Revolutionizing Healthcare: Cloud-Based Health Information Exchange and Disease Prediction

Sai Nithin Nayini1*, Vishwas Gade1, Mani Prakash Reddy Gudapareddy1, Varshith Peddineni1 and Dr. Beena B.M.1

1 Department of Computer Science & Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Bangalore, Karnataka, 560035, India

Abstract

A cloud-based system called Health Information Exchange (HIE) is hosted on Amazon Web Services (AWS). This system provides logins, for doctors, administrators and patients. The primary objective of this work is to revolutionize healthcare data sharing by providing a convenient platform. Patients are empowered to manage their health by selecting their symptoms and receiving disease predictions. AWS is utilized to orchestrate the cloud-based infrastructure, which includes SageMaker for constructing machine learning models RDS for a PostgreSQL database, EC2 for hosting the Django server, Lambda for deploying machine learning model APIs CloudWatch for resource monitoring and IAM, for access control. The core element of this system is the implementation of a classification framework that optimizes the organization and retrieval of data. This innovative approach enhances data accessibility while streamlining healthcare processes to facilitate decision making and patient care. By integrating AWS services, it also creates a scalable foundation that supports the growth of a dynamic retail environment.

Keywords

Amazon Web Services (AWS), Symptoms, Cloud Computing, Diseases Prediction, Machine learning.

1. Introduction

The introduction of cutting-edge technologies is driving a revolutionary transformation in the modern healthcare environment. One of the advancements, in healthcare is the introduction of a Cloud Based Health Information Exchange (HIE) system [1]. This study focuses on this groundbreaking system, which is intelligently hosted on the infrastructure of Amazon Web Services (AWS) [2]. Healthcare stakeholders are increasingly recognizing the importance of easily accessible and well-organized platforms to facilitate communication and data sharing. This article aims to address these needs by offering a user-friendly solution.

The main element of this system is a user website with login portals for administrators, physicians and patients. This promotes collaboration. Enables management of health information [3]. This unique approach emphasizes inclusivity in order to enhance communication and data exchange while acknowledging the roles within the healthcare ecosystem. By implementing these customized logins for user types, the system caters to the requirements and responsibilities of patients seeking proactive health management physicians requiring seamless access, to patient information and administrators overseeing overall platform operations.
One notable feature of this proposed system is its ability to empower patients through insights. Patients have the opportunity to input their symptoms and subsequently receive forecasts regarding conditions. This feature not only engages individuals, in managing their health but also contributes to the broader transformation of healthcare towards a patient centered approach. By empowering patients with capabilities, the system aims to encourage understanding of health matters and cultivate a sense of personal responsibility for one’s well-being.

A significant innovation of this system is its classification scheme, for organizing information. This technology enhances the retrieval and analysis of health data, such as scans and medical reports by categorizing the information. This categorization process improves efficiency by reducing the time spent searching for information thereby enhancing the quality of healthcare services and decision making. In summary with its patient prediction capabilities and categorization system this Cloud Based HIE system represents a groundbreaking advancement, in redefining healthcare information management while fostering a [5] well-organized and responsive healthcare environment.

The remaining work is presented as follows: Section II contains Literature Survey of this work. Section III clearly talks about the Datasets for different diseases. Section IV contains AWS Services in this work and Section V explains information regarding the methodology that has been followed in the work. Section VI of this paper explains the Results and Sections VII summarizes the conclusion. Section VIII provides the future scope of the work. Section IX lists all the references referred as part of this work.

