | Numeric       |
| Al        | NO                        | Specific gravity     | Numeric       |
| Su        | Yes                       | Albumin              | Nominal       |
| rbc       | NO                        | Sugar                | Nominal       |
| Pc        | NO                        | Red blood cells      | Nominal       |
| pcc       | Yes                       | Pus cell             | Nominal       |
| Ba        | NO                        | Bacteria             | Nominal       |
| bgr       | Yes                       | Blood glucose random | Numeric       |
| Bu        | Yes                       | Blood urea           | Numeric       |
| Sc        | Yes                       | Serum creatinine     | Numeric       |
| sod       | Yes                       | Sodium               | Numeric       |
| pot       | Yes                       | Potassium            | Numeric       |

(Continued)

**Table 3 (continued)**

| Attribute | Employed in our framework | Description             | Type of value |
|-----------|---------------------------|-------------------------|---------------|
| hemo      | Yes                       | Haemoglobin             | Numeric       |
| pcv       | NO                        | Packed cell volume      | Numeric       |
| wc        | NO                        | White blood cell count  | Numeric       |
| Rc        | NO                        | Red blood cell count    | Numeric       |
| htn       | Yes                       | Hypertension            | Numeric       |
| dm        | NO                        | DM                      | Nominal       |
| cad       | Yes                       | Coronary artery disease | Nominal       |
| appet     | NO                        | Appetite                | Nominal       |
| Pe        | NO                        | Pedal edema             | Nominal       |
| ane       | Yes                       | Anemia                  | Nominal       |
| class     | NO                        | Class                   | Nominal       |

### 3.3 Feature Selection & Training and Testing Methodology

A description of the dataset's features can be found in [Table 3](#). CKD is a global health problem. Yearly there is an increase in the total number of patients with CKD leading to ESRD. The etiology of CKD is diverse. Most cases are secondary to DM and hypertension. Other causes include glomerular diseases, leading to proteinuria KD. Because there is no treatment to reverse the renal damage caused by DKD and hypertensive nephrosclerosis, the current treatment objective is risk factor control. This is also the case for any type of CKD. Numerous studies including patients with Type 1 and Type 2 DM, hypertension, and KD originating from other causes with high-grade proteinuria have demonstrated that altering these factors can modify the CKD progression. Hence, clinicians should thoroughly assess CKD risk factors. This includes assessment of diabetes control with glycosylated hemoglobin, blood pressure using sphygmomanometers, and proteinuria using urine samples. Because atherosclerosis is a significant contributor to CKD progression, hyperlipidemia treatment is also an important objective of the therapy. Hyperlipidemia is treated by reducing the low-density lipoproteins. Hence, in the clinical scenario, the management of all CKD patients, irrespective of the etiology, includes a good assessment of the underlying risk factors and provision of treatment accordingly. CKD diagnosis is established by the presence of kidney damage (urinary albumin excretion of 30 mg/day or more) or decreased kidney function, which is defined by a reduced eGFR for 3 months or more, irrespective of the etiology. CKD staging provides a guide for its management. The hyper parameters which have employed, are shown in [Table 4](#).

**Table 4:** Different ML models with hyper parameters to be tuned by the grid search technique

| ML models | Hyperparameters                                                                                                                                                                                                   |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SVM       | #Define model<br>#Define kernel (poly, rbf, sigmoid)<br>#Define misclassified point as $C = [50, 10, 1.0, 0.1, 0.01]$<br>#Define gamma = ['scale'] controls the distance of influence of a single training point. |

(Continued)

