Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

An Efficient and Privacy-Preserving Hybrid Classification model for Patient-Centric Clinical Decision Support System

S. P. Rathinaeaswari[1,*], V. Santhi[2]

[1] Department of Computer Science and Engineering, Veerammal Engineering College ,Tamil Nādu, India
[2] Department of Computer Science Engineering, PSG College of Technology, Coimbatore, India

ARTICLE INFO

Article history:
Received 14 June 2023
Received in revised form 28 July 2023
Accepted 4 October 2023
Available online 19 October 2023

Keywords:
Data Mining; Double Encryption; Rotation Forest; Clinical Support System; Encryption

ABSTRACT

A lot of focus has recently been placed on clinical decision support systems, that use advanced data mining methods to assist clinicians in making wise decisions. Along with increasing diagnosis accuracy, clinical decision support systems (CDSS) have the added benefit of speeding up diagnosis. Data security is crucial in this system. In this research, we offer EPPCD (Efficient and Privacy preserving Patient-centric Clinical Decision) support system to assist physicians in predicting illness risks of patients in a privacy-preserving manner. This system is proposed to solve the privacy difficulties present in the CDSS. The fine-grained access control enabled by the novel Double Encryption Ciphertext Policy Attribute-Based Encryption (DE-CPABE) technique is found to be a potential solution to this issue. In the proposed system, the past patients' historical data are stored in cloud and can be used to train the hybrid Rotation Forest and AdaBoost classifier. Furthermore, extensive simulations used to evaluate performance show that our technology is capable of quickly and accurately determining a patient's disease risk while maintaining their privacy. The proposed system model is divided into five parties: Cloud Platform (CP), Trusted authority (TA), processing unit (PU), data provider (DP), and undiagnosed patient (PA).

1. Introduction

CDSSs combine a significant amount of clinical data to give doctors the information they need to support medical decisions and help manage treatment protocols [1]. The improvement of patient care through data-driven, patient-centered decision support systems is a growing area of interest for medical informatics. Predictive analytics can now be employed regularly by physicians thanks to the increasing integration of artificial intelligence (AI) algorithms into CDSSs. These systems' main goal is to enhance medical decision-making through the use of data-driven methods. CDSS attempts to improve healthcare delivery by merging particular patient data, clinical knowledge, and other health information into medical decisions.

* *Corresponding author.*
E-mail address: rathinaeaswarisp@gmail.com

https://doi.org/10.37934/araset.33.1.299316

Software is used to develop patient-specific assessments or recommendations after matching a patient's features to a computerised clinical knowledge base. These recommendations are then sent to the doctor for consideration. This is known as a classic CDSS [2]. Nowadays, CDSSs are frequently used at the point of treatment, enabling the doctor to combine their knowledge with information or suggestions from the CDSS. And yet many CDSS are being developed, some of which may make use of information and observations that would otherwise be inaccessible to or unclear to people [3]. CDSS currently frequently use web apps or interface with electronic health record (EHR) and computerised provider order entry (CPOE) systems. They can be controlled using smartphones, tablets, and computers, and other tech tools such as wearable medical equipment and biometric sensors which might or might not have the ability to output data directly to the device or connect to EHR databases [4]. The CDSS provides physicians with access to patient medical records that can assist in making a diagnosis. The patient's symptoms, medical history, pertinent test results, and relevant research findings are all included here.

The use of computerised decision support systems (CDSSs) can help practitioners and healthcare professionals make better decisions by providing rapid access to EHR. A CDSS interfaces with practitioners and electronic medical record systems to accept patient data as input and to provide cautions or reminders for patient diagnosis. A CDSS must have access to healthcare data and information stored in data and knowledge bases [5]. By foreseeing behaviours and future trends on everything, the amazing progress of data mining method over the past two decades has had a significant impact on the revolution of human living. This technique can transform stored data into relevant information. These methods are excellent for offering decision support in a medical setting. A new system in the healthcare sector should be viable to provide a much quicker and less expensive method for diagnosis in order to reduce diagnosis time and increase diagnosis accuracy. A lot of attention has recently been paid to CDSS, which use a variety of data mining techniques to help doctors diagnose patient diseases that have similar symptoms [6]. Preventing the unauthorised exposure of patient medical data is one of the primary challenges [7-10].

A wide range of healthcare stakeholders may be interested in the use of medical data. One online direct-to-consumer service provider, for instance, provides disease risk prediction for specific patients. Without adequate security, a patient could be reluctant to provide CDSS access to his medical records out of concern that they would be misused or leaked. Therefore, it is essential to safeguard patient medical information. However, maintaining the medical data privacy is insufficient to propel the CDSS as a whole forward and prosper. Since the classifier is regarded as the service provider's own asset, it cannot be shared with third parties in order to anticipate the patient's sickness. If not, outside parties might misuse the classifier to forecast a patient's illness, which would hurt the service provider's bottom line. So, for the CDSS, protecting service providers' privacy is just as important as protecting the privacy of medical data about patients. In this research, we offer EPPCD (Efficient and Privacy preserving Patient-centric Clinical Decision) support system that is based on naive Bayesian classification to assist physicians in predicting illness risks of patients in a privacy-preserving manner. This system is proposed to solve the privacy difficulties present in the CDSS.

The promise of attribute-based encryption (ABE) cryptography to address the scalability problem of modern system is well-known. By assigning unique labels to the secret key and ciphertext, it provides a granular access control mechanism with great flexibility and efficiency. The ABE system adds an extra degree of security to big data processing and analytics by protecting users' data and privacy. Quantum computing, however, is emerging as a new technology with the potential to drastically alter the state of online security and personal data protection. Similarly to traditional ABE methods, modern encryption is not designed to be immune to the effects of quantum computing. Recent studies have mostly focused on the broad strokes of pairing-based ABE schemes, namely their

scalability, adaptability, data secrecy, and scope; characteristics like user revocation, scalability, and flexibility have gotten less attention.

The key contributions of this work are more precise.

1. First, we provide a safe and EPPCD that uses Double Encryption CPABE algorithm to maintain data privacy without disclosing any personal health information.
2. We provide a novel hybrid classifier called RandRotBoost to classify diseases.
3. The service provider can utilise the trained classifier to discreetly diagnose the patient's ailments based on his symptoms.
4. Simulation results shows that the proposed system functions well in terms of all privacy requirements and performance requirements.

The rest of the paper is organized as follows: Section 2 shows the related works, Section 3 shows the preliminaries, Section 4 shows the System Model, Section 5 shows the Proposed Double Encryption CPABE (DE-CPABE) system, Section 6 shows the Proposed RandRotBoost classifier, Section 7 shows the performace analysis and the paper is concluded in Section 8.

