



## DNN-Apache Hive Environment for Big Data in the Health Sector with Data Private Protection

5024

S. Angel Latha Mary<sup>1</sup>, Dr.K.Ashwini<sup>2</sup>, Dr.N.Bharathiraja<sup>3</sup>  
P.ShanmugaPriya<sup>4</sup>

Professor , Department of Computer Science and Business Systems, Sri Eshwar College of  
Engineering, Coimbatore<sup>1</sup>

Associate Professor, School of Commerce JAIN (Deemed to be) University, Bengaluru<sup>2</sup>

Associate Professor ,Department of Biomedical Engineering, Vel Tech Multi Tech

Dr.RangarajanDr.Sakunthala Engineering College, Avadi<sup>3</sup>

AssistantProfessor ,Department of Electronics and Communication Engineering, Rajalakshmi  
Engineering College, Thandalam, Chennai<sup>4</sup>

xavierangellatha@gmail.com<sup>1</sup>, ashwinikdeepak@gmail.com<sup>2</sup>, bharahtinatarajan78@gmail.com<sup>3</sup>  
shanmugapriya.p@rajalakshmi.edu.in<sup>4</sup>

### Abstract

Clinical data are the most important data as the data stored in the healthcare records are about the patient's health status and personal information. These clinical data are gathered and are further used by the doctor professionals to analyze the patient's health condition. This consists of clinical health records with the enormous clinical type decisions and diagnosis. Big data includes about the medical stakeholders and other pharmaceutical resources. It reviews different challenges which include the data resources using the big data. Large data is stored in the electronic manner, which includes terabytes to perabyte of data. Data are stored in the relational database management system which reads, writes or stores the data. In this proposed patent, the patient's health condition report is diagnosed to analyze the chronic illness or early diagnosis of health condition of patients. Apache Hive is used to standardize the health-related massive data which are generated. These health reports are organized in an efficient manner which stores and analyses the data. This framework creates the big data to collect these data in a resourceful manner. This Apache hive manages and collects the data for analysis with the minimal amount of time. Deep learning analysis is used to diagnosis the medical data analysis and for the visualization. Health monitoring with the big data and deep learning process the data for processing, cleansing and trained datasets. These deep learning techniques represent the high level of extracting process which abstracts the data in a hierarchical way of manner. Extraction of the complex patterns streams the data diagnosis, which simplifies the processes. Unstructured data are diagnosed into the structured manner which correlates the termed output. Big data applications are used in the healthcare and IOT applications for diagnosing the medical illness, medical record and other clinical data. Apache Hive Query analysis and provides the file distributed system and databases. It integrates the hive SQL which transforms the query database and processes it. DNN gives good and accurate performance with the different tools and techniques. These frameworks improve to diagnose the patient's health condition and diagnose the chronic illness. Big data analytics enhances the outcomes of the patient, personalized care for patients and reduces the time consumption. Homomorphic encryption encrypts the data in the server in the cloud where the practitioner/physician gives the request to access the information in the encrypted mode.