**Table 4 (continued)**

| ML models | Hyperparameters                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|           | <pre>#Define grid search # define grid search grid = dict (kernel=kernel, C=C, gamma=gamma) #Define cross validation as RepeatedStratifiedKFold (n_splits = 10, n_repeats = 3, random_state = 1) Call GridSearchCV with the above mentioned parameters such as estimator=model, grid, n_jobs=-1, cv, scoring='accuracy' and error_score = 0 grid_result = grid_search.fit (X, y)</pre>                                                                                                                                                               |
| RF        | <pre>#Define model #Define n_estimators = [10, 100, 1000] #Define features = ['sqrt', 'log2'] #Define define grid search as n_estimators=n_estimators, max_features=max_features) #Define cross validation CV as RepeatedStratifiedKFold (n_splits = 10, n_repeats = 3, random_state = 1) Apply gride search using earlier define parameters as estimator=model, param_grid=grid, n_jobs=-1, cv=cv, scoring='accuracy', error_score = 0) Run grid_search.fit (X, y)</pre>                                                                            |
| KNN       | <pre>#Define model = KNeighborsClassifier () Specify range of neighbors as 1, 21, 2 Define weights as 'uniform', 'distance' Use metric as ['euclidean', 'manhattan', 'minkowski'] Apply grid search grid = dict (n_neighbors=n_neighbors, weights=weights, metric=metric) Use cross validation as RepeatedStratifiedKFold (n_splits = 10, n_repeats = 3, random_state = 1) Run GridSearchCV (estimator=model, param_grid=grid, n_jobs=-1, cv=cv, scoring='accuracy', error_score = 0) Use the most efficient by running grid_search.fit (X, y)</pre> |
| DT        | <pre>#Define the mode Choose the estimators as [10, 100, 1000] Define grid search using dict (n_estimators=n_estimators) Apply Cross validation as RepeatedStratifiedKFold (n_splits = 10, n_repeats = 3, random_state = 1) Call GridSearchCV (estimator=model, param_grid=grid, n_jobs=-1, cv=cv, scoring='accuracy', error_score = 0) Fit the model by calling grid_search.fit (X, y)</pre>                                                                                                                                                        |
| NB        | <pre>#Define objective function as x, noise = 0.1): , noise = normal (loc = 0, scale=noise) Allocate grid-based sample of the domain [0, 1] with rage arange (0, 1, 0.01) Find the optimal result using argmax (y) from the library from numpy import argmax</pre>                                                                                                                                                                                                                                                                                   |

Among the 25 attributes used in DKD prediction, age and sex are attributes belonging to the personal information of the patient. Hence, they do not have any impact on the prediction. However, the remaining 23 attributes are clinical values, which are absolute readings used in the DKD prediction. Hence, we used all these attributes for our selected classifiers, which are discussed in the Results and Discussion section. However, selecting the most significant attribute according to the American Association of Kidney

Patients and other relevant kidney associations suggests that albumin, serum creatinine, and anemia are the indicators of ESRD. DKD can silently advance towards end stage renal disease, early detection and timely intervention is critical, screening of serum creatinine, albuminuria and anemia is recommended. High serum creatinine level and amount between 30 to 300 mg of albumin loss per day was considered to be insignificant previously, but presently with research advances it is now considered to be significant by current clinical practice that emphasize albuminuria. Patients with CKD are at increased risk of anemia; hence hemoglobin should be measured in patients with CKD, specifically when renal function declines. Normochromic normocytic anemia develops from the decreased renal synthesis of erythropoietin, the hormone responsible for bone marrow stimulation for red blood cell (RBC) production. Uremia-induced platelet dysfunction enhances the bleeding is considered to be one of the other causes of anemia in DKD. Complete Blood Count shows normochromic normocytic anemia.

Age and Sex are considered in every patient when diagnosing the DKD, but present study includes more specific parameters related to kidney function to intervene the early detection of renal failure progression. Present is not involving the patient's cross-sectional study, but it is rather a study which gives methods to early intervention of renal failure progression to prevent end stage renal failure and renal replacement. Relation of age and sex with interventional method described in present study is not the scope of present study. Rather present study provides an interventional method for early detection of renal failure to prevent end stage renal failure.

