How to Cite:
Sangeetha, M., Devi, R. M., Sagana, C., Suruthi, B. S. S., Thejesvika, S. S., & Vaibhav, R. (2022). Privacy preserving disease risk assessment model using machine learning classifiers. *International Journal of Health Sciences*, *6*(S3), 9712–9727. https://doi.org/10.53730/ijhs.v6nS3.8549

Privacy preserving disease risk assessment model using machine learning classifiers

Ms. M. Sangeetha
AP(SRG)/Dept of CSE, Kongu Engineering College, Erode, India

Dr. R. Manjula Devi
ASP/Dept of CSE, Kongu Engineering College, Erode, India

Ms. C. Sagana
AP(SRG)/Dept of CSE, Kongu Engineering College, Erode, India

B. S. Swarna Suruthi
Dept of CSE, Kongu Engineering College, Erode, India

S. S. Thejesvika
Kongu Engineering College, Erode, India

R. Vaibhav
Kongu Engineering College, Erode, India

***Abstract*---**In the pre-digital era, medical diagnosis used to consume a lot of time and human resources. But in this digital age, the entire process can be efficiently done with the help of machines. So a new term called prognosis is introduced in this modern era where scientific prediction of the likely development of a disease and its outcome can be done. Machine learning techniques are utilized for prognosis. Though machine learning algorithms solve many problems in the healthcare field, they cannot flourish further without privacy and security assurances as the healthcare field consists of more sensitive data. Our application addresses this issue and provides a fully secured disease risk assessment using an encryption algorithm called RSA.The main objective here is to build a system with the following aspects, Functional Aspect: Disease prediction using machine learning technique decision trees. Non-Functional Aspect: Providing privacy and confidentiality of data using encryption techniques like RSA. Data privacy and confidentiality is provided with the help of the RSA algorithm. RSA is the first successful public key cryptographic algorithm.

***Keywords*---**random forest, SMOTE, RSA, encryption, decryption, prognosis, risk assessment.

Introduction

For years, machine learning has aided the health-care industry in a variety of ways. Machine learning is being utilized in an assortment of medical services settings, from case the executives of normal constant sicknesses to utilizing patient health data. Technology becomes more valuable as our understanding of precautions and diseases grows. The method of predicting diseases from a list of attributes, for example, has played a significant role in the accurate prediction of diseases such as kidney diseases, heart attacks, chronic disease risk, cancers, and so on. Machine Learning, in particular, will aid in the development of classifiers capable of accurately foretell the position of a patient's illness in the in future will aid in the development of classifiers capable of accurately foretell the position of a patient's disease in the near future.

Disease prediction is a method of identifying patient health by using data mining and machine learning techniques on patient treatment history. Disease risk assessment can be characterized as a coordinated assessment and acknowledgment of hazard factors that are liable for a sickness, the assessment of hazard levels, and the disclosure of possible ways of forestalling the development and spread of a sickness in the populace. Online disease risk assessment is one of the most popular e-healthcare applications, because it can sense a risky situation before it turns into a disease or problem, and also the cost of interpose is significantly lower than the ultimate cost of operation. By and large, illness risk appraisal has two sections: model preparation and infection risk expectation. A web-based medical care supplier can give a web-based illness risk expectation administration to a client with a forecast model prepared during the infection risk expectation period, which will essentially work on the viability of clinical therapy and individuals' personal satisfaction.

For the purpose of disease prediction, medical data are collected from several medical centers. All these data are used for training the model and disease prediction can be done. The system is divided into two major phases: model building and disease prediction. Model building involves pre-processing the distributed datasets in a way suitable to use. Being an age-old problem, there have been multiple approaches to provide an optimal solution to disease risk assessment (model prediction). The most common and adapted approach involves the use of a machine learning model like decision trees while also maintaining data privacy and confidentiality using encryption techniques. Security is an important aspect in medical field. Equal amount of importance should be given to both disease prediction as well as maintaining the privacy of the collected data. In information systems, privacy and security are major concerns. The modernization of healthcare services provides access to the patient information from any internet connected web browser, anywhere in the world. One of the many advantages of having medical information available everywhere is that it raises patients' awareness of the diseases from which they suffer and improves medical treatment.

Users typically have no thought how their personal data is handled. The availability of healthcare data in the cloud has piqued the interest of cybercriminals, who target systems in order to obtain confidential information and profit financially. Each year, a large number of data breaches occur. There are different encryption algorithms to secure the privacy and confidentiality of data. Encryption is the method of modifying the form of a message so that it cannot be read by anyone. Encryption is of two types, symmetric and asymmetric. Symmetric- key encryption encrypts the data with a key and decrypts it with the same key, making it simple to use but on the other hand it's less secure. It also necessitates a secure method of transferring the key from one party to another. The two major problems in symmetric encryption are the lack of confidentiality of the message and the lack of authentication.

