# EFFICIENT AND ACCURATE CLOUD-ASSISTED MEDICAL PRE-DIAGNOSIS WITH PRIVACY PRESERVATION

<sup>1</sup>Bibi Alaida Akhil Momin, <sup>2</sup>Sujata. A. Gaikwad (Prof) <sup>1,2</sup>TPCT'S College of Engineering, Osmanabad <sup>1</sup>tasmiyamomin48@gmail.com

Abstract: This project presents NAIAD, a cloud-based medical pre-diagnosis system that uses outsourced ML models to offer diagnostic services that are safe and respect privacy. It protects patient privacy by encrypting medical records and using a new, secure way to compare them. NAIAD uses a "Secure Mahalanobis Distance-based Comparison (SMDC)" method, which is backed by matrix encryption, to keep sensitive medical information safe during KNN-based diagnosis. The system makes encrypted queries (trapdoors) that let you safely engage with the outsourced KNN-based pre-diagnostic model without giving up any of your data. Medical records are encrypted, grouped together, and put into a hierarchical index tree. This structure makes it easier and faster to find the right records throughout the diagnostic process, which improves the overall performance of the system. The project looks at encrypted datasets and concludes that KNN is more accurate than SVM. Adding the Random Forest and XGBOOST algorithms to the system also makes it more accurate, bringing the diagnostic accuracy to above 98%. In situations where data is encrypted, these enhancements greatly improve the system's performance and dependability.

**"Index Terms -** Privacy-preserving medical diagnosis, matrix encryption, trapdoor-based retrieval, cloud computing, Mahalanobis Distance, k-Nearest Neighbor (KNN), secure disease classification, homomorphic encryption, Internet of Medical Things (IoMT), and AI-driven healthcare.".

## 1. INTRODUCTION

The healthcare business is using more and more complex technology, such cloud computing, to make services better, help patients get well, and make operations run more smoothly. computing offers significant advantages, including scalability, flexibility, and cost-effectiveness, enabling healthcare organizations to handle vast datasets and intricate algorithms for medical diagnostics without depending on massive onpremises infrastructure (Hossain and Muhammad, 2016). [6]. Even while these benefits exist, using cloud-based solutions in healthcare raises big worries about data privacy and security. Revealing private patient information can have serious effects, as identity theft and breaches of confidentiality (McKeon, 2023). [9].

To reduce these risks, data protection strategies are needed to keep patient data confidential and simultaneously enable immediate medical diagnosis. This problem has been proposed to be solved using different cryptography and ML-based solutions. Indicatively, privacy-sensitive illness diagnosis systems that take advantage of secure distance evaluation models such as Mahalanobis distance have been developed to ensure that the patient data remain confidential and easy to make valid comparisons (Zhang et al., 2022) [5]. Moreover, the method of safe multiparty computation (Dong et al., 2023) [1] and homomorphic encryption (Brakerski et al., 2013) [21] are promising approaches to safe data processing in clouds.

In addition, privacy assuring ML techniques, including privacy-preserving support vector machines (Zhu et al., 2017) [11] and privacy-preserving KNN classification (Park and Lee, 2018) [12], have been used in secure online medical pre-diagnosis. New advances have explored privacy-aware keyword search methods (Wang et al., 2022) [2] and dynamically evolving skyline queries in medical settings (Zhang et al., 2022) [4], which makes retrieving data relevant and efficient at the same time be ensure privacy. Such practices ensure the safety of patients records, as well as simplified decision-making and diagnosis.

This project aims at developing a privacy preserving system of making safe comparisons of medical records on the cloud. The proposed solution is expected to enhance the best accuracy and efficiency of medical pre-diagnosis as well as improving patient privacy based on the application of the latest cryptography tools and strategies to compare similarities. This study will enhance the existing endeavors on safe cloud-tailed medical care services by resolving the underlying challenge on data utility and high levels of privacy.

# 2. RELATED WORK

Many works have explored privacy-sensitive medical diagnosis systems, applying safe encryption, cloud computing, and ML to preserve the anonymity of the patient. According to Dong et

al. [1], privacy is imperative in processing medical information, and the article presents a safe multiparty computation architecture to extensive genome-wide association studies. Another method of diagnosing diseases without compromising on privacy is PPDDS proposed by Zhang et al. (2022) [5] and is based on the secure Mahalanobis Distance model, which ensures that medical records can be safely compared.