**Keywords:** Big Data, DNN, SQL, Apache Hive Framework, Diagnosing Illness, Homomorphic Encryption

DOI Number: 10.14704/nq.2022.20.10.NQ55478

NeuroQuantology 2022; 20(10): 5024-5034

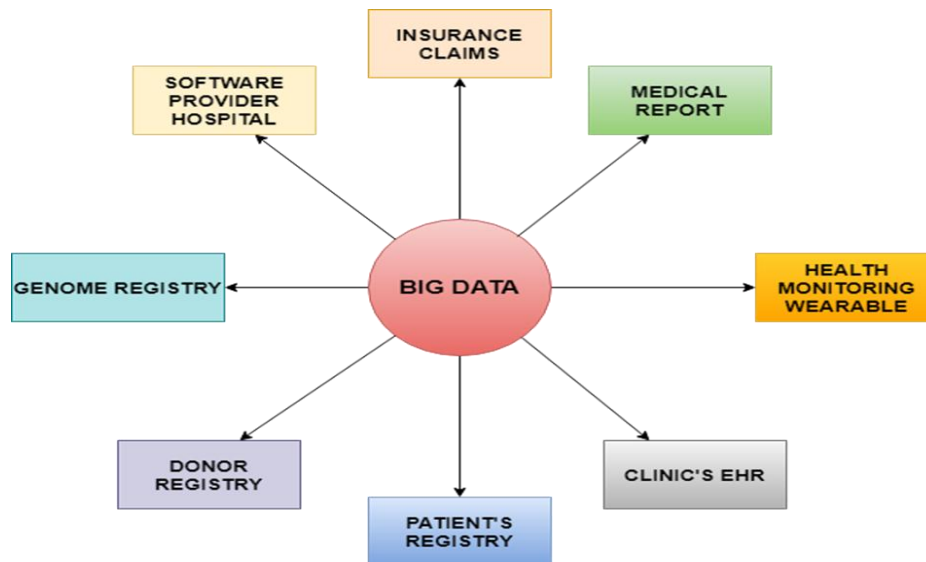


## 1. Introduction

Big data is used in the healthcare sector where the patient's health condition and personalized care is given. These analytics improves the outcomes of the patient which are stored in the electronic health record (EHR) which stores the patient's health record. In the healthcare sector, the data are characterized in the refined form where it categorizes based upon their features. Based upon the structured, semi-structured and unstructured manner, the records are determined. Big data establishes patient's data based upon their sources. Healthcare is one of the prominent industries as the diagnosis and treatment of patients takes place. To avoid the complexity over the data which causes the irregular complexity in the data? These can cause the patient's data to get corrupted. To predict the early prediction over the patient's health condition, EHR is implemented. EHR consists of a vast amount of data which is stored in those records which is handled using big data. In this digital technology, the data which are recorded are pre-processed, cleansed, and trained to attain the accurate results. The complexity and the speed are structured based upon their structured and unstructured data which are generated. The massive collection of complex data is processed at the highest possible response time. Different parameters of the patient's record such as vital parameters and chronic illness of the patient are identified as early intervention of patients. Gathering of the information plays an important key role in the development of the organization. Critical information of the organization is generated using the big data. These big data are

categorized based upon their volume, variety and its velocity based upon their response time effectively. Compared to the EHR, big data can store the entire gene driven by the patients and their record in a structured manner. Comparatively, big data reduces the cost-effective and time-consuming by improving the health outcomes. Even in the biological gene which exhibits the molecular biology experiments and genetic components, information is provided by using the big data. The patient's data has to be encrypted and the data has to be preserved in the encrypted manner. The encrypted data preserves the data from compromising it. A healthcare record incorporates with the electronic medical records such as about the patient's health record details. These specifically carry out all the data by detecting the early abnormalities of the patient's health deterioration. Big data specifically analyze the statistics of the data by reducing the costs and prevention of onset diseases. Big data analytics improves the patient-based services by monitoring the quality of the medical examiners. These DNN incorporates with the big data to analyze the vast data by discovering the patterns. This potential relies on the way of detecting the patterns and knowledge accordingly. The complexity of the big data arises from the electronically capturing the information. Big data identifies the irregular changes in the patient's data. The predictive analysis of the data is by using the big data to improve the public surveillance and assess the large volume of information. Specific health data are drastically checked to avoid the epidemic and cure of the disease.





**Figure 1: Big Data in HealthCare Sector**

Big data collects massive data with a large set of information. These records collect the patient's health which gathers the complicated set of information of all datasets. Enormous range of data are categorized according to the generation of data such as it generates, exploits the massive amount of data. Based upon the different sources of information, such as internet-connected devices or an IOT device transmits their data into the digital system. This digital system diversely computes the larger amount of data with less computational cost. In the healthcare industry, the advanced state of technology is pivoted in the healthcare sector. To enhance the productivity and growth in the service quality, the big data is initiated. This new technology categorizes the data based upon the generation, storage and transmission of the data. It incorporates with the healthcare data of the patients which includes the medical history, biometric of the patient and disease prediction of the patient. To handle the complicated datasets, the vast amount of data is handled using the conventional type of database. Digital technology growth evolves vastly in the healthcare sector and other computing techniques. Day-to-day data are collectively gathered and processed. These data which