Asymmetric data encryption is a technique that uses both public and private keys to encrypt data. It encrypts and decrypts the message using two different keys. It's more secure than symmetric key encryption. As it has two different keys the key distribution problem is eliminated. And also there is no need for sharing private keys. The concept of digital signature can also be given up. Different asymmetric encryption algorithms are Diffie-Hellman, ECC, El Gamal, DSA and RSA algorithms. Data privacy and confidentiality is provided with the help of the RSA algorithm. RSA is the first successful public key cryptographic algorithm. The RSA algorithm has been used to encrypt the data from both the user's point of view and the medical center's point of view. RSA is an asymmetric key encryption technique and one of the most widely used public-key algorithms. It depends on the modular exponentiation of integers. Private keys are generated that ensure effective data security. The main objective here is to build a system with the following aspects, Functional Aspect: Disease prediction using machine learning technique decision trees. Non-Functional Aspect: Providing privacy and confidentiality of data using encryption techniques like RSA. The system consists of three types of users, medical centers/Hospitals, E-Healthcare systems, and Customers.

Related Work

Dindayal Mahto et al... RSA and ECC: A Comparative Analysis - In the ongoing scene situation and public-key cryptography fragment, most of the organizations are satisfied by RSA based cryptosystems. These days we live in a computerized existence where a greater part of our messages or data gets traded between conveying clients or frameworks. Yet, as the web is an open-finished engineering there are a few blemishes through which busybodies perform digital assaults on conveyed messages. In the composition, a part of the writers have presented the comparative execution and security assessment of RSA and ECC with different limits of assessments. Gura et al have contemplated the point increment action of an elliptic curve among RSA and ECC on two 8-cycle processor PC systems and have seen that ECC-160-point duplication is more capable over RSA-1024 private key movement. Bos et al have assessed the bet of use of an essential considering the basic length of RSA and ECC, and they assume that till 2014 that the usage of 1024-cycle RSA gives some little bet while 160-piece ECC over a great field may safely be used for a fundamentally more extended period. Kute et al have contemplated that RSA is speedier than ECC, but security-wise ECC beats RSA.

Alese et al suggested that at this point, RSA is more grounded than ECC notwithstanding the way that they in like manner displayed ECC beats RSA in near future. Mahto et al showed that ECC beats functional productivity and security over RSA.

Nirjharini Mohanty et al... in their paper dealt with binary classification problems. The outcome of the problem mentioned in this paper is to identify whether the person is affected by diabetes or not. The authors have given two solutions using two different machine learning algorithms namely Gradient Boosting Classifier(GrB) and Extra Tree Classifier(ExT). They have compared the performances of the above-mentioned algorithms and have concluded that GrB is best suited for real-time apps which are based on this problem because GrB outperforms ExT in the following measures namely Average precision, Precision and Recall.

S. NO	Title of the paper	Purpose	Algorithm	Advantages	Disadvantages
1	Abhishek Guru et al ... in [15]	To increase the level of security and reliability by modifying the RSA encryption algorithm	RSA	• RSA is an asymmetric key encryption method and perhaps the most generally utilized public-key calculation. • It relies upon the secluded exponentiation of numbers. • Confidential keys are created that guarantee powerful information security.	• The calculation part of modified RSA encryption algorithm is complex
2	Xue Yang et al ... in [24]	To provide disease risk prediction while ensuring privacy preservation.	Bloom Filter technique, Okamoto–Uchiyama Cryptosystem, homomorphic cryptographic algorithm, naïve Bayesian classifier.	• For model preparation, clinical information is gathered from infection affirmed patients and are utilized to prepare the credulous Bayesian classifier. • Then, at that point, the prepared classifier is utilized to extricate the side effect vector set of every infection. • In view of the separated preparation results, the forecast of end-clients should be possible.	• There is a need for access control and authorization as the medical users are unfamiliar with the diagnosed disease.