Wang et al. [2] developed a privacy-saving query searching model that renders the search of various kinds of queries in the cloud secure and unproblematic. This will permit secure web prediagnosis. Xie et al. (2022) [3] proposed an outsourceable SVM-based diagnosis system of the IoMT as well, demonstrating the safety of the use of ML in remote medical treatment. All these studies highlight the growing need of privacy-aware computational procedures in medical data analysis.

Packed ciphertexts homomorphic encryption was introduced in Brakerski et al. (2013) [21], and has found a wide range of uses in the security of cloud-based ML. Xu et al. (2022) [17] contributed to the development of privacy studies by developing reliable top-k disease matching algorithms that are applied to E-healthcare systems to enable safe and effective matching of patients and doctors through encrypted queries.

McLachlan (1999) [19] has discussed the Mahalanobis Distance measure of classification based techniques which have widely been used in medical diagnosis models. The Mahalanobis Distance-based clustering demonstrated the ability of Mahalanobis Distance to improve the accuracy of arrhythmia classification in mobile health monitoring systems (Haldar et al., 2017) [22]. The privacy-preserving KNN algorithm used by Ehealth cloud applications by Park and Lee (2018) [12] also presents a privacy-preserving algorithm that proves that accurate diagnosis can be delivered through the use of secure computation methods.

In addition, studies on disease prediction models of the cloud have been very numerous. Ahmad, Hailul and Al-Shamiri (2020) [14] proposed CREDO, an efficient and secure multi-level medical pre-diagnosis model based on ML-KNN, that ensures efficient and secure disease classification. Kumar et al. (2018) [7] applies to healthcare systems applications, and the authors emphasize that the availability of secure AI-based diagnostics is essential.

The proposed solution is based on this work and is based on a privacy-conscious ML solution that supports safe matrix encryption, recalls of trapdoor-based access, and accurate recognition of cloud-assistance environments in a cloud assistant configuration.

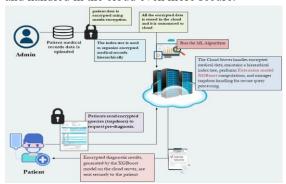
#### 3. MATERIALS AND METHODS

The proposed answer presents an modern version of secure sickness diagnosis, primarily based totally at the approach of system getting to know however that specialize in affected person privateness. It employs a matrix encryption method to make sure protection of the clinical information towards unauthorized get right of entry to and guard privateness of the records (Brakerski et al., 2013). [21]. This encryption method permits one to paintings with outsourced clinical records with a excessive stage of safety and decrease the danger of privateness violations with out compromising processing speed (Dong et al., 2023). [1].

To make sure correct pre-diagnosis, the machine combines KNN with Mahalanobis Distance, as a result improving the accuracy of predictions because of rationalization of complicated institutions of the affected person capabilities and sickness types (McLachlan, 1999). Mahalanobis Distance has been used efficaciously in privateness-retaining sickness detection, to allow similarity-primarily secure based comparisons (Zhang et al., 2022) [5] and cellular fitness monitoring (Haldar et al., 2017). [22]. KNN category of cloud assisted clinical machine is an powerful and privateness retaining method to fitness care analytics (Liu et al., 2019). [16]; (Park and Lee, 2018) [12].

The outsourced version lets in sufferers to securely have interaction with the cloud-primarily based totally platform and reap pre-diagnostic records in addition to defend their very own clinical records (Zhu et al., 2017). [11]. It is executed via secure choice tree category (Liang et al., 2021) [8] and encrypted similarity retrieval procedure (Zhu et al., 2020). [14]. The recommended solution can help to minimize the security risks associated with cloud-based medical diagnosis (Xu et al., 2022) [17] through the methods of data-processing that ensure privacy.

This means that it is more affordable to diagnose the illnesses very fast and properly, besides that, the relevant privacy concerns are covered, which creates the confidence in cloud-based solutions in healthcare. The privacy-conscious searchable encryption algorithms (Wang et al., 2022) [2] and the privacy-conservative access control systems (Barua et al., 2011) [10] make medical data stored and handled in the cloud even more secure.