are collected are massive in the state, which causes the data to be missing and inefficient. This insufficient space can cause complexity where the important gathered data are not processed. The structured manner of the data exhibits the data knowingly for the large massive amount of data. In the industry sector, finance and healthcare sector, the big data plays a pivotal role in storing and gathering the data in a structured manner. These big data analyse the information and attains the accurate results and performance based on the decision-making techniques. Since the data are complex in nature due to the vast amount of data; the data are advanced using the level of response time. Apache Hive framework reduces the time-response and experiments the data which have been recorded. Hive used in the healthcare dataset collectively gathers the set of information in a processed manner. Along with the Hadoop, the data which are collectively gathered are acknowledged for the constructive manner. Sensitive information about the patient's health condition is effectively handled by preserving the data in an encrypted manner. The data which are gathered in the datasets are stored in the collective manner where it enhances the privacy of the data by preserving it. To



improve the privacy over the data records in the big data, the data privacy preservation is enhanced. Each data is defined in a constructive manner. It identifies the symptoms of the patients by diagnosing the diseases in early intervention. Based upon their structural classification of the data, they are developed accordingly. In this big data analysis, the abundant types of potential are supported using the decision support techniques. Personalized care is used for the patients with the rapid advanced techniques. These frameworks support the development by organizing in a structured manner. Diverse complexities of the data are existed in the healthcare sector which computationally handles the medical data. The purpose of the patient's data is characteristically monitored, integrated and predicted using the different models. The compression of the medical data is predicted in a constructive manner. These can enrich the data and distils the knowledge such as diagnosing, medical images and medical reports. Big data is expanding its advancement in healthcare technologies at a larger vast space. The analysis of the data uses different data aspects such as data modeling, pre-process and visualization of the data. The each and every health information is monitored spontaneously without any delay in the period of time. Big data carries out the information based upon the various types of data productivity. Patient's data are analyzed using the relevant type of their personal information. These are stored in the electronic record applications which imply about the medical data records of the patient. These develop the sensor nodes where the different accumulated results are used. In the healthcare industry, big data implies on the various diagnoses, drug ailments prediction, disease prediction of the patients. Big data plays an important and crucial role in the different environments such as different

sectors such as healthcare, networking and other industries. These are generally classified in different criteria based upon their velocity, volume, variety, veracity and value. These different criteria are used to represent the data which analyses and collects the data analytics within it. It analyses the massive amount of data and predicts the high value and determines the precise results. In the upcoming years, big data can be used in different applications which give diverse knowledge and initiate new knowledge. In some of the applications, the patient data are analyzed and monitored. Using the frameworks and data which are stored the healthcare, are delivered in the big data for further processing. It checks and emphasizes the physicians to analyze the data of the patient and gives satisfactory results by giving accurate and good performance.

By using the big data with the DNN, the efficiency gets increased. Due to the efficiency, the performance is evaluated which attains the easier way to use. IoT-type devices can be used in the remote type of analyzing the health condition of the patient. Big data applications can be used in the public health sector, research development, operating clinics and analyzing the patient profile. Remote monitoring of the patient's health is monitored using the big data applications. Patient's health and safety are empowered and checked using the big data. IoT gives the greater impact with the big data which reduces the efficiency and consumes lesser time consumption. Prediction over the analysis of the healthcare is used for the future prediction of the patient's health outcomes within the clinical or non-clinical way of patterns within the data. This deep learning is used to attain the prediction of the chronic illness in the patient's health and discovery of the drugs dosage to the patients. Medical complications are reduced where the



diagnostic references are minimized due to the usage of the big data. Based upon the medical history, patient's data history, the efficient treatment is given to the best treatment to the patients. By the way of the responses, the patient's mortality rate is analyzed. IoT gives a major impact on the reduction in the healthcare costs and improves the analyzing of the patients monitoring. These give the prominent response in the healthcare applications of the patient's health status. The predictive model is characterized based upon the frameworks of the health state. It examines some of the clinical research state of the patients by analyzing the insurance claims, clinical set of registries. These analytics give the predictive analytics over the health condition of the patients, which includes the disease prediction, drug dosage and medical ailments.

One of the important domains in healthcare is to predict the disease prediction based upon the varied high state of the patients state i.e., health condition. Massive datasets are characterized based upon the clinical information which are assigned from the physician's order set of entries. Some of these datasets are generally complex in nature which is assigned in the fragments. Analytics over the prediction of the healthcare life and providers are assigned. Many different applications are emphasized to attain the high range of optimization as a result. Deep learning is used to minimize the complexity in the data by extracting the features with their known set of patterns using the improvised tools. These techniques reveal only the relevant set of information which can assist the physicians to identify the abnormalities in the medical reports of the patient's health. This deep learning hides the information and assists with the decision-making process.