3	Evelyn P Whitlock et al ... in [31]	To survey proof related to information gaps recognized by the 2002 recommendation and to think about local area execution of screening endoscopy, includi ngmistreat.	Newer FOBTs, Fecal DNA Test, CT Colonography	• To expand the take-up of and benefits from suggested colorectal disease screeni ng, scientists have tried to work on the exactness, agreeableness, or openness of screening by presenting new tests or improving existing tests.	• The precision and damages of screening tests were checked after just a solitary application.
4	J. Qiu et al ... in [1]	To enable medical care to counter the security challenges in the shrewd urban communities.	Artificial Intelligence, Machine Learning, Sensor Network	• Blockchain innovation has worked for free from even a hint of harm stockpiling of the patient's data in the medical care framework.	• Scalability is an issue in this.
5	A. Abbas et al ... in [3]	A cloud-based system that efficiently manag es health-associated Big-data while taking advantage of the Internet and ubiquity ofsocial media.	Softwareas a service	• Trial results show that the schema proposed has accomplished high accuracy when contrasted with the state-of-the-art approaches as far as illness risk evaluation and master client suggestion.	• The framework does not facilitate the mobile users.
6	S. Perveen et al ... in [4]	The goal of this examination is a Machine Learning based strategy to recognize people at an expanded gamble of creating NAFLD utilizing risk elements of ATP III clinical rules refreshed in 2005.	Decision Tre e algorithm, CPCSSN dataset	• It consolidates different indicators in a basic bit by bit way, whose syntax are naturally understood and simple to explain for specialists, as they can view the design of choices in the ordering system	• Limitation is expected in TRG rules as the reference a n incentive for deciding NAFLD chance and output speculation while managing other populace.
7	L. Jena et al ... in [5]	The objective of this work is mainly to predict the risk in	Biomedical dataset	• Focuses on the implementationof classification algorithms	• No feature reduction

		chronic diseases using machine learning strategies such as feature selection and classification.		in medical data and bioinformatics.	using genetic searc h algorithm.
8	R. Bocu et al ... in [9]	An integrated personal health information system that permits secure capacity and handling of clinical information in the cloud by utilizing a comprehensive homomorphic encryption model to protect information.	Comprehensive homomorphic encryption model	• The framework gathers the client information through a client application module • It is introduced on the client's cell phone or smartwatch, and safely ships the information to the cloud backend fueled by IBM Bluemix.	• The occasion based controllers are not set off by the IBM OpenWhisk programming administration.
9	J. Hua et al ... in [11]	It is based on Naive Bayes Classification, and has suggested PDiag, a proficient an d privacy-preserving healthcare essenti al diagnosis plan.	Naïve baye s Classification	• Execution assessments by means of carrying out PDiag on cell phones and PCs exhibit PDiag's adequacy concerning genuine environment.	• PDiag guarantees clients' wellbeing data and specialist organization's forecast model are retain secret, and has altogether lower calculation and correspondence upward tha n on going plans.

Existing Work

The two main phases of disease risk assessment are model training and disease risk estimation. In the model training stage, the medical data are gathered from multiple sources and aggregated. Then the model gets trained based on the machine learning algorithms to become an efficient disease risk estimation model. To enhance the standards of people's lives and to provide efficient treatment to the people, disease risk prediction services are provided to the people by the e-healthcare provider. The collected medical data will contain sensitive information about the people which should be kept highly confidential. Also the disease risk inquiry requests and outputs are highly confidential as they contain personal data about the particular user.

CARER is an effectual and privacy-preserving online disease risk assessment system over multi- outsourced vertical datasets. This schema provides a disease risk prediction service which is very well secured. The model is trained over vertically distributed dataset and it can also be dynamically updated. The model user for prediction can also be updated by using a strategy called model updating strategy. The privacy is preserved in both the training stage and the disease risk estimation phase. To train the model securely a modified Paillier cryptosystem has been proposed so that the sensitive data will not be disclosed to any others. The disease risk prediction phase and the query requests and results are protected using a random masking technique. By using these techniques, the data has been secured in an efficient

Data preprocessing is done for reducing the communication overhead. Naive Bayesian classification is used to provide high efficiency in disease risk assessment. The four parts of the system are the trusted authority, users, medical data centers, e-healthcare providers. Preprocessing and securing the data are done by the medical centers. Disease risk assessment is provided by the online healthcare organizations like e- healthcare centers. Firstly they aggregate the collected data then encrypt the data then offer the disease risk prediction service which is privacy preserved. Users are the patients and the doctors. The building block of CARER is the Paillier cryptosystem. It is a most commonly used public key cryptography technique. A security parameter, two big prime numbers are chosen. The product of the two prime numbers is calculated and a random number which satisfies the gcd condition is chosen. Finally, the public key and the parallel hidden key is obtained.

The CARER schema has four stages: framework initialization, medical data preprocessing and encryption, secure data aggregation and training, and disease risk prediction. The medical center performs data preprocessing and encryption on collected data so that it can be used to tutor the model for disease risk estimation. The user-supplied symptom vector is encrypted, and disease risk inquiry requests and disease risk prediction output are generated using the trained model. These two tasks are done by the online health care benefactor. The variable N of the public key is divided into m parts for producing secret keys during the encryption phase of a modified Paillier cryptosystem. Even if the user has access to the secret key, this cannot be decrypted. The consolidated data is only available to the online health care benefactor. During the disease risk estimation stage, users can calculate disease risk query responses using a protected two-party inward item calculation convention.