2. Related Works

A CDSS system based on Deep Neural Network that aids in disease diagnosis using a patient's Electronic Health Record (HER) is proposed by Khan *et al*., [11]. Because it is a generalist system that may detect numerous diseases at once and is not intended for the diagnosis of a single disease, the system is different and unique from other CDS systems. Natural language processing is used to translate a patient's medical and family history into a structured text that may be processed further. Multiple diseases can be recognised simultaneously using multi-label categorization. In order to categorise patients' symptoms, a classification model is constructed using data mining methods and previous medical data. Alabdulkarim *et al*., [12] proposed a CDSS for identifying new symptoms without subjecting patient information to various network threats. To secure user data, a homomorphic encryption cypher is employed. Since everyone will be using the same key pair, the technique also employs nonces to prevent one party from decrypting the data of other parties. This approach outperformed the Naive Bayes algorithm by 46.46%.

A CDSS was created by Sreejith *et al*., [13] to help doctors monitor Polycystic Ovarian Syndrome (PCOS). The suggested approach makes use of a classification model with machine learning algorithms and optimization techniques. Any classifier's performance can suffer by the presence of irrelevant features. This study examines how these features affect classification accuracy while also automatically locating and eliminating the unnecessary features from the dataset. For feature selection, a wrapper method is used. The best characteristics are identified using the Red Deer Algorithm (RDA), and they are assessed using a random forest (RF) classifier. The roaring and mating habits of red deer served as the basis for the optimization method known as RDA. A 50-estimator RF classifier is used to train and test the best dataset. The justifications for choosing RDA as the search strategy are its better exploration and exploitation capabilities. The proposed CDSS has 89.81% accuracy.

Through the incorporation of monitoring devices like medical devices and sensors for remote patient observation, cloud computing (CC) and IoT environments are widely used in a number of healthcare applications. Instead of depending on constrained storage and processing resources, the data created by IoT devices used in the medical area can be explored in the CC environment to provide better healthcare services [14]. In order to dramatically lower the mortality rate, early detection of chronic kidney disease is important. For the purpose of diagnosing CKD in the Internet of Things, Alsuhibany *et al*., [15] created an ensemble of deep learning-based CDSS (EDL-CDSS). It

uses medical data gathered by IoT devices and benchmark repositories to identify and categorise various stages of CKD. It also calls for the creation of an Adaptive Synthetic technology for the outlier detection procedure.

Bashir *et al*., [16] provide an ensemble-based CDSS voting mechanism to effectively predict cardiac disease. For testing and analysis, four benchmark heart disease datasets from the UCI repository were considered. The usefulness of the suggested ensemble scheme is demonstrated by comparing its performance with that of individual classifiers to that of five different ensemble systems utilising a variety of parameters. In comparison to existing ensemble schemes and individual classifiers, the suggested ensemble scheme has greater average accuracy (83%) according to the results analysis.

A cloud and IoT based clinical decision support system for the prediction and observation of Chronic Kidney Disease (CKD) and its severity level is proposed by Lakshmanaprabu *et al*., [17] in their study. The suggested framework gathers patient data from IoT devices that are attached to the user and stores it in the cloud along with the pertinent medical records that are obtained from the UCI repository. Additionally, DNN classifier is used to predict CKD and its severity degree. The performance of the DNN classifier is also enhanced using a feature selection technique based on Particle Swarm Optimization (PSO). Casal-Guisande *et al*., [18] described a CDSS to be utilised in the diagnosis of breast cancer with the goal of increasing evaluation accuracy and decreasing evaluation uncertainty. The system can learn from and interpret the patient's medical-healthcare data by integrating data augmentation techniques, and classification methods like bagged and k-neighbors trees.

A conceptual framework for creating data-driven (DD) predictive modeling–based clinical decision support systems (CDSSs) was proposed by Kovalchuk *et al*., [19] in their study. We have made an effort to bridge the gap between solutions that are acceptable to doctors in routine practise and experimental CDSS implementations that are extensively addressed in the literature with the suggested framework. The framework is built on a three-stage process in which the description of the DD model is completed together with references to real-world norms and an explanation of the predictions' outcomes and suggestions.

The machine learning algorithm in CDSSs classifies or forecasts the collected data. The trained model can be used to make predictions or make classifications as fresh data come in, and it can then be applied to a variety of real-world applications. The medical profession, which creates large volume of data is one of the popular fields. Higher number of data and features might result in a noticeable rise in noise and time loss when using machine learning algorithms as well as ensemble decision trees [20].

A supervised learning approach called ensemble learning [21] uses multiple learning models to improve prediction accuracy. It typically performs better than a standalone learning approach since its overall performance is better. One of the most well-known ensemble learnings is the Forest, which combines weak learners to create a stronger and more accurate learner. Strong learners can be further split into two types, boosting and bagging, in accordance with the sample methodology. Random sampling and voting on the training sample set constitute the fundamental principles of bagging. Random Forest [22] is an illustration of something that is usual. Another boosting ensemble learning technique is the AdaBoost algorithm [23]. It is a well-liked technique for raising quality and builds strong classifiers by linearly combining weak classifiers. CART (Classification And Regression Trees) are used as the foundation classifier in the boosting decision tree methodology. The decision tree algorithm in this instance can be thought of as a specific example of AdaBoost. Each weak classifier is trained individually throughout the training phase, and after obtaining their weights, the misclassified data are dynamically corrected [32]. The weight will change depending on whether the

preceding weak classifiers correctly identified the samples or not. By changing the weight, the most beneficial samples will get better with each new iteration [33]. The weighted sum of the prediction results produced by the weak classifiers, in the end, constitutes the final classifier's prediction result. Its sensitivity to noisy data needs to be solved. Rotation operation via PCA may be the solution to this noise sensitivity [34].

3. Preliminaries

The background and the hardness issues that will be used throughout the study are covered in this section.

- Bilinear group

The security of the proposed system is based on the bilinear map.
Definition 1
Consider cyclic multiplicative group M, M_1 of prime order p and generators m, m_1 respectively. A deterministic bilinear map function $f: M \times M \rightarrow M_1$ with following requirements.
•Bi-linearity : For all $x \in M, y \in M, a, b \in Z_p$, $f(x^a, y^b) = f(x, y)^{ab}$.
•Non-degeneracy: $f(m, m) \neq 1$.
•Efficiency :f must be a time efficient function.
Definition 2
(Discrete Logarithm Problem (DLP)). Given two group elements m and h, Find an integer x$\in Z_p$ such that h=m^x whenever such integer exist.
Definition 3
(Decisional Bilinear Diffie Hellman (DBDH) Problem): In prime order group M with generator g, on input m,m^a, m^b, m^c ∈M, check if c=ab or not.