2. Related Work

The importance of individuals having control, over their health data during exchanges is emphasized in the paper [1]. The paper discusses two protocols for Health Information Exchange (HIE); one that ensures the backup of health data to trusted cloud storage and another that enables the exchange of health information between citizens and healthcare professionals in emergency situations. The paper evaluates these protocols efficiency and practicality using a scenario. It acknowledges limitations related to evolving interoperability and security requirements. Assures readers that these challenges are addressed comprehensively within the manuscript. Additionally, the paper examines the adoption of cloud technologies [2] for storing and processing health data within HIEs. It specifically focuses on introducing an Enterprise Architecture (EA) designed to facilitate the transition of HIEs to a cloud infrastructure. The EA serves as a blueprint outlining how resources should be strategically placed in the IT environment to support core business functions. Successful implementation of this architecture is expected to provide organizations with an understanding of their resources align with business goals and guide them towards achieving those goals through effective IT support. Finally key findings, from the research highlight how performance expectancy, effort expectancy, social influence, facilitating conditions, data security and information sharing all influence individual’s behavioral intentions regarding HIE usage [3]. However, the impact of cloud-based health information did not show any significance. The study ends with a warning, for policymakers highlighting the need for evaluation before adopting cloud-based health centers due to challenges, in integration. The increasing use of cloud computing, in the healthcare industry driven by its ability to handle the growing volume and variety of healthcare data [4] underscores the importance of data acquisition, integration and large-scale analysis provided by cloud infrastructures. Factors such as security, availability and disaster recovery play a role in driving this transition.
However, there are challenges involved in migrating healthcare workloads to the cloud due to factors like healthcare data standards, heterogeneity, sensitivity and regulatory constraints. The paper emphasizes the significance of acquiring healthcare data through cloud-based solutions. Provides insights into the challenges, requirements, use cases and best practices for developing an advanced healthcare data ingestion service on the cloud. The study also explores the adoption of Health Information Exchange (HIE) [5]. Recognizes its benefits for healthcare professionals. Although information technology has facilitated HIE adoption among hospitals and healthcare systems thus far. Additionally discussed is a Cloud Based Personal Health Record (PHR) system along with a Personalized Indian Health Network (PIHN) which connects users with conditions [6]. PIHN serves as a platform for patients to share experiences and store their records specifically tailored for the population. It includes a self-tracking system that monitors health conditions well as a recommender system that connects individuals, with similar profiles. User privacy is of importance. This system is built using Django and Firebase enabling us to present user data in a manner providing quick and valuable health insights. Patient recommendations, between individuals utilize filtering algorithms providing a solution for personalized health management and efficient support in healthcare. The challenges related to maintaining privacy in cloud based Personal Health Records (PHRs) emphasize the importance of safeguarding information and addressing concerns such as data leakage [7]. Author focus lies in securing PHRs that are outsourced to Cloud Service Providers (CSPs) introducing the Rail Fence Data Encryption (RFDE) algorithm to enhance privacy standards. RFDE employs a transposition cipher also known as zigzag encryption to protect PHR files from access. This algorithm ensures security, flexible accessibility and effective management of user privacy risks while addressing the challenges posed by cloud storage. Their proposed algorithm showcases performance when compared to methods contributing towards improved control over data access and privacy within PHRs. To enhance modernization efforts, they have explored the integration of technology into hospital information management systems. This entails outlining the framework as hardware and software components of the system with a focus on real time transmission and cloud-based sharing of medical information [8]. The study establishes a platform for sharing data based on cloud computing. Through this scheme author demonstrate its effectiveness in monitoring while enhancing both flexibility and security, within hospital services. Bitbox Introducing [9] a user web application created to simplify the exchange of medical imaging data. It tackles the obstacles posed by the amount and varied complexity of data guaranteeing both privacy protection and data integrity. Bitbox allows for the transfer of imaging and non-imaging information, from external locations, to a centralized server. The functionality of the system is demonstrated in the COVID 19 Clinical Neuroscience Study (COVID CNS) work, which shows how it effectively helps investigate the neuropsychiatric effects of COVID 19 infections in a scale multi-site study. As cloud computing continues to advance and businesses increasingly rely on servers [10], for data handling there are concerns about security and privacy. To address these concerns, a proposed approach for sharing information ensures both the security of cloud storage data and the privacy of its owner. This approach offers flexibility in data usage while also tackling challenges related to isolation and security in data sharing. The examination of this plan reveals its effectiveness in enhancing security and preserving privacy within cloud storage systems. The use of mobile cloud technology and the Internet of Medical Things (IoMT) [11] is explored for automated diagnosis and health monitoring specifically focusing on tracking the progression of a disorder through a motor coordination test. This scheme leverages cloud server computing and storage capabilities to assess severity levels based on measurements. An Android application is utilized for data acquisition and communication with the cloud. Furthermore, the system is integrated with a network to ensure reliable data exchange, among healthcare users.
The results of the experiments prove that the suggested system is both feasible and effective highlighting its use in healthcare applications. One of the obstacles, to sharing data is the political and financial factors involved. To address this issue a cloud based integrated clinical information system [12] has been proposed as part of the Hospital Information System. The system allows for sharing and exchange of data in large hospitals. By utilizing cloud computing it simplifies the management of hospital information saving both time and money while also improving analysis and accessibility of data. Incorporating intelligent authentication strengthens the security of information stored on clouds providing a solution, for hospital administration [13]. This research highlights the challenges faced in terms of password security and data privacy when it comes to cloud-based health services. It emphasizes the importance of transmitting health records (PHRs) and suggests a combined authentication approach using RFDE models along with encrypting PHR files using Rail Fence Data Encryption (RFDE) to enhance confidentiality. The proposed technique shows efficiency and security compared to methods addressing concerns regarding information leakage and regulatory compliance when storing sensitive health data on the cloud. Furthermore, this study addresses challenges related to patient monitoring (RPM) devices by introducing an edge cloud computing architecture that integrates RPM data through edge computing and uploads latent representations for AI assisted decision making [14]. The model incorporates edge modules for medical image analysis as time series analysis utilizing machine learning algorithms for anomaly detection and severity classification. Additionally, the cloud telehealth management module employs networks (CNN) recurrent neural networks (RNN) along with attention models to create personalized treatment plans. The platform has been tested on RPM devices. Shows improved speed and accuracy in diagnosing conditions. This paper highlights the growing importance of healthcare information technology with a focus on standardizing the exchange of data and interoperability, between Health Information Systems (HISs) and e health applications using a cloud-based service [15].