Even if an attacker can listen in on EP and user communications and obtain all packets sent and received between the two, it will be unable to obtain any useful information because query requests and responses contain massive random numbers. The security of clients' side effect vectors and inquiry output is hence guaranteed. CARER accomplishes coordinated security assurance in both the model preparation and sickness risk forecast stages. As far as calculation cost and correspondence upward, CARER is also effective. Other types of attacks on e-healthcare systems include denial of service (DoS) and poisoning attacks. The system will be protected from these attacks in the future.

Proposed Framework

Architecture

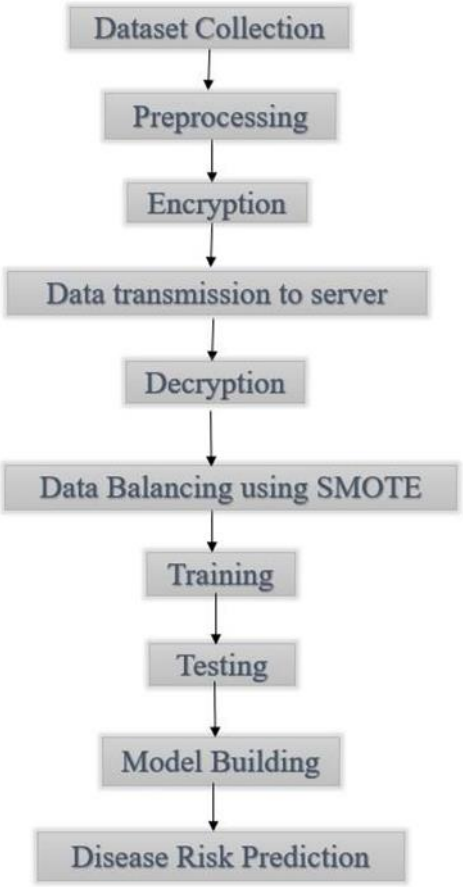

Fig 3.1 Architecture flow

Model Building

A model is built using the machine learning algorithm called Random Forest classifier. Random forest classifier algorithm is chosen because Decision trees can work well even with less features and they take into attention only the significant features for classification. It also handles higher dimensionality data very well. The dataset collected from the user is first checked for Data imbalance. It is handled through the over-sampling technique called SMOT which stands for Synthetic Minority Over-sampling Technique. Now the dataset becomes balanced. Finally, the model is trained and tested.

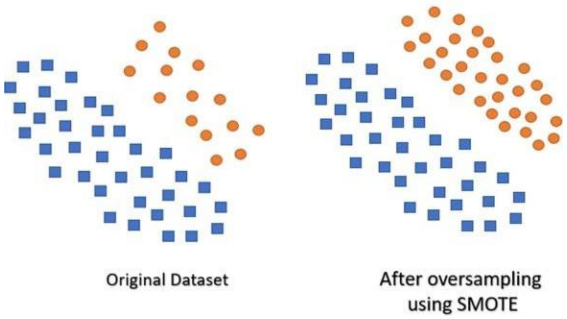

Fig 3.2 SMOTE

Random Forest Classifier

Random Forest algorithm is an ensemble algorithm. Ensemble methods are used to increase the probability in the models by using various models to achieve a dependable model. Bagging, boosting, and stacking are the commonly used methods. An ensemble is a cluster of elements that are considered as a whole not individually. These techniques help to increase the wellness / simplification of the model. It is a supervised learning algorithm. It builds a forest of decision trees using a bagging method. This combines many decision trees into a single tree to get more accurate predictions. The steps followed in a Random Forest Algorithm:

- Random samples are selected from the dataset
- Decision trees are built for all the samples
- Prediction results of every decision trees are obtained
- Votes are calculated for all the predicted results
- The prediction result which has the majority of the votes is taken as the final output