- Classifiers

Let 'L' be a binary classification training data set, where L = {(x_1, y_1), (x_2 , y_2), · · · , (x_N , y_N)}, $x_i \in X \subseteq R^N$, y_i= {−1, +1}, N represents number of data, 'F' represents number of features, 'K' represents number of subsets, 'D' represents initial weight distribution, where D_i = (w_{1n}, · · · w_{1N}), 'i' represents iteration, 'S' represents sample, and 't' represents classifiers.

a. RotForest

F is randomly divided into K subgroups for the initial S in RotForest [24], and these subsets may intersect or be unconnected. After obtaining K, each feature subset has M features, where $M = F/K$. If the number of features cannot be divided, the rest of the features are added to the prior feature set. The new samples that are generated are based on how the feature subsets are divided, and this influences how the ensemble classifiers differ from one another. After K samples are resampled to form a bootstrap sample, PCA is then used to analyse the sample subset of classes. Principal component analysis (PCA) retains the principal component coefficients $\mathrm{x}_{v,j}^{(1)}$, $\mathrm{x}_{v,j}^{(2)}$, · · ·,$\mathrm{x}_{v,j}^{(M)}$ in order to preserve the relevant portions of the data and variability information. The data set is rotated along its K-axis after PCA transformation processing on each of the K sample subsets, and the coefficients are then merged to create the sparse rotation matrix R_i.

$$R_i = \begin{pmatrix} \mathrm{x}_{v,j}^{(1)}, \mathrm{x}_{v,j}^{(2)}, \cdots, \mathrm{x}_{v,j}^{(M)} & \cdots & [0] \\ \vdots & \ddots & \vdots \\ [0] & \cdots & \mathrm{x}_{v,j}^{(1)}, \mathrm{x}_{v,j}^{(2)}, \cdots, \mathrm{x}_{v,j}^{(M)} \end{pmatrix} \quad (1)$$

where R_i is rearranged to produce R_i^a according to the original features' order, making XR_i^a the new training set for the classifier.
By getting the new training set, the whole process of RotForest [7] is performed as shown below.

Input: L, loss: l.
For $t = 1, 2, \cdots, T$
1. Subdivide the F feature set into K subgroups, with $M = F/K$ features in each subset;
2. To obtain the principal component coefficients, do PCA on a bootstrap sample of each subset;
3. Repeat steps 1 and 2 for K times and add the K principal component coefficients to the rotation matrix.
4. Rearrange the rotation matrix in accordance with the order of the original feature set, and the training set for the classifier G_t (x) is XR_t^a.
5. Discover the classification outcome of G_t (x).
Output G(x) = $\frac{1}{M}\sum_{t=1}^{T} G_t(\mathrm{x})$ (2)

b. AdaBoost
One of the key components of our suggested methodology is the use of AdaBoost to connect each rotation result [25]. The approach that follows demonstrates the standard, in-depth processes of AdaBoost.

$Input: L, loss: l, initial\ weight\ D$
$For\ t = 1, 2, \cdots, T$:
1. From l, find the classifier error of $G_t(\mathrm{x})$ on training set:
$e_t = \mathrm{P}((G_t(x_i) \neq y_i)) = \sum_{i=1}^{N} w_{ti} \mathrm{I}\ G_t\ (x_i)\ \neq y_i)$ (3)
2. Find the classification of G_t (x):
$\alpha_t = \frac{1}{2} \log \frac{1 - \mathrm{e}_t}{\mathrm{e}_t}$ (4)
3. Update the weight distribution:
$\mathrm{D}_{t+1} = (\mathrm{w}_{t+1,n}, \cdots, \mathrm{w}_{t+1,N})$ (5)
4. Build $f(x) = \sum_{t=1}^{M} \alpha_t G_t(x)$ (6)
where $\mathrm{w}_{t+1,n} = \frac{\mathrm{w}_{t,n}}{Z_t}\ exp(-\alpha_t y_n G_t(x))$, D_{t+1} is a probability distribution.
Output: $G(x) = sign(f(x)) = sign \sum_{t=1}^{T} \alpha_t G_t(x)$ (7)

c. RotBoost
By just fusing RotBoost and AdaBoost, RotBoost [26] has a similar aspect. To train several base classifiers at once, AdaBoost minimises the residual from the sum of all classifier outcomes. According to the explanations of RotForest and AdaBoost in the previous two subsections, the following explanation
1. To get the training data, follow the same procedures as in RotForest steps 1 to 4.
2. Next, use the decision tree as the base classifier.
3. Calculate the residual from the output of $G(x)$ in steps 1 through 3, then update the weights of each classifier based on the overall performance of all classifiers.

4. System Model

In our approach, we primarily concentrate on safely training the RandRotBoost classifier and utilise them to determine a patient's ailment without disclosing any of the patient's personal information. To be more precise, we categorise EPPCD into five parties in order to describe the system model: the data provider (DP), trusted authority (TA), processing unit (PU), undiagnosed patient (PA), and cloud platform (CP). Figure 1 depicts the overall system model of our EPPCD.

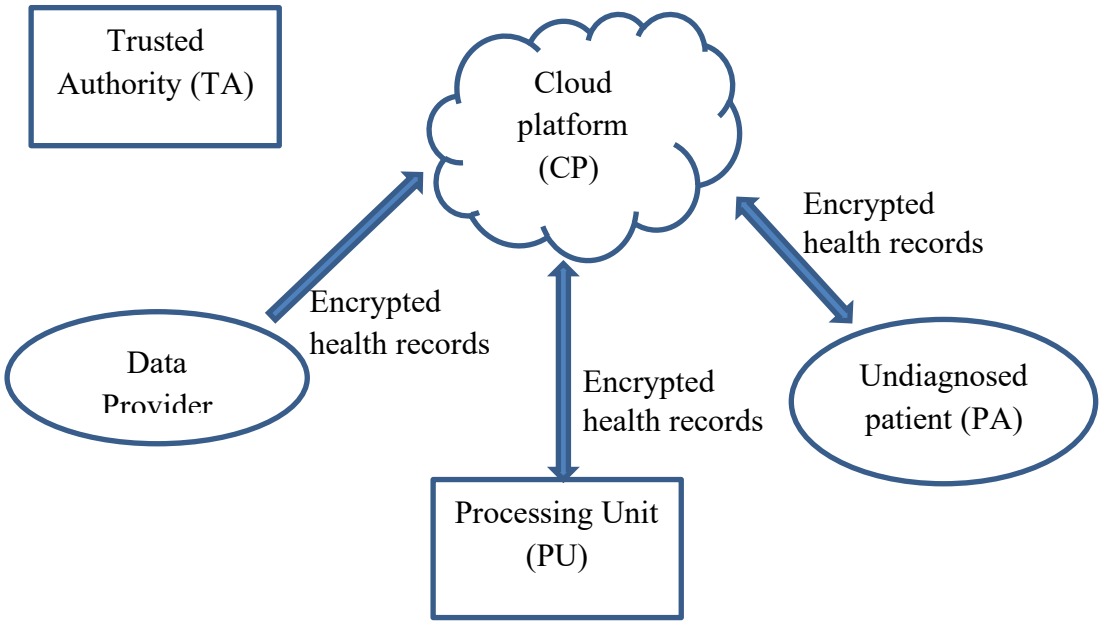

Fig. 1. Proposed system model

a) Trusted Authority:

All parties involved in the system place their trust in TA, which is in charge of managing all of the private keys employed within. A trustworthy third party that can verify the authenticity of a certificate issued by another system is required for verification purposes. As a result, Certificate Authorities are non-governmental organisations (NGOs) that issue certificates as trustworthy third parties. A certificate authority will verify the identity of the requesting party prior to providing a certificate. After the certificate has been granted, its details will be stored in a database that anybody may view. Users may verify the authenticity of their certificates by looking them up in the central database. There are varying degrees of security certifications that may be issued by certificate authority. For instance: Authentication of Secure Email Clients and Authentication through a Server etc.