To preserve the data from other intruders, the standard encryptions are implemented. This

encryption encrypts the data from any intruder or any misleading of the data. By using the location awareness, the patient data are accessed by tracking the information. Rather than using the pharmaceuticals way of identifying the disease prediction, the deep learning with the big data can give more accurate results. Enhancement of medical services can improve the patient's satisfaction. These resolve the problems and reduce the complexity by innovative type analytics. Big data analyses the structured and unstructured set of data by realizing the data within it. Data are getting increased in the healthcare sectors as the population is getting increased. To capture all the information, big data is used as it gives the complex structural view of the patient medical reports and records in the insurance claims of the patient. To attain the convenient type of performances in the data in the healthcare, the big data enhances the technologies using the cloud infrastructure systems. It determines based upon the casual relationships of the patient's condition and their disease prediction. Data which are collected using the IoT sensors are identified based upon the risk factors. Hive queries are used to analyze the datasets using the HDFS methodology. In the healthcare sector, the relational database of the big data of the healthcare is analyzed and monitored. Dnn trains up the model by minimizing the latency and improves the granularity in the trained set of datasets.

## 2. Literature Review

Boulemtafes[1] proposed on deep learning in healthcare applications where it has received a lot of interest recently from the scientific community. These are used to analyse large numbers of health records in sophisticated cloud environments. Nevertheless, once private data is delivered to a distant server, concerns about privacy surface. Throughout



this work, we concentrate on ubiquitous wearable health apps that allow remote patients to be anywhere at any time and from any location, including earlier Covid-19 identification, sleeping apnea identification, and the identification of cardiac disorders. Since client-side environments used by ubiquitous healthcare monitoring are typically limited, it is crucial to address such limitations while developing private information alternatives. Thus, the point of the study is to evaluate the suitability of current privacy-preserving deep learning models in the context of widespread monitoring systems.

I. Razzak [2] researched paper on the federated learning has lately come to light as an impressive platform for enabling the distributed development of art deep learning networks involving various types of people while maintaining the privacy of user information. Federated learning has the advantage of enabling all stakeholders to build solid predictions in the absence of enough training data. Most respondents are permitted to maintain their anonymity during the procedure. Online communication has dramatically expanded since then, owing to its mobility and capacity for performing several everyday chores. The sensitive keypad on a cell phone produces sensations while you type that might be exploited to identify actual typed keys, facilitating side-channel assaults. Patient records, health-related data, usernames, and credentials are some examples of this data. In order to safeguard this data, a research paper suggests a unique approach that relies on federated learning to enable side-channel detection attacks. Ten Android smartphones who have been prompted [2] to speak on the device's soft keyboard provided us with a dataset. To enable two clients to train local models, we divide each dataset into two panes of five individuals each. The DNN classifier is

constructed separately on the data using the federated have to learn architecture, which collects modelling changes supplied by two customers. Every customer looks through the results three times in order to minimise the over-fitting issue. The experimental findings demonstrate that the DNN model does have a greater accuracy of 80.09%, demonstrating the effectiveness of the suggested architecture in detecting side-channel assaults.

Vadrevu,[3], suggested to the state-of-the-art in data publishing while respecting privacy is the main topic of this study. Generally, data is divided into categories, including tabular, procedural, set-value, narrative, and networking site data. Both surveillance footage and health information are taken into account in this program. Each document is split into two sections. Firstly, several anonymization techniques for surveillance footage that protect individuals are presented. Furthermore, techniques for publishing health information that protect individual privacy are provided. The benefits and drawbacks of current privacy-preserving approaches in monitoring and ongoing medical information release are eventually examined.