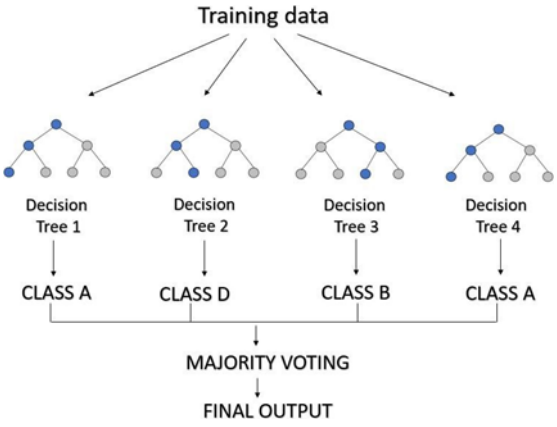

Fig 3.3 Random Forest Classifier

Disease Prediction

Input symptoms are received from the user and passed as input for the model. Sample input symptoms include anxiety, lethargy, acidity, cough, sweating, dehydration, indigestion, headache, nausea, diarrhea, phlegm, congestion, cramps, obesity, blister etc. These input symptoms are then passed to the trained model for predicting the probability of the disease (prognosis)

User Interface

There are 3 UI pages. One is for uploading the dataset in the form of CSV files. Then the uploaded dataset is used for model training. And the second is for getting the input symptoms from the user. Here a Drop- down list is provided with all the input symptoms of the dataset and the user chooses the symptoms that he is experiencing. The symptoms given by the user are used for the disease prediction(prognosis). The third page displays the prognosis results in the form of a Bar chart with the disease having the highest probability in the first place (Descending order of probability).

Encryption

The algorithm used for encryption is RSA which stands for Rivest-Shamir-Adleman.

- The CSV file is transformed into a string (of text)
- RSA algorithm works with numbers, so each of the characters in the string is converted to the corresponding ASCII value.
- ASCII values are encrypted using the RSA algorithm formula,

C = me mod N
where,
m - message in terms of ASCII, e - encryption key,
N - a product of two large prime numbers,
C - ciphertext
The ciphertext 'C' is then transmitted to the E-healthcare provider.

Decryption

Process of decryption using RSA algorithm. The received ciphertext at E-healthcare provider is decrypted using the RSA algorithm formula,
M = cd mod N
where,
m - message in ASCII format, c - ciphertext,
d - decryption key,
N - a product of two large prime numbers
1. The decrypted message (m) will still be in ASCII
form. So, it converted back to normal text (ASCII to character conversion)

Now, the string of text is transformed again to CSV drop-down format that the ML model can use for training.

System Implementation
Dataset Collection

Dataset Name: Disease Prediction UsingMachine Learning

Use Machine Learning and Deep Learning models to classify 42 diseases. Complete Dataset consists of 2 CSV files. One of them is training and the other is for testing the model. Each CSV file has 133 columns. 132 of these columns are symptoms that a person experiences and the last column is the prognosis. These symptoms are mapped to 42 diseases. The diseases that can be predicted are fungal infection, allergy, GERD, Chronic cholestasis, Drug reaction, Peptic ulcer disease,AIDS, Diabetes, Gastroenteritis, Bronchial Asthma, Hypertension, Migraine, Cervical spondylosis, Paralysis, Jaundice, Malaria, Chicken pox, Dengue, Typhoid, Hepatitis - A,B,C,D,E,Alcoholic hepatitis, Tuberculosis, Common cold, Pneumonia, Dimorphic Hemorrhoids, Heart attack, Varicose veins, Hypothyroidism, Hyperthyroidism, Hypoglycemia, Osteoarthritis, Arthritis, Paroxysmal positional vertigo, Acne, Urinary tract infection, Psoriasis, Impetigo.

Data Balancing Using SMOTE

In [a, b], a refers to the original no of availablesamples and b refers to the total no of samples belonging to that class after sampling. The no of minority class samples are increased to match the no of majority class samples thereby solving the data imbalancing problems using an oversampling technique called SMOTE.

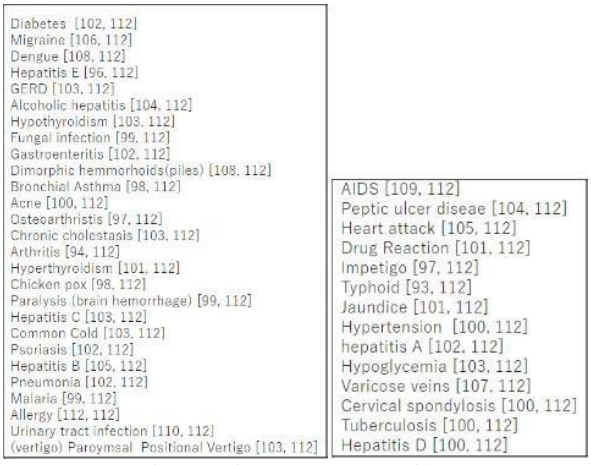

Fig 4.1 Before, After Oversampling(SMOTE)

Once the dataset is uploaded, preprocessing of the dataset is done to make it ready for encryption. Then the encryption phase occurs and then the data is transmitted to the server. Final phase is the decryption phase. After decryption, the model will be trained, tested and made available for use. The input symptoms are received from the user. The user can select up to a maximum of 6 input symptoms. After selecting the input symptoms disease risk assessment is done and the prognosis results are
displayed in the form of a bar chart.