The proposed service acts as a centralized entry point for retrieving Electronic Medical Records (EMR) from different HISs and serves various e-health applications. Additionally, the research introduces a unified secure platform enabling access to a framework for retrieving and managing medical records and Personal Health Records (PHR).

3. Methodology

The system's approach consists of a number of processes that include data collection, analysis, model development, and prediction within a structured framework.

3.1. Data Collection and Preparation

In this work different diseases datasets are collected from Kaggle and trained using machine learning models.

A. The dataset for Brain Tumor Detection is collected from Kaggle which contains 1300 images of two classes namely, no tumor and positive tumor.
B. The dataset for kidney stone detection is collected from Kaggle which includes 13,691 images with conditions divided into three categories: cyst, normal, stone.
C. For Skin Disease Classification, the extensive dataset of 37,961 instances covers diverse skin conditions such as 'Benign,' 'Malign,' 'Enfeksiyonel,' 'Ekzama,' 'Akne,' and 'Pigment'.
D. The disease prediction using symptoms dataset consists of 133 symptoms with 5000 samples in csv format. Based on the symptoms it predicts the disease the person is suffering with.
3.2. AWS Services

In this study a Health Information Exchange (HIE) system is introduced that is built on the infrastructure provided by Amazon Web Services (AWS). This system represents an advancement, in the field of healthcare offering users a website with separate logins for administrators, physicians and patients to promote active engagement in managing their health [16]. By selecting their symptoms patients can leverage the efficiency of cloud technology to anticipate diseases. The main objective is to revolutionize the way healthcare data is exchanged, allowing for sharing and retrieval of medical reports and scans. The cloud architecture incorporates classification techniques to enhance organization and make data readily accessible for decision making [17]. This work aims to establish standards in accessible health information management setting a high benchmark for the digital era of healthcare.

Amazon Web Services (AWS) provides Cloud Computing services such, as Amazon Elastic Compute Cloud (EC2) AWS Lambda, AWS Relational Database Service (RDS) Sagemaker, Cloud Watch and Amazon Identity and Access Management (IAM).

Amazon EC2 is one of the services provided by AWS that offers computing power in the cloud. In this work it is used as a hosting environment, for the Django server. This allows us to easily scale the resources based on demand and ensure performance for the web application. AWS Lambda is another service from AWS that is utilized to host the APIs for machine learning models. Its serverless architecture enables code execution in response to events providing a cost scalable solution for hosting the backend of the machine learning functionalities.