Once the set of features are selected, the models for this research are constructed. We conduct a ten-fold cross-validation to validate the performance of the models. Our UCI dataset is divided into numerous subsets, which were processed ten times. Approximately 90% of the dataset is used for training, and the remaining 10% is used for testing. Finally, the results are obtained by averaging each of the ten repetitions. The datasets are arranged according to stratified sampling. The following neural network layers are available to build sophisticated architectures:

- ConvolutionLayer
- DenseLayer
- SubsamplingLayer
- BatchNormalization
- LSTM
- GlobalPoolingLayer
- OutputLayer: generates classification/regression outputs

#### 4 Results

The performance of the proposed classifier techniques were measured using performance measures: accuracy, F-measure, recall, precision, and confusion matrix. Accuracy is represented as percentage and is the ratio of the correctly predicted attributes to the total observations. The prediction model is acceptable if the accuracy is high, and if the dataset is symmetric, its attribute values of false positives and false negatives are similar in number. The F-score or F-measure is the weighted mean of the precision (P) and recall (R). P is the ratio of the correctly predicted positive attributes to the total predicted positive attributes, whereas R is the number of positive attribute predictions from the total positive examples in the dataset. The confusion matrix is a  $2 \times 2$  matrix containing four outcomes produced by a binary classifier. The four outcomes are classified as true positives: correct positive prediction, false positive: incorrect positive prediction, true negative: correct negative prediction, and false negative: incorrect negative prediction. This section presents the performance of the tests. The 25 attributes, selected featured attributes, 1 selected attribute, and all the 6 classifying techniques applied to identify suitable features

were analyzed. The 6 classifiers were applied to 9200 combinations of the attributes, and the performance was measured and reported for each classifier. The classifier performance of the six ML methods on various the combinations of the attributes was investigated sequentially. All the investigation outcomes, particularly the accuracy, P, and F-measure of each technique, were collected for the analysis. [Tables 5 and 6](#) list the accuracy, P, R, and F-measure of each classifier, with only a few key attributes directing the prediction of the early stage of DKD. Thus, the best accuracy of 87.5% was attained by SVM, RF, and KNN with the key features, and accuracy of 87% was attained by DT. The NN and NB techniques yielded good results and could also be implemented in the predication process; however, they lagged in comparison to the SVM, RF, and KNN techniques. [Table 7](#), summarizes the techniques considering the features of age, blood pressure, cardiovascular disease, and anemia. It reveals that DL is the best performer when multiple attributes are considered. The DL accuracy was 86.1%, recall was 86.2%, and precision was 84.2% with <86% correctly identified or classified attributes. The performance matrix of RF was similar to that of DL, but with respect to precision, it lagged by 0.030. NB, SVM, DT, and KNN did not perform well, and their accuracy, recall, and precision performance measurements were below 85%. [Table 8](#), include the complete attributes mentioned in [Table 2](#). SVM, DL, and RF attained 99% accuracy with almost correctly identified attributes, whereas NB, DT, and KNN attained <95% accuracy ([Table 8](#)). Therefore, increasing the number of features in the framework can yield better performance with most classifying techniques.

**Table 5:** Classifiers on only one selected feature

| Techniques | Features | Correctly classified attributes in % | Incorrectly classified attributes in % |
|------------|----------|--------------------------------------|----------------------------------------|
| NB         | SC       | 83.5                                 | 16.5                                   |
| SVM        | SC       | 87.5                                 | 12.5                                   |
| DL         | SC       | 83                                   | 17                                     |
| RF         | SC       | 87.5                                 | 12.5                                   |
| DT         | SC       | 87                                   | 13                                     |
| KNN        | SC       | 87.5                                 | 12.5                                   |

**Table 6:** Classifiers on only one selected feature

| Techniques | Features  | Precision    | Recall       | F-measure    | Accuracy     |
|------------|-----------|--------------|--------------|--------------|--------------|
| NB         | SC        | 0.871        | 0.835        | 0.838        | 0.835        |
| SVM        | <b>SC</b> | <b>0.891</b> | <b>0.875</b> | <b>0.877</b> | <b>0.875</b> |
| DL         | SC        | 0.844        | 0.830        | 0.832        | 0.83         |
| RF         | <b>SC</b> | <b>0.896</b> | <b>0.875</b> | <b>0.877</b> | <b>0.875</b> |
| DT         | SC        | 0.893        | 0.870        | 0.872        | 0.87         |
| KNN        | <b>SC</b> | <b>0.896</b> | <b>0.875</b> | <b>0.877</b> | <b>0.875</b> |