b) Cloud Platform:

All of the system's data can be stored and managed on CP's limitless storage capacity. Other parties who have a limited amount of storage space might contract with CP to store their data. Additionally, CP includes some computation capabilities that allow it to make computations over the data that is stored.

c) Data Provider:

The proposed RandRotBoost classifier may be trained using historical medical data from DP, which includes patient symptoms and diagnoses. All the information are outsourced to CP and stored in it.

d) Service Provider or Processing Unit (PU):

PU can be a business or a medical facility that offers direct-to-consumer internet services that offers individualised risk assessments for a variety of diseases based on client symptoms. RandRotBoost classifier is created by PU using previous medical data, and the model is then used to forecast the disease risk of patients who have not yet been diagnosed.

e) Undiagnosed Patient (PA):

During doctor visits or directly from patients, PA receives certain symptom information like heart rate, blood pressure and weight. For disease diagnosis, these symptoms are forwarded to PU.

Privacy Requirements

The success of the patient's ailment diagnosis depends on privacy. According to our privacy model, DP is trustworthy because it gives accurate past medical information. The internal party PU is seen as curious-but-honest because it is curious about the personal medical histories of both DP and PA but rigorously adheres to the system's standards. In this arrangement, PA and CP are equally curious but trustworthy parties. While CP is curious about all the data in the system, PA is curious about PU's classifier. Additionally, an outside enemy can listen in on all system communications and is curious about the data being exchanged. Therefore, the following privacy requirements in EPPCD should be met in order to prevent both internal party from external eavesdropping and information leakage.

- **DP's Privacy:**

A documented case record of a patient's symptoms and a proven ailment can be found in DP's previous medical data. Using these historical medical data, the proposed RandRotBoost classifier can be trained. These specific data include certain delicate information that is closely tied to patients' privacy. During transmission and storage, it cannot be directly accessed by unreliable parties. Otherwise, due to the privacy information leak, DP does not send its own information to other parties. As a result, our system should protect DP privacy.

- **PU's Privacy:**

The proposed RandRotBoost classifier is trained by PU using historical medical data, and conditional probabilities about the classifier are obtained. These possibilities are regarded as a resource for PU and are not disclosed to patients or third parties while a condition is being diagnosed.

- **PA's Privacy:**

Some of the symptom data in PA are sensitive and are not exposed to other parities directly. Additionally, the diagnosis results are also very private information that must not be disclosed to third parties. If more information is required for processing, PA may permit the authorised to communicate the diagnosis results.

Design Goals

Our system design will provide the following performance and privacy assurances in order to help UP receive a secure medical decision.

- **The Proposed System Should Meet the Needs for Privacy Protection:**

As previously said, if CDSS does not take the privacy standards into account, patients' highly sensitive information will be disclosed to CP, PU, and unauthorised parties in the patient's medical decision. It will permit the patient to give CDSS its own data despite not wanting to. Additionally, PU

is always a profitable business, preventing the exposure of his own data to other participants in the system. As a result, the suggested approach ought to concurrently protect both PU and PA's privacy.

- **Computation Efficiency Should Be Achieved by the Proposed System:**

The patient's computational capabilities are constantly finite and cannot support overload processing. The suggested system should take computing efficiency into account to allow patient-centric diagnosis results retrieval from CP in a timely manner. It is crucial to give PA access to real-time diagnosis results as a result.

- **The Proposed System Should Achieve better classification results:**

After training the system using the proposed RandRot Boost classifier, it should achieve better classification.

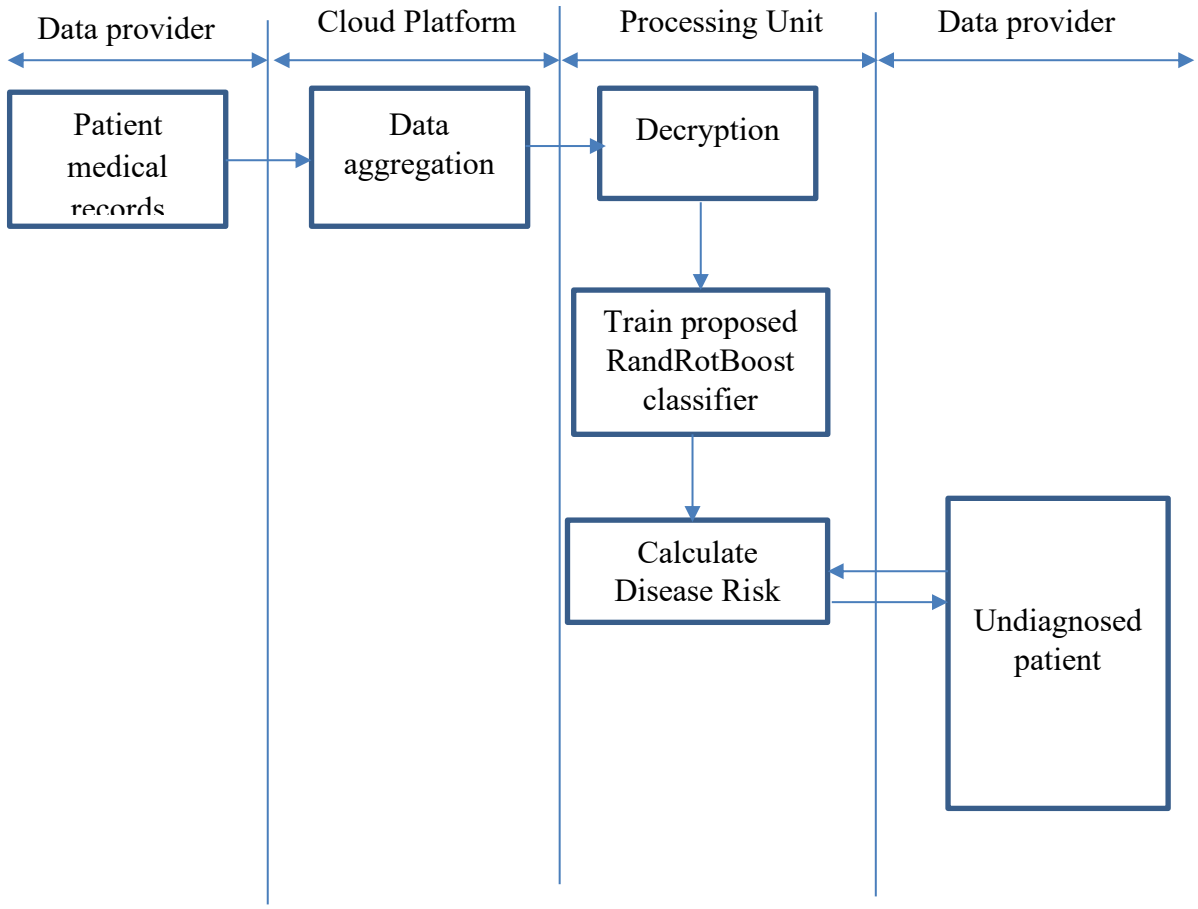

Fig. 2. Overall activity diagram

5. The Proposed Double Encryption CPABE (DE-CPABE) system

The following polynomial algorithms make up the suggested scheme.