Amazon RDS (Relational Database Service) simplifies database management for this work. PostgreSQL is specifically as the database engine with RDS, which ensures easy to manage storage for all applications data, enhancing the overall robustness of the system. When it comes to building, training and deploying machine learning models Amazon Sagemaker is utilized. A service that streamlines these processes and seamlessly integrates with AWS services. For real time application and resource monitoring needs Amazon CloudWatch is needed. It provides insights, into how different parts of system are performing enabling us to proactively take action and maintain a healthy and efficient system. The work utilizes AWS Identity and Access Management (IAM) to manage access control. IAM adds a level of security and governance, to the system by overseeing users and their permissions guaranteeing secure entry, to AWS services.

3.3. Model Development

A. All Web page creation and hosting in cloud

First, a signup page is created for a web application where administrators, doctors and patients can log in separately. Patients can securely access the platform choose their symptoms and receive disease fore-casts based on algorithms. To ensure organization of data a robust categorization framework is also implement for effective classification. Various services are utilized to deploy it on the AWS cloud. The web application is hosted on Amazon EC2 and AWS Lambda manages the APIs for machine learning models using a serverless architecture. Amazon SageMaker handles model development while real time resource monitoring is taken care of by Amazon CloudWatch. By implementing access control through AWS IAM data security is prioritized. This comprehensive approach combines application development, with AWS cloud deployment to leverage the transformative advantages of cloud technology in enhancing accessibility, scalability and reliability of health information.
The workflow as shown in Figure 1 depicts the step-by-step implementation of this work and components used to design. The Application’s Front-end was designed with the developing techniques which are HTML, CSS and JS. The Database used for this is PostgreSQL. It can also be used in any device which makes it compatible to use. After the online application is ready it is hosted in AWS cloud platform using various AWS services available.

To host the application in the cloud, various services of AWS are used as shown in Figure 2. The Django server relies on Amazon EC2 as its core infrastructure, which provides a scalable framework, for user interactions. AWS Lambda effortlessly hosts the APIs of machine learning models ensuring dynamic real time predictions [18]. To ensure storage Amazon RDS is utilized with PostgreSQL database. Additionally, this platform is enhanced with machine learning capabilities, through Amazon SageMaker. AWS IAM provides robust access control measures while AWS CloudWatch constantly monitors resources. This AWS architectural diagram showcases how these components are integrated synergistically [19]. This integration not only used in an era of intelligent and easily accessible healthcare solutions but also facilitates the exchange of healthcare data while ensuring the classification and protection [20] of vital medical information.
B. Various Machine Learning models training

i. Brain Tumor Detection: The dataset has been split into an 80:20 ratio, and SVM and Logistic Regression have been employed to classify brain images into two categories: ‘no tumor’ and ‘positive tumor’. The trained models for each classifier have been individually saved in H5 format. This h5 is stored in AWS lambda to predict the disease of a new patient.

ii. Kidney stone Detection: The dataset consists of 11,202 training and 2,489 validation files, categorized into 'Cyst', 'Normal', 'Stone'. The MobileNetV2 model is utilized for training. MobileNetV2 is a convolutional neural network consisting 53 deep layers. It is a pretrained version of the network trained on more than a million images from the ImageNet database. Similar to the above, here also, the model is trained and saved in h5 format and stored in AWS lambda to predict the disease of a new patient.

iii. Skin Disease Detection: The dataset consists of 30910 images of various skin diseases divided into six classes 'Benign', 'Malign', 'Enfeksiyonel', 'Ekzama', 'Akne', and 'Pigment'. Xception model is used for training. Xception is a convolutional neural network consisting 71 deep layers. It is an also pretrained model trained on millions of images.

iv. Disease Prediction using Symptoms: The dataset consists of 133 symptoms, based on the symptoms selected from the patient the disease is predicted. This dataset is trained using Naive bayes classifier. It is a classification algorithm based on Bayes theorem.

Figure 3. Model training flowchart

The main aim of the model training is to create a model that generalizes well on the unseen new data. So, different models are trained on different datasets as shown in Figure 3, to make predictions to patients about the diseases they are suffering with.
4. Experimental Results

The Application has been deployed on AWS Platform using the AWS Services which are present in the AWS Architecture.

The application is hosted on Aws using EC2, it provides scalable and customizable virtual servers, allowing users to deploy and manage application easily. In this work, symptoms-based disease prediction model is hosted in EC2 for the prediction.

Amazon RDS is a managed database service on AWS. It is used for hosting and operating relation databases like MySQL, PostgreSQL, etc. In this work PostgreSQL is used to store login credentials data of different users.