"Fig.1 Proposed Architecture"

The image (Fig.1) presents a secure cloud-based medical diagnosis platform that involves the use of encrypted data and machine learning. The medical records of patients are uploaded by the admin and encrypted to the matrix and stored in the cloud. The index tree is used to organize the encrypted data. The encrypted questions (trapdoors) are sent by patients to request a diagnosis in advance with a doctor. The cloud server has XGBoost model to perform diagnosis, maintains an index tree of hierarchy, and is in charge of secure query processing. The results of the encrypted tests are sent back to the patient in a secure manner. This solution is safe, private, and at the same time allows you to perform medical diagnosis with the help of cloud-based ML.

#### i) Admin Module

Admin Module will be responsible in running the system, and it involves management of datasets, encryption and machine learning. To prevent unauthorized access of others to the system, the admin would log in using secure authentication credentials (Barua et al., 2011). [10]. Once registered, the administrator uploads a heart disease dataset that artificial intelligence algorithms will use to assist in the pre-diagnosis (Zhu et al., 2017). [11]. To ensure the privacy of the data, the system applies matrix encryption to the data and sends it to the DRIVEHQ cloud. It also possesses a trapdoor facility allowing users to safely access, as well as compute the information (Brakerski et al., 2013). [21]; Dong et al. (2023) [1]. This encryption method prevents inappropriate people who should not be accessed to do this, but still allows for privacy calculations (Wang et al., 2022). [2].

After the protected data is protected, the administrator ML algorithm is applied to the

encrypted data and observe how it is performed in the disease diagnosis. Data protection presentation classification algorithms such as KNN and Mahalanobis distance are used by the system to ensure that predictions are accurate and safe (McLachlan, 1999) [19]. (Park and Lee, 2018) [12]. It also assists computer model which enables individuals to make choices and maintain patient data confidential (Zhu et al., 2020). [14]. The administrator draws a computation graph so as to observe the system efficiency. This figure demonstrates the cost of a trapdoor building and processing of data (Xu et al., 2022) [17]. This step assists us in determining the functionality of the system and enhance purveying procedures of healthcare applications in clouds that protect privacy.

Admin Module ensures that data security, algorithm implementation and performance analysis are all performed amicably. This assists in developing a robust and privacy-based cloud-based healthcare model.

## ii) User Module

The User Module allows the patients to securely utilize the cloud-based solution of the disease diagnosis system even retaining their privacy. Patients start with registration and obtain a table of private keys and disease indexes. This allows for secure depiction of encrypted medical data and diagnostic services (Brakerski et al., 2013). [twenty This data protection presentation one]. authentication system prevents unauthorized access and ensures that only registered users can obtain appropriate medical information (Barua et al., 2011). [10].

Patients can access diagnostic services after later registration and use (Zhu et al., 2017). [11]. After registration, the customer has the option to visit the diagnosis in the disease section and enter the results of the test. To ensure that patient data remains confidential, the input data is encrypted and executed by trapdoor technology. This allows avoiding the unsafe comparison of similarity and disease classification without disclosing the actual medical data (Wang et al., 2022). [2]; (Xu et al., 2022) [17]. The system then applies privacy preserving ML algorithms of KNN and Mahalanobis Distance to examine the encrypted test data and offer a precise diagnosis (McLachlan, 1999) [19]; (Park and Lee, 2018). [12].

The User Module ensures that patients can access medical information with their private health information being held securely using secure encryption, privacy-sensitive machine learning and cloud-based diagnostic solutions. Such an approach generates a sense of trust in cloud-based healthcare systems, therefore, it is possible to conduct medical pre-diagnosis within a short time, without violating the privacy rights (Zhang et al., 2022) [5].

## iii) Extension:

The change is an addition to the proposed system since the ensemble ML algorithms, such as Random Forest and XGBoost are added to the separate ML models, KNN and SVM. This improvement will aim at simplifying and improving the diagnosis process of diseases using encrypted datasets. It is an attempt to achieve improved performance rates without sacrificing privacy-preserving capabilities of the original system.