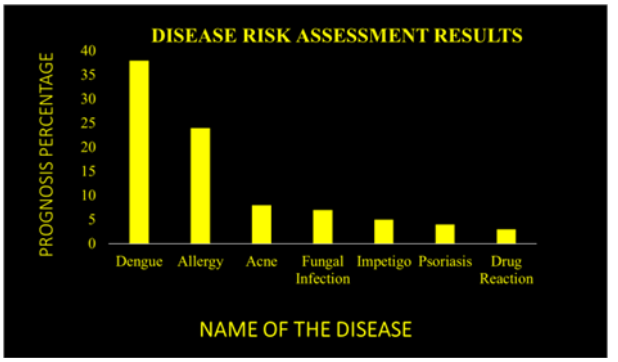

Fig 4.2 Disease Risk Assessment Results

Fig 4.5 shows the prognosis results in the form of a bar chart for the given set of input symptoms by the user. Higher the length of the bar chart means that the probability of the specified disease is higher.

Performance Metrics

Model approval is alluded to as the interaction where a prepared model is assessed with a testing informational index. The testing informational index is a different part of similar informational collection from which the preparation set is derived. The reason for model approval is to checked the accuracy and execution of the model in light of the past information for which we as of now have actuals. Metrics considered for evaluation:

- Confusion Matrix
- Accuracy
- Precision
- Recall

Confusion Matrix

Confusion matrix is used for describing the performance of the model. It's done by comparing the values of the actual and the predicted classes.

- TP: True Positive - Predicted value (positive) of the model is same as that of the actual value
- FP: False Positive - Predicted value is incorrect. Negative values are estimated by the model as positive values
- FN: False Negative - Predicted value is incorrect. Positive values are estimated by the model as negative values
- TN: True Negative - Predicted value (negative) of the model is same as that of the actual value

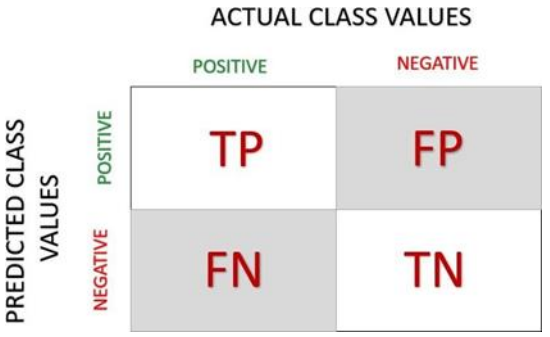

Fig 4.3 Confusion Matrix

Accuracy Score

Accuracy is a measurement for assessing models. Accuracy is the representation of the number ofright predictions per complete number of estimations.

Accuracy = (TP + TN) / (TP + FP + TN + FN)

```
Accuracy score 1.0
[[16  0  0 ...  0  0  0]
 [ 0 18  0 ...  0  0  0]
 [ 0  0 17 ...  0  0  0]
 ...
 [ 0  0  0 ... 16  0  0]
 [ 0  0  0 ...  0 18  0]
 [ 0  0  0 ...  0  0 20]]
```
Fig 4.4 Accuracy Score and Confusion matrix

From the confusion matrix we can calculate the accuracy score, precision, recall as well as F1 score. From the above confusion matrix, we can conclude that the accuracy of the model is 100% which is 1.

Precision

Precision is the ratio of true positives to total estimated positives. It is the proportion of predicted positives correctly classified. When a model classifies an observation as positive, it measures how precise it is. It also indicates how effectively our model reduces type I error (False positive).

Precision=TP/(FP+TP)

Recall

The recall ratio is the number of true positives partitioned by the absolute number of actual positives. It represents the proportion of true positives that were correctly classified. It assesses a model's ability to recall the actual positive classes. It also indicates how effectively our model reduces type II error (False negatives). It is also known as sensitivity because it assesses a model's sensitivity to the positive class.