Amazon Lambda, a serverless computing solution, makes it possible to run code without having to worry about maintaining servers. In this work, saved h5 files of first 3 diseases are hosted in the lambda for reliability and faster prediction.

Sagemaker, a machine learning service, it makes the process of building, training and deploy machine learning models easier at scale. In this work, models are trained in sagemaker.

CloudWatch is a monitoring and observability service, it allows users to collect and track metrics, monitor logfiles, set alarms.

Below are the images of the application hosted in AWS.

![Home page of the website](image1.png)

**Figure 4. Home page**

Home page of the website as shown in Figure 4 has some information about hospital, it is similar to both doctor and patient after logging in.

![Doctor and Patient login](image2.png)

**Figure 5. Doctor and Patient login**

Login page of both patient and doctor is displayed in the above image as shown in Figure 5, where doctor has options to view consultation history and give feedback and patients has check disease, view consultation history and feedback options.
When patients select some symptoms, he is suffering with then the model predicts the disease with confidence level displayed and it also gives suggestion for which doctor the patient should consult as shown in Figure 6. Patients can also check disease using MRI scans by clicking on the option in the webpage. Naive Bayes model is used to predict the disease.

The Patient also can upload the scan of his Brain, Kidney and Skin image as shown in Figure 7 and the machine predicts the problem whether the person is suffering with any problem if so, which type of disease he is suffering with. The machine is trained with machine learning models to predict and the results for the models are shown below.

When patient clicks on consultant doctor, a chat box will appear where patient can discuss about their problem with doctor as shown in Figure 8.
Using the machine learning model, the model is trained to predict on unseen data and the results are shown in Figure 9 where, brain images with no tumor are classified as no tumor and brain images having tumor are classified as positive tumor.

![Figure 9. Brain Tumor Detection](image)

The model is trained and tested on unseen data, where it makes predictions on patients MRI images of kidney stones as shown in Figure 10, categorizing them into three types: Normal, Cyst, and Stone.

Similarly, Skin disease is trained with Xception model, where it makes predictions on patients uploaded images.

<table>
<thead>
<tr>
<th>Model</th>
<th>Training Accuracy</th>
<th>Testing Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression (Brain Tumor)</td>
<td>0.98</td>
<td>0.95</td>
</tr>
<tr>
<td>SVM (Brain Tumor)</td>
<td>0.99</td>
<td>0.96</td>
</tr>
<tr>
<td>MobileNet V2 (Kidney Stone)</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Xception (Skin)</td>
<td>0.99</td>
<td>0.90</td>
</tr>
<tr>
<td>Naive Bayes (Symptoms based)</td>
<td>0.95</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Table 1. Training and Testing Accuracy for various diseases and models used

Training accuracy and testing accuracy of different models for different diseases is displayed in the table as shown in Table 1. Brain Tumor got better results with SVM, so SVM is used to predict on unseen data in the application.
5. Conclusion

This work titled "Cloud Based Health Information Exchange, with Categorization on AWS" has the potential to revolutionize the healthcare industry. It aims to create a user data driven ecosystem that empowers individuals to take a role in managing their health. By utilizing a categorization system on AWS, this work streamlines processes. Provides healthcare providers with quick access to patient data for making well informed decisions. This not optimizes resource utilization. Also has a significant impact on society by improving the effectiveness, accessibility and customization of the healthcare system. This work places importance on ensuring data security and privacy by leveraging AWS IAM giving users confidence that their confidential medical information is protected. Beyond advancements this work represents a shift towards a more interconnected healthcare infrastructure that aligns seamlessly with responsible data management practices. In addition to enhancing information exchange this cloud-based solution lays [21] the groundwork for a connected and patient centric future, in healthcare.

6. Future Scope

The future of healthcare technology advancements holds promise. The success of this work relies on expanding the types of healthcare data promoting integration, with systems and utilizing AI and machine learning for more precise disease predictions. In addition to aligning with the rising popularity of virtual healthcare services through telehealth and remote monitoring capabilities blockchain technology enhances security and transparency. By incorporating user driven features, a feedback system and global scalability, with interoperability standards this work will adapt to the evolving needs of users. This work will be at the forefront of healthcare innovation and contribute to the development of a more user-centric, connected, and accessible healthcare ecosystem by taking these new directions.

References