## iv) Technologies:

# **Technologies Used in the Admin Module**

The administrator module includes a combination of encryption, cloud computing and machine learning to ensure that disease identification structures are safe and functional. One of the main technologies is matrix cryptography. This allows for safe outsourcing of medical data and avoids unauthorized access (Brakerski et al., 2013) [21]. Another useful feature of the system is Trapdoor technology that ensures simple repetition and calculation of encrypted medical records in the DriveHQ cloud, and maintains privacy (Dong et al., 2023). [1]. This is a cloud-based approach that simplifies the process of scaling and acceleration of computation and goes beyond regional storage and processing obstacles (Wang et al., 2022). [2]. The Administrator module uses ML techniques such as KNN and Mahalanobis distance classification to determine the disease a person needs. Algorithms allow accurate prediction of disease and ensure confidentiality (McLachlan, 1999) [19]; (Park and Lee, 2018) [12]. It is also visually examined using methods to visualize the visualization of arithmetic graphics to visualize the computational costs of trapdoor production and implementation of ML tasks (Xu et al., 2022). [17]. This combination of secure encryption, cloud storage and highly developed AI algorithms represents an effective, privacy and scalable framework for disease diagnosis.

## **Technologies Used in the User Module**

Technology recorded in the user module allows you to securely register, manage medical

information in an encrypted way, and diagnose illnesses without affecting patient privacy. The system begins with a way that patients can register by issuing private keys and disease index tables to safely examine medical documents (Barua et al., 2011). [10]. Data protection motor authentication procedures enable input of unwanted access and guarantee this process by simultaneously accessing health services easily stored in the cloud (Zhu et al., 2017). [11].

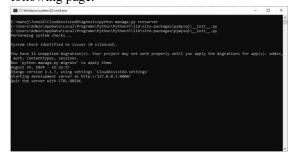
After the user is authenticated, the test data can be sent to the diagnostics. This information is protected by encryption and a secure match on the base secure similarity is used (Wang et al., 2022). [2]. The privacy-preserving ML frameworks, such as KNN and Mahalanobis Distance based classification, used in the system make it easy to predict diseases, and protect sensitive medical information (Haldar et al., 2017). [22]. Besides, protected remote computation in the cloud can be used to execute the ML models, which ensures a privacy-saving and scaling medical diagnostics methodology (Zhu et al., 2020) [14].

The User Module ensures that patients receive their diagnoses of diseases privately, safely and quickly through the application of secure encryption, cloud computing and privacy conscious classification techniques. This instills confidence in the solutions of cloud-aided healthcare (Zhang et al., 2022) [5].

## 4. RESULTS & DISCUSSION

In order to create a database, one should take the contents of the file named Database.txt and insert the contents in the MYSQL server.

In order to begin the project, one should simply run the runServer.bat file by clicking twice to get the following page.



The python web server begins on the following screen. At this point, open your browser and enter the following address: and then make sure you type the writings 127.0.0.1:8000/index.html. Then, hit the enter key to view the site below.



Click the "Admin Login" link on the screen above to get to the login page below.



The admin is logged in on the screen above, and after logging in, they will see the page below.



Admin can click on the "Upload Dataset" link in the screen above to get to the page below.



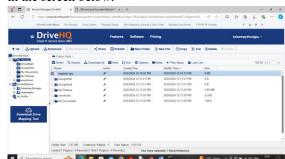
The admin can upload the "heart" dataset file from the screen above and then click the "Open" button to load the dataset and see the page below.



The dataset has been loaded on the screen above, and you can see all of the plain numbers. Now, click on the "Matrix Encryption & Trapdoor Outsource to Cloud" link to encrypt the dataset and send it to the DRIVE HQ cloud.



The screen above shows that all of the dataset values are encrypted. The encrypted trapdoor will be sent to the HQDRIVE cloud, which you can see in the screen below.



You can see the file "trapdoor.npy" saved in the HQDRIVE cloud in the screen above. To see it, go to "https://www.drivehq.com" and log in with the username "kaleempythongpu4@gmail.com" and the password "Offenburg965#." Then, click on the link that says "My Storage."

To train the ML algorithms, click on the "Run KNN & SVM" link in the program. This will give you the output below.