**Table 7:** Classifiers on a combination of selected features

| Techniques | Features                 | Accuracy     | Recall       | Precision    |
|------------|--------------------------|--------------|--------------|--------------|
| NB         | Age, BP, Cad, Ane        | 0.852        | 0.855        | 0.824        |
| SVM        | <b>Age, BP, Cad, Ane</b> | <b>0.855</b> | <b>0.857</b> | <b>0.832</b> |
| DL         | Age, BP, Cad, Ane        | 0.861        | 0.862        | 0.842        |
| RF         | <b>Age, BP, Cad, Ane</b> | <b>0.86</b>  | <b>0.862</b> | <b>0.836</b> |
| DT         | Age, BP, Cad, Ane        | 0.845        | 0.847        | 0.773        |
| KNN        | <b>Age, BP, Cad, Ane</b> | <b>0.802</b> | <b>0.805</b> | <b>0.799</b> |

**Table 8:** Classifiers on all features

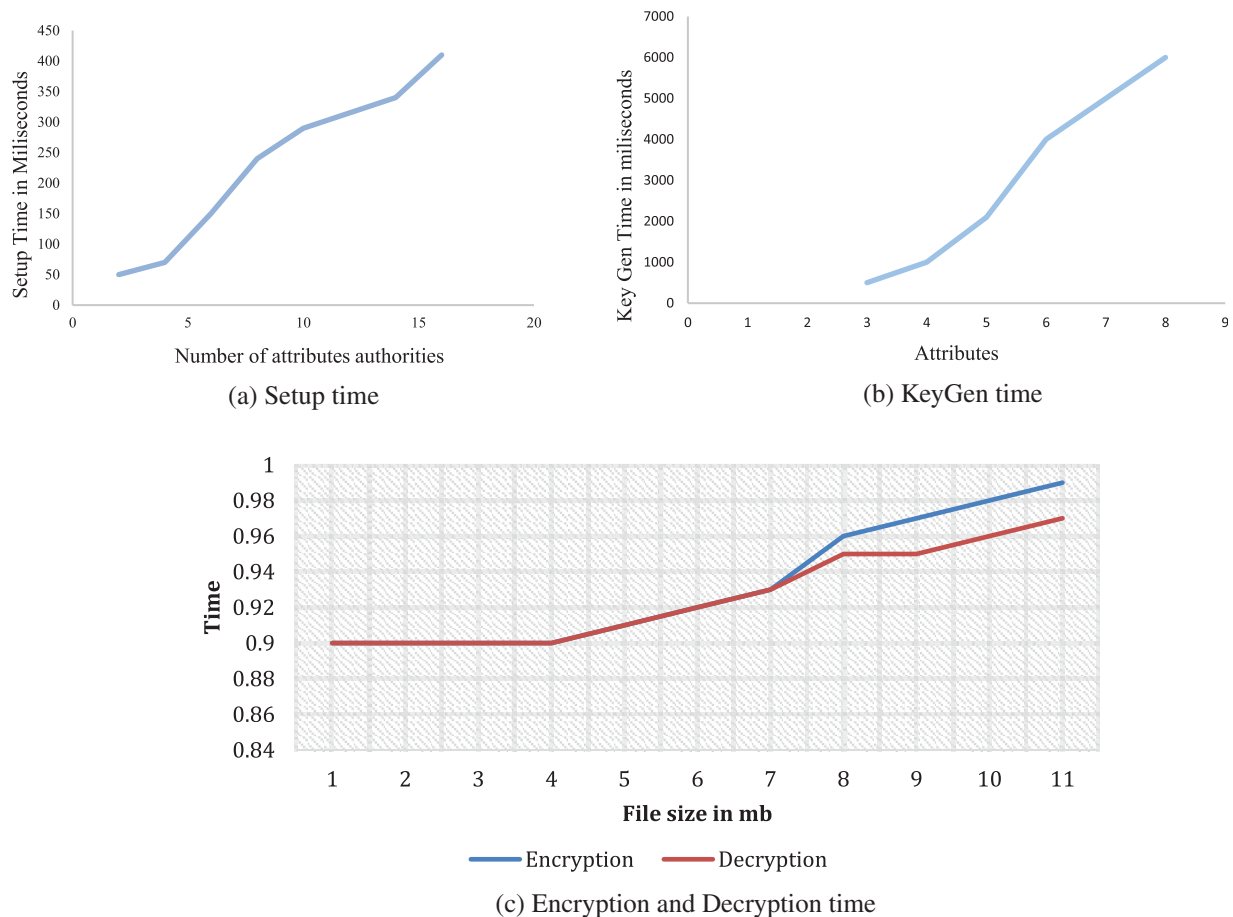
| Techniques | Accuracy | Correctly classified attributes in % | Incorrectly classified attributes in % |
|------------|----------|--------------------------------------|----------------------------------------|
| NB         | 0.95     | 0.95                                 | 0.5                                    |
| SVM        | 0.99     | 0.99                                 | 0.1                                    |
| DL         | 0.99     | 0.99                                 | 0.25                                   |
| RF         | 0.99     | 0.99                                 | 0.1                                    |
| DT         | 0.96     | 0.95                                 | 4.25                                   |
| KNN        | 0.96     | 0.95                                 | 4.25                                   |