•**Setup:** It is controlled by a central authority (CA), which generates the system's private and public parameters. The following is how CA carries it out.

- A bilinear group G_0 of prime order p with generator g is selected.
- The exponents $y, \beta \in_R Z_p$ are selected.
- Compute $h = g^{\beta}, Y = e(g, g)^{y}$
- Compute Master Public Key MPK=G_0, g, h, Y
- Compute Master Secret Key MSK=(β,y)

For each system attribute i

- Chooses exponent $\alpha_i \in_R Z_p$.
- Finds Public Parameter $PK_i = g^{\alpha_i}$
- Finds Secret Parameter $SK_i = \alpha_i$

Keygen (MSK,u) : It is run by the CA to generate the secret key (SK) for user u. The CA uses this process to create a portion of the user's secret key. The following sub-algorithms make up it.

- Chooses $r \in_R Z_p$
- Compute secret key $SK_u = g^{(y+r)/\beta}$
- Compute public key $PK_u = g^r$
- Set attribute list $L_u = \emptyset$
- For each user attribute i
- It computes $D_i = (g^r)^{\alpha_i}$ and $L_u = L_{u+i}$

L denotes attributes list. As can be seen from Keygen, the r value for each user is different which helps to prevent the collusion attack among users.

RKGen(MPK,W,W',SKL) :

To create a re-encryption key for the proxy servers, the user executes RKGen. It is run by the user u containing the attribute set $AS \subseteq L$ and it satisfy access policy W. By using the proxy re-encryption key provided by this approach, ciphertext with access policy W can be changed to ciphertext with access policy W'.

- Generates $d, g_1 \in_R Z_p$.
- Finds $C = Encrypt\ (MPK, g_1^{nd}, W')$
- Computes $R = D \prod_{v_{i,j} \in AS} (D_{i,j} g_1^{d})$
- $RK_{AS \to W'} = \langle C, R, g^r \rangle$

Encrypt (M, W, PK1, PK2, …, PKN):

The sender executes the command to convert the plaintext into ciphertext based on the access policy. This algorithm is composed of the following sub-algorithm. The sender executes it while using message M and the set of attributes for the policy. It takes the next few actions.

- Chooses exponent $s \in_R Z_p$.
- Computes $C_1 = MY^s$
- Computes $C_2 = g^s$
- Computes $C_3 = (\prod_{t \in W} PK_i)^s = (\prod_{t \in W} g^{\alpha_i})^s$
- Computes $C_4 = (h)^s = g^{\beta s}$
- Final ciphertext CT=$\{C_1, C_2, C_3, C_4, W\}$

RKEncrypt (CT_W, $RK_{AS\rightarrow W'}$):
It is run by the proxy server to convert the CT_W to $CT_{W'}$. There exist attribute set AS⊆L and it satisfy access policy W.

$C' = \frac{C_1 e(C_2, g^r)}{e(C_3, R)} = \frac{M}{e(g,g_1)^{nsd\beta}}$

$CT_{W'} = \langle C', C, C_3 \rangle$

Decrypt (SK, CT):
If the access criteria are followed, the receiver utilises this to decipher the plaintext from the ciphertext; otherwise, a random message is transmitted. The receiver controls it by receiving both the CT and the SK as input. Otherwise, a random message is returned if the policy is satisfied. Take for granted that $AS = W$ and $AS \subseteq L$. Based on CT, the proxy's re-encrypted original ciphertext, it was separated into two portions.

$$CT = \frac{C_1 e(C_2, g^r).e(C_3, g^r)}{e\left(C_4, g^{\frac{y+r}{\beta}}\right).e(C_2, \prod_{t\in AS} g^{\alpha_t})^r)}$$

$$= \frac{M.e(g,g)^{ys}.e(g,g)^{rs}}{e\left(C_4, g^{\frac{y+r}{\beta}}\right).e(C_2, \prod_{t\in AS} g^{\alpha_t})^r)}$$

=M

Here $p = \sum_{t\in W} \alpha_t$ and $q = \sum_{t\in AS} \alpha_t$

If CT is a re-encrypted ciphertext then,

g_1^{nd}=Decrypt (SK, CT)

$$= \frac{M.e(C_3, g_1^{nd})}{e((g,g_1)^{nsd\beta})}$$

=M

6. The Proposed RandRotBoost classifier

Based on AdaBoost, ensemble learnings and PCA, the proposed method RandRotBoost is designed as follows.

In order to use the benefits of RotForest and AdaBoost a hybrid method named as RandRotBoost is proposed. The training procedure is such that each rotation component conducts randomly generated featured projections on randomly generated segment subsets before processing PCA conversion. AdaBoost also connects each tree between rotations. Instead of predefining K as the number of rotation sample subsets, this technique enhances the diversity of features by applying a random number to randomly select feature subsets at each round. Following PCA conversion, it is possible to keep features' useful information to the fullest extent while simultaneously successfully removing duplicate information and unimportant features.

Input: L, loss: l, initial weight D For f in F:
1. $Randomly\ select\ K$:
(a) $Split\ the\ feature\ set\ F\ into\ K\ subsets;\ each\ subset\ has\ M = F/K\ features;$
$(b)\ Process\ PCA\ to\ get\ a\ rotation\ matrix;$
$(c)\ Apply\ decision\ tree\ to\ perform\ the\ classification\ and\ calculate\ the\ weight\ of\ each\ tir$
2. $AdaBoost\ connect$:
$(a)\ Calculate\ the\ classifier\ error\ of\ g_f(XR_f^a)\ \ from\ l\ on\ the\ training\ set$:
g_f= P $((g_f(x_i) \neq y_i))$ $= \sum_{i=1}^{N} w_{fi} l(g_f(x_i) \neq y_i)$
$(b)\ Calculate\ the\ classification\ of\ g_f(x)$:

$$\alpha_f = \frac{1}{2} log \frac{1-e_f}{e_f} \quad (9)$$

(c) *Update the weight distribution*:

$$D_{f+1} = w_{f+1,n}, \cdots, w_{f+1,N}) \quad (10)$$

(d) *Build the updated combination*:

$$g_f(x) = sign\left(\sum_{i=1}^{f} \alpha_i G_i(x)\right)$$

$Output$: *The final classifier is as follows*:

$$G(x) = sign(f(x)) = sign\left(\sum_{f=1}^{F} \alpha_f G_f(x)\right)$$

7. Performance Analysis

Using a unique Java-built simulator, we assess the EPPCD's computation costs. On a test computer with a 6 GB of RAM and a single 2.5 GHz two-core processor, this experiment was conducted. We take into account two datasets in the experiment. A genuine dataset termed the acute inflammations dataset (AID) is used from the UCI machine learning repository [27]. Using this dataset and our EPPCD, we evaluate how well the naive Bayesian classifier performs. An expert in medicine developed the AID as a dataset to evaluate the expert system that was used to make a provisional diagnosis of two disorders of the urinary system. There are 120 instances in this dataset. Each instance consists of two decisions (NRPO (nephritis of renal pelvis origin) and IUB (inflammation of the urinary bladder)) and six attributes (occurrence of nausea, temperature, lumbar pain, burning of the urethra, micturition pains, urine pushing, swelling of the urethra outlet, itch). With the exception of temperature, every attribute and decision may be stated as a binary bit 1 (YES) or 0 (NO). In the dataset, the temperature value ranges from 35.5 to 41.5 C. We first train the proposed RadRotBoost classifier using EPPCD, and then we assess the classifier's success rate using this classifier and AID. The result is shown in Table 1.