You can see that both algorithms acquired more than 98% accuracy in the screen above. Now click on the "Comparison Graph" link to see how much it costs to make a trapdoor.



The x-axis on the screen above shows the number of trapdoors generated, and the y-axis shows the cost of computing trapdoors. Now, log out and sign up a new patient.



The patient is entering their signup information on the screen above and then clicking the button to get to the page below.



The patient has finished signing up on the screen above. Now click on the "Patients Login" link to proceed to the page below.



The patient is now logged in on the screen above. After they log in, they will see the page below.



If the patient clicks on the "Diagnose Your Disease" link on the screen above, they will go to the page below.



The patient can enter cardiac test data values on the screen above. You can copy these values from the "Dataset/testData.csv" file and then click the button to get the page below.



The first line of the screen shows the patient's original data, the second line shows the encrypted numbers, and the third line shows the anticipated output, which is "No Heart Disease detected." You can also enter other test data and find out what disease it is.



The screen above shows the input for another test and the output below.



Heart illness was found on the screen above.

## 5. CONCLUSION

The project successfully combines KNN, SVM, and more complex algorithms like RF and XGBoost. It has an excellent accuracy rate of over 98% when diagnosing diseases. This is a big improvement above standard ways of diagnosing. The technology uses a new matrix encryption method to keep patient data private during the diagnosis process. This directly addresses important privacy issues that come up with cloudbased healthcare solutions. The use of a hierarchical index tree also speeds up the searching and retrieval of encrypted medical records, which greatly improves the overall performance and efficiency of the diagnostic system. The outsourced model can be hosted on cloud servers and be exposed to pre-diagnostic results without exposing their personal data to risks. The latest ML algorithms and high-security standards ensure that medical pre-diagnosis is conducted in a manner that will not violate privacy with the help of this new idea. On the whole, the project represents a secure and efficient means of utilizing the cloud to perform pre-diagnosis medically, which will eventually result in improved health outcomes and ensure privacy of the patient information.

## Future Scope:

The future versions of the proposed privacy-saving pre-diagnosis system can be focused on numerous crucial additions to the working capabilities and usability. An example of this achievement is the

application of more sophisticated cryptographic tools, including homomorphic encryption and secure multi-party computation, that allow making complex computations on encrypted data without exposing any patient information. ensuring that it can integrate with other EHR systems and other healthcare applications will simplify the data sharing among healthcare professionals without jeopardizing the data security. It also could be possible to design realtime data processing, which would allow healthcare workers to make fast decisions using the latest information about patients without sacrificing privacy. The strategies and models created in this project can be applied to other areas, including banking, insurance, and smart cities. It is now possible to analyze and make decisions regarding safe data in a broader scope of businesses.