In the proposed framework, we presented that the classifier algorithms could predict the end stage of DKD using ML based on the UCI Machine Learning Dataset Repository. We applied a cloud computing-based framework to improve the ability of the prediction system to detect DKD. Initially, we constructed a cloud computing framework on Google cloud to implement the classifiers and provide a database of patients for storing and retrieving data when required by the end user. Finally, we implemented various ML classification techniques on the dataset and generated results.

Diabetes is a major cause of kidney disorder. The lack of early recognition and laboratory assessment could lead to a transition of CKD in patients with diabetes to ESRD, if not addressed. The objectives of this study were to detect renal function in patients with diabetes in the region of Saudi Arabia and to identify the level of renal impairment. This can lead to more effective healthcare delivery and subsequently prevent the transition of renal damage to ESRD. Identifying the level of renal impairment can bring awareness among people about risk factors and prevent further kidney damage. This can reduce the medical expenses for diabetes patients, and, thus, the financial burden on the health ministry of Saudi Arabia. Consequently, this study can significantly contribute to community health and awareness in the region of Wadi Ad Dawaser and its surrounding areas. In the present study, various parameters of the renal function were measured in patients: blood urea, creatinine, electrolytes (sodium and potassium), creatinine clearance, microalbuminuria, and urine analysis. All the renal function parameters obtained from the laboratory reports of the patients were subjected to statistical analysis and compared to international standard levels. The cloud data transaction of our proposed system is not only stored but also processed in a cloud environment. Hence, we proposed a privacy-preserved transaction between the cloud server and the client. To achieve this setting, the data must be processed in the run time on a virtualized cloud platform. We used the CP-ABE tool kit, which is based on the PBC library and supports

a Linux-based operating system. Fig. 4 shows our proposed algorithm and the key transactions under various conditions of cloud operations, such as setup, key generation, encryption, and decryption.

Fig. 5 consists of three graphs a-c. They display the key setup time, key generation, and time taken by the cloud server, respectively. Our approach considers different file sizes for the encryption and decryption of the transactions between the client and the cloud server under different privileges.