Table 1
Classifier Output

Disease	IUB	NRPO
Yes	100%	100%
No	91%	93%

We can observe from Table 1 that all cases with IUB and NRPO can be effectively diagnosed without receiving a false negative diagnosis. However, AID has a problem with false positive diagnoses. IUB and NRPO diagnoses have false positive rates of 26.23% and 14.29%, respectively.

4.2.1 Accuracy

One of the most common means to evaluate performance is accuracy. The accuracy rate attained by various classifiers is displayed in Figure 3.

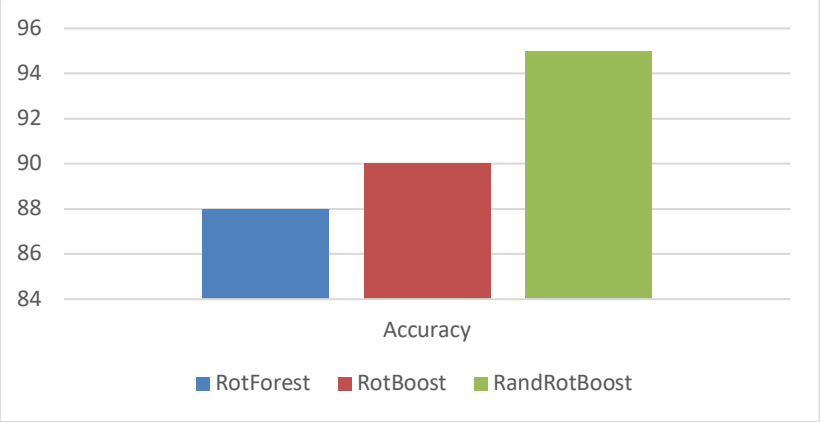

Fig. 3. Accuracy comparison of various classifiers

From Figure 3 it is observed that Rotforest achieved 88% accuracy, Rotboost achieved 90% accuracy and the proposed RandRotBoost achieved 95% accuracy which is better than the other two methods.

Precision

Precision reflects the ability of the model to distinguish between negative samples. Figure 4 shows the precision rate achieved by various classifiers.

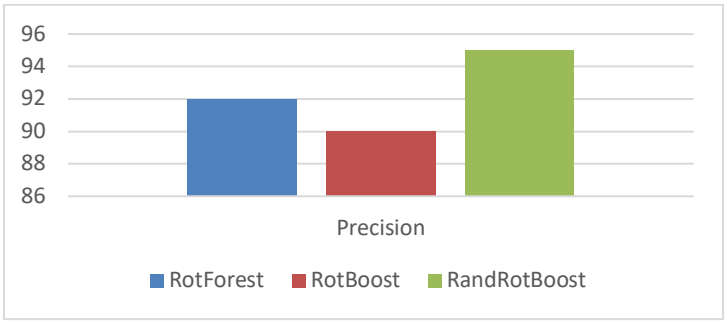

Fig. 4. Precision comparison of various classifiers

From Figure 4 it is observed that Rotforest achieved 92% precision, Rotboost achieved 90% precision and the proposed RandRotBoost achieved 95% precision which is better than the other two methods.

Recall

The assessment of the models' completeness and capacity to classify positive findings is taken into account when evaluating models using Recall. This kind of competence is essential for the medical sector to identify sickness. Figure 5 shows the recall rate achieved by various classifiers.

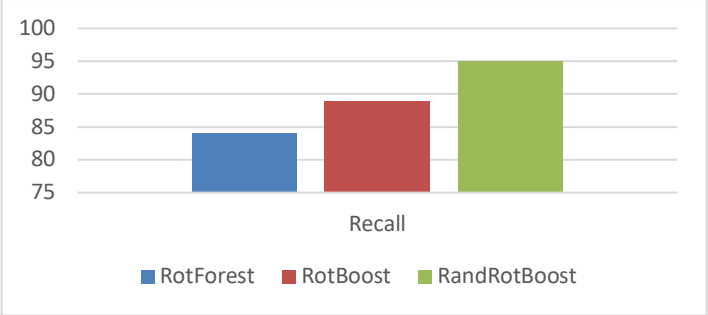

Fig. 5. Recall rate achieved by various classifiers

From Figure 5, it is observed that Rotforest achieved 84% recall, Rotboost achieved 89% recall and the proposed RandRotBoost achieved 95% recall which is better than the other two methods.

F1 Score

After assessing recall, accuracy, and precision, an all-around evaluating factor is the F1-Score, an indicator that balances the capacity for assessing both positive and negative capabilities. Figure 6 shows the F1 Score achieved by various classifiers.

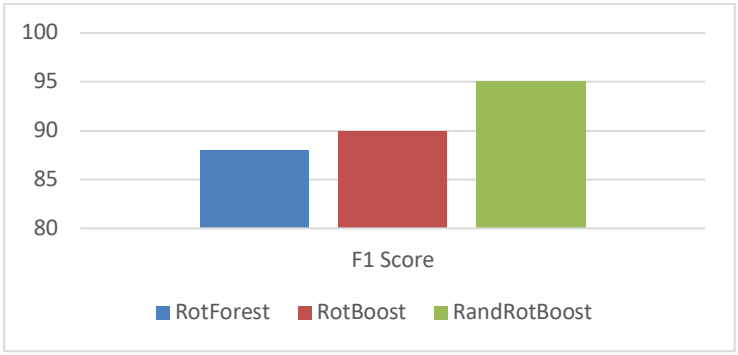

Fig. 6. F1 Score achieved by various classifiers

From Figure 6, it is observed that Rotforest achieved 88% F1 Score, Rotboost achieved 90% recall and the proposed RandRotBoost achieved 95% F1 Score which is better than the other two methods. The comparative analysis based on the size of various parameters in Table 2 and computation time in Table 3.