#### REFERENCES

- [1] C. Dong et al., "Maliciously secure and efficient large-scale genome-wide association study with multi-party computation", IEEE Trans. Dependable Secure Comput., vol. 20, no. 2, pp. 1243-1257, Mar./Apr. 2023.
- [2] X. Wang, J. Ma, M. Yinbin, X. Liu and Y. Ruikang, "Privacy-preserving diverse keyword search and online pre-diagnosis in cloud computing", IEEE Trans. Serv. Comput., vol. 15, no. 2, pp. 710-723, Mar./Apr. 2022.
- [3] B. Xie, T. Xiang, X. Liao and J. Wu, "Achieving privacy-preserving online diagnosis with outsourced SVM in internet of medical things environment", IEEE Trans. Dependable Secure Comput., vol. 19, no. 6, pp. 4113-4126, Nov./Dec. 2022.
- [4] S. Zhang, S. Ray, R. Lu, Y. Zheng, Y. Guan and J. Shao, "Achieving efficient and privacy-preserving dynamic skyline query in online medical diagnosis", IEEE Internet Things J., vol. 9, no. 12, pp. 9973-9986, Jun. 2022.
- [5] M. Zhang, Y. Zhang and G. Shen, "PPDDS: A privacy-preserving disease diagnosis scheme based on the secure mahalanobis distance evaluation model", IEEE Syst. J., vol. 16, no. 3, pp. 4552-4562, Sep. 2022.
- [6] M. S. Hossain and G. Muhammad, "Cloud-assisted industrial Internet of Things (IIoT)—enabled framework for health monitoring", Comput. Netw., vol. 101, pp. 192-202, 2016.
- [7] 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 Gener. Comput. Syst., vol. 86, pp. 527-534, 2018.
- [8] J. Liang, Z. Qin, S. Xiao, L. Ou and X. Lin, "Efficient and secure decision tree classification for cloud-assisted online diagnosis services", IEEE Trans. Dependable Secure Comput., vol. 18, no. 4, pp. 1632-1644, Jul./Aug. 2021.
- [9] J. McKeon, "Houston health department suffers healthcare data breach", [online] Available: https://healthitsecurity.com/news/houston-health-department-suffers-healthcare-data-breach.
- [10] M. Barua, X. Liang, R. Lu and X. Shen, "ESPAC: Enabling security and patient-centric access control for ehealth in cloud computing", Int. J. Secur. Netw., vol. 6, no. 2/3, pp. 67-76, 2011.
- [11] H. Zhu, X. Liu, R. Lu and H. Li, "Efficient and privacy-preserving online medical prediagnosis framework using nonlinear SVM", IEEE J. Biomed. Health Informat., vol. 21, no. 3, pp. 838-850, May 2017.
- [12] J. Park and D. H. Lee, "Privacy preserving K-nearest neighbor for medical diagnosis in E-health cloud", J. Healthcare Eng., vol. 2018, pp. 1-11, 2018.
- [13] J. Hua, G. Shi, H. Zhu, F. Wang, X. Liu and H. Li, "Camps: Efficient and privacy-preserving medical primary diagnosis over outsourced cloud", Inf. Sci., vol. 527, pp. 560-575, 2020.
- [14] D. Zhu et al., "CREDO: Efficient and privacy-preserving multi-level medical pre-diagnosis based on ML-KNN", Inf. Sci., vol. 514, pp. 244-262, 2020.
- [15] D. Zhu, H. Zhu, X. Wang, R. Lu and D. Feng, "An accurate and privacy-preserving retrieval scheme over outsourced medical images", IEEE Trans. Serv. Comput...
- [16] L. Liu et al., "Toward highly secure yet efficient KNN classification scheme on outsourced cloud data", IEEE Internet Things J., vol. 6, no. 6, pp. 9841-9852, Dec. 2019.
- [17] C. Xu, N. Wang, L. Zhu, C. Zhang, K. Sharif and H. Wu, "Reliable and privacy-preserving top-k disease matching schemes for E-healthcare systems", IEEE Internet Things J., vol. 9, no. 7, pp. 5537-5547, Apr. 2022.
- [18] N. Cao, C. Wang, M. Li, K. Ren and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data", IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 1, pp. 222-233, Jan. 2014.
- [19] G. J. McLachlan, "Mahalanobis distance", Resonance, vol. 4, no. 6, pp. 20-26, 1999.

- [20] J. Yuan and Y. Tian, "Practical privacy-preserving mapreduce based K-means clustering over large-scale dataset", IEEE Trans. Cloud Comput., vol. 7, no. 2, pp. 568-579, 2019.
- [21] Z. Brakerski, C. Gentry and S. Halevi, "Packed ciphertexts in LWE-based homomorphic encryption", Proc. Int. Workshop Public Key Cryptogr., pp. 1-13, 2013.
- [22] N. A. H. Haldar, F. A. Khan, A. Ali and H. Abbas, "Arrhythmia classification using mahalanobis distance based improved fuzzy C-means clustering for mobile health monitoring systems", Neurocomputing, vol. 220, pp. 221-235, 2017.
- [23] D. Charalampidis, "A modified k-means algorithm for circular invariant clustering", IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 12, pp. 1856-1865, Dec. 2005.
- [24] S. Kamara, C. Papamanthou and T. Roeder, "Dynamic searchable symmetric encryption", ACM Conf. Comput. Commun. Secur., pp. 965-976, 2012.
- [25] X. Wang, J. Ma, X. Liu, Y. Miao, Y. Liu and R. H. Deng, "Forward/backward and content private DSSE for spatial keyword queries", IEEE Trans. Dependable Secure Comput.