**Figure 5:** Result on our implemented prototype system on Linux system i7 second generation

## 5 Conclusion

In the present study, kidney function in patients with diabetes was assessed. Owing to a lack of early detection and clinical results, diabetes can slowly progress to DKD and eventually to ESRD. The major challenge is to detect it in the initial stage and the DKD onset, thereby improving the DKD outcomes, and thus, reducing cardiovascular morbidity. Our study can strengthen therapeutic interventions for preventing KD, which can reinforce health policies and probably reduce the burden on the health sector and government expenses on DKD. Hence, thorough management of CKD patients includes a good assessment of the risk factors, a complete physical examination, and laboratory studies. Consequently, early detection of DKD through most optimized hyper-parameter machine learning is recommended. We used a grid search technique in our proposed framework to optimize the parameters of different classifiers to drive their learning capabilities. Further research is needed to develop federated learning algorithms that can divide features among different clients in a more powerful and scalable way. The learning

process will also be enhanced by highly efficient homomorphic encryption algorithms. In addition, adaptive learning algorithms that protect privacy, including hybrid algorithms and anti-malicious attack algorithms, should receive more attention.

**Funding Statement:** The author received no specific funding for this study.

**Conflicts of Interest:** The author declares that they have no conflicts of interest to report regarding the present study.

## References

- [1] M. Kolhar, M. M. Abu-Alhaj and S. M. Abd El-atty, "Cloud data auditing techniques with a focus on privacy and security," *IEEE Security & Privacy*, vol. 15, no. 1, pp. 42–51, 2017.
- [2] K. Saranya and K. Premalatha, "Multi attribute case based privacy-preserving for healthcare transactional data using cryptography," *Intelligent Automation & Soft Computing*, vol. 35, no. 2, pp. 2029–2042, 2023.
- [3] K. Anand, A. Vijayara and M. Vijay Anand, "Privacy preserving framework using Gaussian mutation based firebug optimization in cloud computing," *The Journal of Supercomputing*, vol. 1, no. 24, pp. 9414–9437, 2022.
- [4] S. Ben Hamida, H. Mrabet, S. Belguith, A. Alhomoud and A. Jemai, "Towards securing machine learning models against membership inference attacks," *Computers, Materials & Continua*, vol. 70, no. 3, pp. 4897–4919, 2022.
- [5] H. Fang and Q. Qian, "Privacy preserving machine learning with homomorphic encryption and federated learning," *Future Internet*, vol. 13, no. 4, pp. 94–99, 2021.
- [6] M. Sugahara, W. L. W. Pak, T. Tanaka, S. C. Tang and M. Nangaku, "Update on diagnosis, pathophysiology, and management of diabetic kidney disease," *Nephrology*, vol. 26, no. 6, pp. 491–500, 2021.
- [7] N. Altemtam, J. Russell and M. El Nahas, "A study of the natural history of diabetic kidney disease (DKD)," *Nephrology Dialysis Transplantation*, vol. 27, no. 5, pp. 1847–1854, 2021.
- [8] R. J. MacIsaac, E. I. Ekinici and G. Jerums, "Markers of and risk factors for the development and progression of diabetic kidney disease," *American Journal of Kidney Diseases*, vol. 63, no. 2, pp. S39–S62, 2014.
- [9] R. Thorman, M. Neovius and B. Hylander, "Clinical findings in oral health during progression of chronic kidney disease to end-stage renal disease in a Swedish population," *Scandinavian Journal of Urology and Nephrology*, vol. 43, no. 2, pp. 154–159, 2009.
- [10] J. Zhou, Y. Zhou, B. Wang and J. Zang, "Human–cyber–physical systems (HCPSs) in the context of new-generation intelligent manufacturing," *Engineering*, vol. 5, no. 4, pp. 624–636, 2019.
- [11] H. Li, X. He, J. He, P. Vijayakumar, X. Zhang *et al.*, "A verifiable privacy-preserving machine learning prediction scheme for edge-enhanced HCPSs," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 8, pp. 5494–5503, 2021.
- [12] S. Feng, L. Jin, Y. Wang, W. Zhao and H. Peng, "An efficient algorithm based on spectrum migration for high frame rate ultrasound imaging," *CMES-Computer Modeling in Engineering & Sciences*, vol. 126, no. 2, pp. 739–754, 2021.
- [13] M. Kato and R. Natarajan, "Epigenetics and epigenomics in diabetic kidney disease and metabolic memory," *Nature Reviews Nephrology*, vol. 15, no. 6, pp. 327–345, 2019.
- [14] J. P. Weng and Y. Bi, Y., "Epidemiological status of chronic diabetic complications in China," *Chinese Medical Journal*, vol. 128, no. 24, pp. 3267–3269, 2015.
- [15] D. Ren, W. Kang and G. Xu, "Meta-analysis of diagnostic accuracy of retinopathy for the detection of diabetic kidney disease in adults with type 2 diabetes," *Canadian Journal of Diabetes*, vol. 43, no. 7, pp. 530–537, 2019.
- [16] P. Rossing, "Diabetic nephropathy: Worldwide epidemic and effects of current treatment on natural history," *Current Diabetes Reports*, vol. 6, no. 6, pp. 479–483, 2006.
- [17] T. A. Reutens, "Epidemiology of diabetic kidney disease," *Medical Clinics*, vol. 97, no. 1, pp. 1–18, 2013.
- [18] B. C. Astor, S. I. Hallan, E. R. Miller III, E. Yeung and J. Coresh, "Glomerular filtration rate, albuminuria, and risk of cardiovascular and all-cause mortality in the US population," *American Journal of Epidemiology*, vol. 167, no. 10, pp. 1226–1234, 2008.