Recall=TP/(FP+TP)

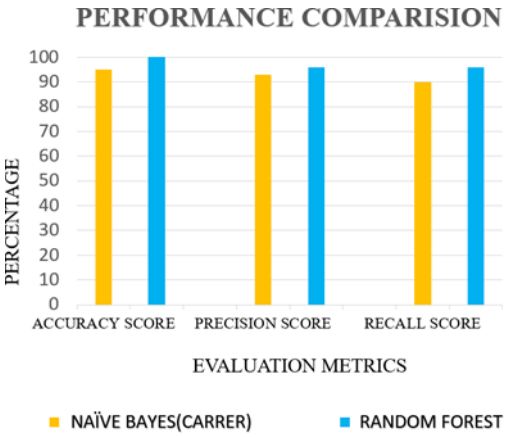

Fig 4.5 Performance Comparison ofexisting and proposed System

From the figure we can conclude that the accuracy of the carer schema is around 95% and the accuracy of the proposed system (Random Forest classifier using RSA) is nearly 100%. Similarly, the Precision value of the CARER schema is around 92% and the precision value of the proposed system is around 95%. Also the recall score of the proposed system is high when compared to the CARER schema.

Conclusion and Future Work

The goal of this research is to build a privacy preserving disease risk assessment system for the users. We also provide an insight on how the system can be extended for future work along with some of the useful works referred while building the system. The accuracy of these predictions is tested against the test data and any necessary scope for improving the resultant accuracy are handled. These predictions could prove to be helpful to elevate the overall efficiency and assist medical prognosis. The system can predict 42 diseases at present based on the symptoms. In future, the system can be extended by predicting the disease using many more Machine learning algorithms and the prediction results can be compared. Also, the system can be made to provide a detailed report of the patient, the medicines that the user may need for the disease, and also the nearest hospital where the user can be admitted if the disease is severe. This system can also be set up inside the medical center for easy use of users.

References

1. J. Qiu, X. Liang, S. Shetty, and D. Bowden, “Towards secure and smart healthcare in smart cities using blockchain,” in IEEE International Smart Cities Conference. IEEE, 2018, pp. 1–4.
2. J. S. Lin, C. V. Evans, E. Johnson, N. Redmond, E. L. Coppola, and N. Smith, “Nontraditional risk factors in cardiovascular disease risk assessment: Updated evidence report and systematic review for the us preventive services task force,” Journal of the American Medical Association, vol. 320, no. 3, pp.

281–297, 2018.
3. A. Abbas, M. Ali, M. U. S. Khan, and S. U. Khan, “Personalized healthcare cloud services for disease risk assessment and wellness management using social media,” Pervasive and Mobile Computing, vol. 28, pp. 81–99, 2016.
4. S. Perveen, M. Shahbaz, K. Keshavjee, and A. Guergachi, “A systematic machine learning based approach for the diagnosis of non-alcoholic fatty liver disease risk and progression,” Scientific reports, vol. 8, no. 1, pp. 1–12, 2018.
5. L. Jena, S. Nayak, and R. Swain, “Chronic disease risk (cdr) prediction in biomedical data using machine learning approach,” in Advances in Intelligent Computing and Communication, 2020, pp. 232–239.
6. C. Xu, N. Wang, L. Zhu, K. Sharif, and C. Zhang, “Achieving searchable and privacy-preserving data sharing for cloud-assisted e-healthcare system,” IEEE Internet of Things Journal, vol. 6, no. 5, pp. 8345–8356, 2019.
7. W. Tang, J. Ren, K. Deng, and Y. Zhang, “Secure data aggregation of lightweight e-healthcare iot devices with fair incentives,” IEEE Internet of Things Journal, vol. 6, no. 5, pp. 8714–8726, 2019.
8. L. Yang, Q. Zheng, and X. Fan, “RSPP: A reliable, searchable and privacy-preserving e-healthcare system for cloud-assisted body area networks,” in 2017 IEEE Conference on Computer Communications. IEEE, 2017, pp. 1–9.
9. R. Bocu and C. Costache, “A homomorphic encryption- based system for securely managing personal health metrics data,” IBM Journal of Research and Development, vol. 62, no. 1, pp. 1:1–1:10, 2018.
10. M. Martinez-Arroyo and L. E. Sucar, “Learning an optimal naive bayes classifier,” in 18th International Conference on Pattern Recognition (ICPR 2006). IEEE Computer Society, 2006, pp. 1236–1239.
11. J. Hua, H. Zhu, F. Wang, X. Liu, R. Lu, H. Li, and Y. Zhang, “CINEMA: efficient and privacy-preserving online medical primary diagnosis with skyline query,” IEEE Internet of Things Journal, vol. 6, no. 2, pp. 1450–1461, 2019.
12. X. Liu, H. Zhu, R. Lu, and H. Li, “Efficient privacy- preserving online medical primary diagnosis scheme on naive bayesian classification,” Peer-to-Peer Networking and Applications, vol. 11, no. 2, pp. 334–347, 2018.
13. R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, “Toward efficient and privacy-preserving computing in big data era,” IEEE Network, vol. 28, no. 4, pp. 46–50, 2014.
14. A. Asuncion and D. Newman, “Uci machine learning repository,” 2007. [Online]. Available: https://doi.org/10.1109/MNET.2014. 6863131
15. P. Li, J. Li, Z. Huang, C. Gao, W. Chen, and K. Chen, “Privacy preserving outsourced classification in cloud computing,” Cluster Computing, vol. 21, no. 1, pp. 277–286, 2018.
16. Abhishek Guru and Asha Ambhaikar, “Development of “RSA” Encryption Algorithm for Secure Data Transmission,” in 2020 Research Journal of Computer and Information Technology Sciences, Vol. 8(1), 9-12.
17. Dindayal Mahto and Dilip Kumar Yadav, “Performance Analysis of RSA and Elliptic Curve Cryptography,” in 2018 International Journal of Network Security, Vol.20, No.4, PP.625-635.
18. Laiphrakpam Dolendro Singh and Khumanthem Manglem Singh, “Implementation of Text Encryption using Elliptic Curve Cryptography,” in 2015 Procedia Computer Science 54 PP.73-82.
19. Dindayal Mahto, Dilip Kumar Yadav, “RSA and ECC: A Comparative Analysis” in International Journal of Applied Engineering Research ISSN 0973-4562