Table 2
Size of parameters for CP-ABE schemes

Scheme	MPK	MSK	SK	CT
[28]	–	–	$O(n)\|G\|$	$O(r_1)\|G\|$
[29]	–	–	$O(n)\|G\|$	$O(r_1)\|G\|$
[30]	$O(n)\|G\|$	$O(n)\|Z\|$	$O(r_2)\|G\|$	$O(1)\|G\|$
[31]	$O(1)\|G\|$	$O(1)\|G\|$	$O(r_2)\|G\|$	$O(1)\|G\|$
Proposed	$O(n)\|G\|$	$O(1)\|Z\|$	$O(r_2)\|G\|$	$O(1)\|G\|$

Table 2 shows that a specific parameter is not necessary. We consider each authority to be in charge of just one attribute. Table 2 shows that the suggested technique is compatible with ciphertexts of constant length.

Table 3
Computational comparison for proposed schemes

Scheme	Encryption	Decryption
[28]	$O(n)(T_{mul}+T_{Pairing})$	$O(n)(T_{exp}+T_{Pairing})$
[29]	$O(n)(T_{mul}+T_{Pairing})$	$O(n)(T_{exp}+T_{Pairing})$
[30]	$O(n)(T_{mul}+T_{Pairing})$	$O(1)(T_{exp}+T_{Pairing})$
[31]	$O(n)(T_{mul}+T_{Pairing})$	$O(1)(T_{exp}+T_{Pairing})$
Proposed	$O(1)T_{exp}+O(r_1)T_{mul}$	$O(r_1)T_{mul}+O(r_1)T_{pairing}$

Additionally, Table 3 shows that the pairing procedures continue to be constant as a result of the constant length ciphertext strategy.

Table 4
Feature based comparative analysis

Scheme	Constant Length Ciphertext	Proxy Re-Encryption
[28]	No	Yes
[29]	No	Yes
[30]	Yes	No
[31]	Yes	No
Proposed	Yes	Yes

We have provided a feature-based comparison study of the suggested design against current schemes in Table 4.

8. Conclusion

An effective method for the multicasting security feature is CP-ABE. Proxy re-encryption and ciphertext length are two crucial aspects that the fundamental CP-ABE method lacks. For each feature, the authors' research has suggested many approaches. None of the systems offered all of these features at once, though. As a result, we have developed a method in this work that offers all-encompassing properties, making it as adaptable to a variety of contexts as its predecessors. The suggested method successfully combines Rotation Forest with AdaBoost's benefits to provide a novel tree-based technique called Random RotBoost. A less reliable classification method is not only unable to achieve acceptable classification performance when dealing with sparse features in medical data

sets, but it can also quickly lead to the issue of data distortion while processing medical data. Practical evidence from this work demonstrates that the suggested method, when compared to other novel ensemble methods, may achieve a reliable classification performance in simulation trials. Since RandRotBoost does not optimise the input parameters together, it is less likely to overfit. RandRotBoost may be used to increase the precision of shaky classifiers. These days, instead of using it for binary classification issues, RandRotBoost is employed for text and picture classification.

Although clinical decision support technologies have the potential to improve patient care by providing clinicians with access to relevant data, difficulties with alarm fatigue and adherence can exacerbate existing concerns in healthcare systems, such as intrusive alarms, weary physicians, and missed diagnoses. The increasing prevalence and danger associated with AI increases the possibility of diagnostic mistakes and data omissions.

Patient care and the treatment of chronic diseases are being transformed in four ways thanks to clinical decision support systems (CDSSs): diagnostic imaging, therapeutic routes, and individualised healthcare. Data-driven administration of patients, cognitive healthcare systems, monitoring patients remotely, sensor-based systems, and value-based healthcare are just a few examples of emerging fields where CDSSs might be used. These systems rely on highly advanced technology, but their effective implementation also depends on a number of human and organisational variables. As a future work, any one of these areas can be focussed.

References

[1] Sutton, Reed T., David Pincock, Daniel C. Baumgart, Daniel C. Sadowski, Richard N. Fedorak, and Karen I. Kroeker. "An overview of clinical decision support systems: benefits, risks, and strategies for success." *NPJ digital medicine* 3, no. 1 (2020): 17. https://doi.org/10.1038/s41746-020-0221-y

[2] Osheroff, Jerome A., Jonathan M. Teich, Donald Levick, Luis Saldana, Ferdinand Velasco, Dean F. Sittig, Kendall M. Rogers, and Robert A. Jenders. *Improving outcomes with clinical decision support: an implementer's guide*. CRC Press, 2012.

[3] Middleton, B., D. F. Sittig, and A. Wright. "Clinical decision support: a 25 year retrospective and a 25 year vision." *Yearbook of medical informatics* 25, no. S 01 (2016): S103-S116. https://doi.org/10.15265/IYS-2016-s034

[4] Dias, Duarte, and João Paulo Silva Cunha. "Wearable health devices—vital sign monitoring, systems and technologies." *Sensors* 18, no. 8 (2018): 2414. https://doi.org/10.3390/s18082414

[5] El-Sappagh, Shaker H., and Samir El-Masri. "A distributed clinical decision support system architecture." *Journal of King Saud University-Computer and Information Sciences* 26, no. 1 (2014): 69-78. https://doi.org/10.1016/j.jksuci.2013.03.005

[6] Karthikeyan, G., G. Komarasamy, and S. Daniel Madan Raja. "An Efficient Method for Heart Disease Prediction Using Hybrid Classifier Model in Machine Learning." *Annals of the Romanian Society for Cell Biology* (2021): 5708-5717.

[7] Abouali, Meryem, Kartikeya Sharma, Oluwaseyi Ajayi, and Tarek Saadawi. "Performance Evaluation of Secured Blockchain-Based Patient Health Records Sharing Framework." In *2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, pp. 1-7. IEEE, 2022. https://doi.org/10.1109/IEMTRONICS55184.2022.9795759

[8] Semantha, Farida Habib, Sami Azam, Bharanidharan Shanmugam, and Kheng Cher Yeo. "PbDinEHR: A Novel Privacy by Design Developed Framework Using Distributed Data Storage and Sharing for Secure and Scalable Electronic Health Records Management." *Journal of Sensor and Actuator Networks* 12, no. 2 (2023): 36. https://doi.org/10.3390/jsan12020036

[9] Lee, Sejong, Jaehyeon Kim, Yongseok Kwon, Teasung Kim, and Sunghyun Cho. "Privacy Preservation in Patient Information Exchange Systems Based on Blockchain: System Design Study." *Journal of medical Internet research* 24, no. 3 (2022): e29108. https://doi.org/10.2196/29108

[10] Almaghrabi, Nada Saddig, and Bussma Ahmed Bugis. "Patient confidentiality of electronic health records: A recent review of the Saudi literature." *Dr. Sulaiman Al Habib Medical Journal* 4, no. 3 (2022): 126-135. https://doi.org/10.1007/s44229-022-00016-9

[11] Khan, Shahzeb, and Jawwad Ahmed Shamsi. "Health Quest: A generalized clinical decision support system with multi-label classification." *Journal of King Saud University-Computer and Information Sciences* 33, no. 1 (2021): 45-53. https://doi.org/10.1016/j.jksuci.2018.11.003