- [19] S. I. Hallan, E. Ritz, S. Lydersen, S. Romundstad, K. Kvenild *et al.*, “Combining GFR and albuminuria to classify CKD improves prediction of ESRD,” *Journal of the American Society of Nephrology*, vol. 20, no. 5, pp. 1069–1077, 2009.
- [20] L. U. Mailloux, A. G. Bellucci, B. M. Wilkes, B. Napolitano, R. T. Mossey *et al.*, “Mortality in dialysis patients: Analysis of the causes of death,” *American Journal of Kidney Diseases*, vol. 18, no. 3, pp. 326–335, 1991.
- [21] N. Sultan, “Making use of cloud computing for healthcare provision: Opportunities and challenges,” *International Journal of Information Management*, vol. 34, no. 2, pp. 177–184, 2014.
- [22] J. Jankowski, J. Floege, D. Fliser, M. Boehm and N. Marx, “Cardiovascular disease in chronic kidney disease: Pathophysiological insights and therapeutic options,” *Circulation*, vol. 143, no. 11, pp. 1157–1172, 2021.
- [23] N. Lascar, J. Brown, H. Pattison, A. H. Barnett, C. J. Bailey *et al.*, “Type 2 diabetes in adolescents and young adults,” *The Lancet Diabetes & Endocrinology*, vol. 6, no. 1, pp. 69–80, 2018.
- [24] U. Eduok, A. Abdelrasoul, A. Shoker and H. Doan, “Recent developments, current challenges and future perspectives on cellulosic hemodialysis membranes for highly efficient clearance of uremic toxins,” *Materials Today Communications*, vol. 27, no. 1, pp. 1021–1083, 2018.
- [25] W. Pan and Y. Kang, “Gut microbiota and chronic kidney disease: Implications for novel mechanistic insights and therapeutic strategies,” *International Urology and Nephrology*, vol. 50, no. 2, pp. 289–299, 2018.
- [26] M. G. Shiraz, R. H. Khokhar and R. Buyya, “A review on distributed application processing frameworks in smart mobile devices for mobile cloud computing,” *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1294–1313, 2012.
- [27] B. Dhiyanesh, K. Karthick, R. Radha and A. Venaik, “Iterative dichotomiser posteriori method based service attack detection in cloud computing,” *Computer Systems Science and Engineering*, vol. 44, no. 2, pp. 1099–1107, 2023.
- [28] M. Kolhar, F. Al-Turjman, A. Alameen and M. A. Abualhaj, “A three layered decentralized IoT biometric architecture for city lockdown during COVID-19 outbreak,” *IEEE Access*, vol. 8, no. 1, pp. 163608–163617, 2020.
- [29] A. Abdelaziz, A. S. Salama, A. M. Riad and A. N. Mahmoud, “A Machine Learning Model for Predicting of Chronic Kidney Disease Based Internet of Things and Cloud Computing in Smart Cities,” *Gewerbestrasse, Cham, Switzerland: Springer*, 2019. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-3-030-01560-2\\_5](https://link.springer.com/chapter/10.1007/978-3-030-01560-2_5).
- [30] P. M. Kumar, S. Lokesh, R. Varatharajan, G. C. Babu and P. Parthasarathy, “Cloud and IoT based disease prediction and diagnosis system for healthcare using fuzzy neural classifier,” *Future Generation Computer Systems*, vol. 86, no. 1, pp. 527–534, 2018.
- [31] P. K. Sahoo, S. K. Mohapatra and S. L. Wu, “Analyzing healthcare big data with prediction for future health condition,” *IEEE Access*, vol. 4, no. 1, pp. 9786–9799, 2020.
- [32] R. Jayaram and S. Prabakaran, “Onboard disease prediction and rehabilitation monitoring on secure edge-cloud integrated privacy preserving healthcare system,” *Egyptian Informatics Journal*, vol. 22, no. 4, pp. 401–410, 2021.
- [33] A. Abdelaziz, M. Elhoseny, A. S. Salama and A. M. Riad, “A machine learning model for improving healthcare services on cloud computing environment,” *Measurement*, vol. 119, no. 12, pp. 117–128, 2018.
- [34] I. Kavakiotis, O. Tsave, A. Salifoglou, V. Maglaveras, I. Vlahavas *et al.*, “Machine learning and data mining methods in diabetes research,” *Computational and Structural Biotechnology Journal*, vol. 15, no. 12, pp. 104–116, 2017.
- [35] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju *et al.*, “Predicting diabetes mellitus with machine learning techniques,” *Frontiers in Genetics*, vol. 9, no. 2, pp. 515–522, 2018.
- [36] P. M. Shakeel, S. Baskar, V. S. Dhulipala and M. M. Jaber, “Cloud based framework for diagnosis of diabetes mellitus using K-means clustering,” *Health Information Science and Systems*, vol. 6, no. 1, pp. 16–25, 2018.
- [37] M. Makino, R. Yoshimoto, M. Ono, T. Itoko, T. Katsuki *et al.*, “Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning,” *Scientific Reports*, vol. 9, no. 1, pp. 1–9, 2019.