Volume 12, Number 19 (2017) pp. 9053-9061.
20. Mohit D. Singanjude and Prof. R. Dalvi, “Literature Survey: Secure transmitting of data using RSA public key implemented with Vedic method” in 2016 International Journal of Computer Applications Technology and Research, Volume 5–Issue 10, 675- 677.
21. Neal Koblitz, “Elliptic Curve Cryptosystems”, in 1987 Mathematics of Computation Volume 4x. number 177. PP 203-209.
22. Nirjharini Mohanty, Soumen Nayak, Monarch Saha,
23. Vishal Baral and Imlee Rout, “Classification of diabetes disease using machine learning algorithms,” in 2021 ICTACT Journal On Data Science And Machine Learning, Volume: 02.
24. Padma Bh, D.Chandravathi, P.Prapoorna Roja, “Encoding And Decoding of a Message in the Implementation of Elliptic Curve Cryptography using Koblitz’s Method,” in 2010 (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 05, 1904-1907.
25. Ximeng Liu, Student Member, IEEE, Rongxing Lu, Member, IEEE, Jianfeng Ma, Le Chen, and Baodong Qin, “A Privacy-Preserving Patient-Centric Clinical Decision Support System on Naive Bayesian Classification,” in 2016 IEEE Journal of Biomedical and Health Informatics, Volume: 20, Issue: 2.
26. Xue Yang, Rongxing Lu, Senior Member, IEEE, Jun Shao, Xiaohu Tang, Member, IEEE, and Haomiao Yang, Member, IEEE, “An Efficient and Privacy-Preserving Disease Risk Prediction Scheme for E- Healthcare,” in 2019 IEEE Internet Of Things Journal, Vol. 6, No. 2, April 2019.
27. D. Zhu, H. Zhu, X. Liu, H. Li, F. Wang, H. Li, and D. Feng, “CREDO: efficient and privacy-preserving multi- level medical pre-diagnosis based on ml-kNN,” Information Sciences, vol. 514, pp. 244–262, 2020.
28. P. Paillier, “Public-key cryptosystems based on composite degree residuosity classes,” in Advances in Cryptology - EUROCRYPT ’99, J. Stern, Ed., vol. 1592. Springer, 1999, pp. 223–238.
29. M. Shen, X. Tang, L. Zhu, X. Du, and M. Guizani, “Privacy-preserving support vector machine training over blockchain-based encrypted iot data in smart cities,” IEEE Internet of Things Journal, vol. 6, no. 5, pp. 7702–7712, 2019.
30. F. Wang, H. Zhu, X. Liu, R. Lu, J. Hua, H. Li, and H. Li, “Privacy-preserving collaborative model learning scheme for e-healthcare,” IEEE Access, vol. 7, pp. 166 054–166 065, 2019.
31. H. Zhu, X. Liu, R. Lu, and H. Li, “Efficient and privacy-preserving online medical prediagnosis framework using nonlinear SVM,”IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 3, pp. 838–850, 2017.
32. Z. Ma, J. Ma, Y. Miao, and X. Liu, “Privacy-preserving and high-accurate outsourced disease predictor on random forest,” Information Science, vol. 496, pp. 225–241, 2019.
33. Evelyn P Whitlock, Jennifer S Lin, Elizabeth Liles, Tracy L Beil, Rongwei Fu, "Screening for Colorectal Cancer: A Targeted, Updated Systematic Review for the U.S. Preventive Services Task Force," in 2008 Annals of internal medicine 149 (9), PP 638-658.