[12] Alabdulkarim, Alia, Mznah Al-Rodhaan, Tinghuai Ma, and Yuan Tian. "PPSDT: A novel privacy-preserving single decision tree algorithm for clinical decision-support systems using IoT devices." *Sensors* 19, no. 1 (2019): 142. https://doi.org/10.3390/s19010142
[13] Sreejith, S., H. Khanna Nehemiah, and A. Kannan. "A clinical decision support system for polycystic ovarian syndrome using red deer algorithm and random forest classifier." *Healthcare Analytics* 2 (2022): 100102. https://doi.org/10.1016/j.health.2022.100102
[14] Alabdulkarim, Alia, Mznah Al-Rodhaan, Yuan Tian, and Abdullah Al-Dhelaan. "A Privacy-Preserving Algorithm for Clinical Decision-Support Systems Using Random Forest." *Computers, Materials & Continua* 58, no. 3 (2019). https://doi.org/10.32604/cmc.2019.05637
[15] Alsuhibany, Suliman A., Sayed Abdel-Khalek, Ali Algarni, Aisha Fayomi, Deepak Gupta, Vinay Kumar, and Romany F. Mansour. "Ensemble of deep learning based clinical decision support system for chronic kidney disease diagnosis in medical internet of things environment." *Computational Intelligence and Neuroscience* 2021 (2021). https://doi.org/10.1155/2021/4931450
[16] Bashir, Saba, Abdulwahab Ali Almazroi, Sufyan Ashfaq, Abdulaleem Ali Almazroi, and Farhan Hassan Khan. "A knowledge-based clinical decision support system utilizing an intelligent ensemble voting scheme for improved cardiovascular disease prediction." *IEEE Access* 9 (2021): 130805-130822. https://doi.org/10.1109/ACCESS.2021.3110604
[17] Lakshmanaprabu, S. K., Sachi Nandan Mohanty, Sujatha Krishnamoorthy, J. Uthayakumar, and K. Shankar. "Online clinical decision support system using optimal deep neural networks." *Applied Soft Computing* 81 (2019): 105487. https://doi.org/10.1016/j.asoc.2019.105487
[18] Casal-Guisande, Manuel, Alberto Comesaña-Campos, Inês Dutra, Jorge Cerqueiro-Pequeño, and José-Benito Bouza-Rodríguez. "Design and development of an intelligent clinical decision support system applied to the evaluation of breast cancer risk." *Journal of personalized medicine* 12, no. 2 (2022): 169. https://doi.org/10.3390/jpm12020169
[19] Kovalchuk, Sergey V., Georgy D. Kopanitsa, Ilia V. Derevitskii, Georgy A. Matveev, and Daria A. Savitskaya. "Three-stage intelligent support of clinical decision making for higher trust, validity, and explainability." *Journal of Biomedical Informatics* 127 (2022): 104013. https://doi.org/10.1016/j.jbi.2022.104013
[20] Moshkov, Mikhail. "On the depth of decision trees with hypotheses." *Entropy* 24, no. 1 (2022): 116. https://doi.org/10.3390/e24010116
[21] Opitz, David, and Richard Maclin. "Popular ensemble methods: An empirical study." *Journal of artificial intelligence research* 11 (1999): 169-198. https://doi.org/10.1613/jair.614
[22] Breiman, Leo. "Random forests." *Machine learning* 45 (2001): 5-32. https://doi.org/10.1023/A:1010933404324
[23] Freund, Yoav, and Robert E. Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting." *Journal of computer and system sciences* 55, no. 1 (1997): 119-139. https://doi.org/10.1006/jcss.1997.1504
[24 Rodriguez, Juan José, Ludmila I. Kuncheva, and Carlos J. Alonso. "Rotation forest: A new classifier ensemble method." *IEEE transactions on pattern analysis and machine intelligence* 28, no. 10 (2006): 1619-1630. https://doi.org/10.1109/TPAMI.2006.211
[25] Caruana, Rich, and Alexandru Niculescu-Mizil. "An empirical comparison of supervised learning algorithms." In *Proceedings of the 23rd international conference on Machine learning*, pp. 161-168. 2006. https://doi.org/10.1145/1143844.1143865
[26] Zhang, Chun-Xia, and Jiang-She Zhang. "RotBoost: A technique for combining Rotation Forest and AdaBoost." *Pattern recognition letters* 29, no. 10 (2008): 1524-1536. https://doi.org/10.1016/j.patrec.2008.03.006
[27] Acute inflammations data set, UCI machine learning repository. (2009). [Online]. Available: https://archive.ics.uci.edu/ml/datasets/ Acute+Inflammations/
[28] Xu, Zhenwu, Jinan Shen, Fang Liang, and Yingjie Chen. "Fine-grained access control scheme based on improved proxy re-encryption in cloud." *Journal of Advanced Computational Intelligence and Intelligent Informatics* 25, no. 2 (2021): 170-176. https://doi.org/10.20965/jaciii.2021.p0170
[29] Pareek, Gaurav, and B. R. Purushothama. "KAPRE: Key-aggregate proxy re-encryption for secure and flexible data sharing in cloud storage." *Journal of Information Security and Applications* 63 (2021): 103009. https://doi.org/10.1016/j.jisa.2021.103009
[30] Kothari, Rakshit, Naveen Choudhary, and Kalpana Jain. "CP-ABE scheme with decryption keys of constant size using ECC with expressive threshold access structure." In *Emerging Trends in Data Driven Computing and Communications: Proceedings of DDCIoT 2021*, pp. 15-36. Springer Singapore, 2021. https://doi.org/10.1007/978-981-16-3915-9_2
[31] Zhang, Zhishuo, Wei Zhang, and Zhiguang Qin. "Fully constant-size CP-ABE with privacy-preserving outsourced decryption for lightweight devices in cloud-assisted IoT." *Security and Communication Networks* 2021 (2021): 1-16. https://doi.org/10.1155/2021/6676862

[32] Gonzalez Sepulveda, Juan Marcos, F. Reed Johnson, Shelby D. Reed, Charles Muiruri, Carolyn A. Hutyra, and Richard C. Mather III. "Patient-preference diagnostics: adapting stated-preference methods to inform effective shared decision making." *Medical Decision Making* 43, no. 2 (2023): 214-226. https://doi.org/10.1177/0272989X221115058

[33] Kaskovich, Samuel, Kirk D. Wyatt, Tomasz Oliwa, Luca Graglia, Brian Furner, Jooho Lee, Anoop Mayampurath, and Samuel L. Volchenboum. "Automated Matching of Patients to Clinical Trials: A Patient-Centric Natural Language Processing Approach for Pediatric Leukemia." *JCO Clinical Cancer Informatics* 7 (2023): e2300009. https://doi.org/10.1200/CCI.23.00009

[34] Subashini, P., T. T. Dhivyaprabha, M. Krishnaveni, and M. B. Jennyfer Susan. "Smart Intelligent System for Cervix Cancer Image Classification Using Google Cloud Platform." In *Enabling Technologies for Effective Planning and Management in Sustainable Smart Cities*, pp. 245-281. Cham: Springer International Publishing, 2023. https://doi.org/10.1007/978-3-031-22922-0_10