- [38] C. Esposito, A. De. Santis, G. Tortora, H. Chang and K. K. R. Choo, "Blockchain: A panacea for healthcare cloud-based data security and privacy," *IEEE Cloud Computing*, vol. 5, no. 1, pp. 31–37, 2018.
- [39] H. Löhr, A. R. Sadeghi and M. Winandy, "Securing the e-health cloud," in *Proc. of the 1st acm Int. Health Informatics Symp.*, New York, USA, pp. 220–229, 2010.
- [40] P. Deshmukh, "Design of cloud security in the EHR for Indian healthcare services," *Journal of King Saud University-Computer and Information Sciences*, vol. 29, no. 3, pp. 281–287, 2017.
- [41] C. Thota, P. Sundarasekar, G. Manogaran, R. Varatharajan and M. K. Priyan, "Centralized fog computing security platform for IoT and cloud in healthcare system," in *Fog Computing: Breakthroughs in Research and Practice*, vol. 18. India: IGI global, pp. 365–378, 2018. [Online]. Available: <https://www.igi-global.com/book/fog-computing-breakthroughs-research-practice/190550>.
- [42] A. D. Dwivedi, G. Srivastava, S. Dha and R. Singh, "A decentralized privacy-preserving healthcare blockchain for IoT," *Sensors*, vol. 19, no. 2, pp. 326–335, 2019.
- [43] K. N. Griggs, O. Ossipova, C. P. Kohlios, A. N. Baccarini, E. A. Howson *et al.*, "Healthcare blockchain system using smart contracts for secure automated remote patient monitoring," *Journal of Medical Systems*, vol. 42, no. 7, pp. 130–145, 2018.