### 3. Methodology

#### **Big Data**

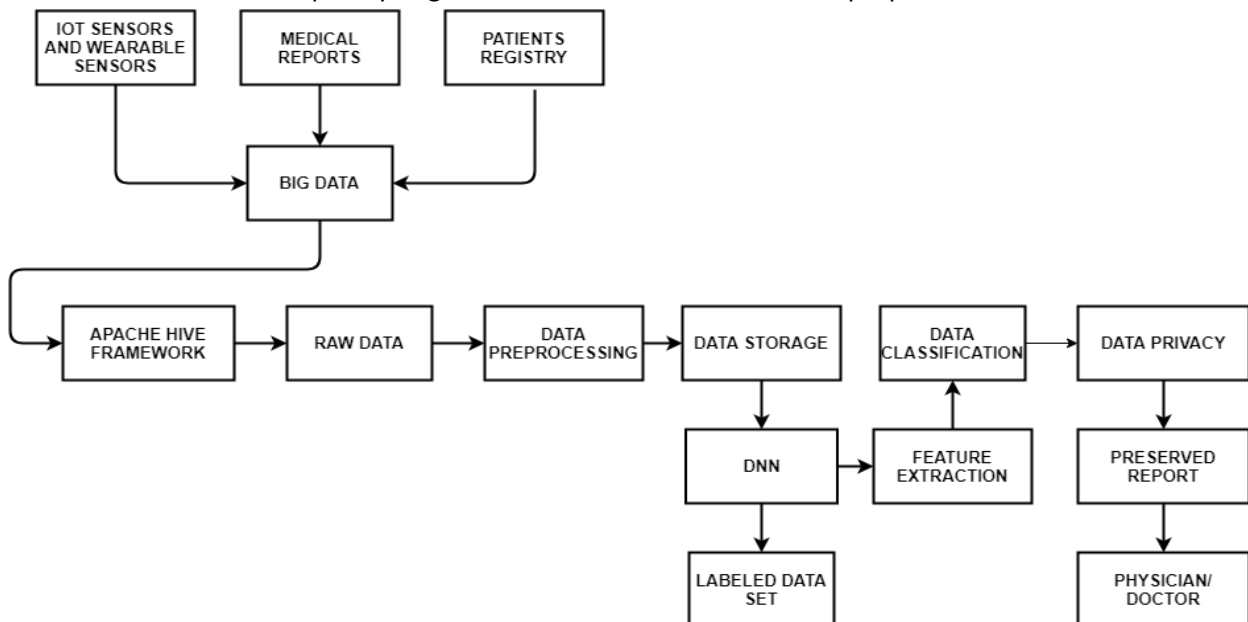
Big data plays an important role in the healthcare applications to improve the clinical research and emphasizes the healthcare of the patients. It analyses the transformation of the health resources which includes the biomedical images, medical reports and genetic data. These data are analyzed and discovered by attaining the patterns. To retrieve the information is by analyzing the useful set of information. It discovers the knowledge with the patterns by the large volumes of the clinical research. To predict the analytics of the outcomes, specific



medications for the patients are determined. Data extraction and collection of the data storage of the health is by preserving the privacy of the reports.

These analytics use the heterogeneous type data which covers the health information, relevant information of the patients. Some of the relevant information is monitored using the IOT devices which gather the raw information. These un-cleansed data are processed and are generated using the Apache hive framework. Apache hive framework generates the data in a sophisticated manner where it checks the data in the generated manner. Deep learning is used to a minimal set of correlated input which reduces the complexity. Big data is used

to allow the physicians to check the patient's records and other critical types of information. Big data checks and analyses of the statistical and mathematical approach to stimulate the data, validate and analyze the forecast of predicting futuristic incidents. It reduces the large datasets by using complex statistical methods. These use different sets of information using the descriptive analysis by collecting the data into a meaningful set of information. In the traditional warehouse of the healthcare sector, the data are maintained in the periodic way where the load of the data gets maximized. In the healthcare sector, the data should be updated up-to-date so that the information can be used later on purpose.



**Figure 2: Architecture Diagram**

**DNN**

Big Data analyses the transformation of the health resources which includes biomedical images, medical reports and genetic data. Big data is used to allow the physicians to check the patient's records and other critical types of information. DNN reduces the large datasets by using complex statistical methods. Prediction over the analysis of the healthcare is used for the future prediction of the patient's health

outcomes within the clinical or non-clinical way of patterns within the data.

Encryption encrypts the data from any intruder or any misleading of the data and the location awareness of the patient data is accessed by tracking the information. Hive used in the healthcare dataset collectively gathers the set of information in a processed manner. Patient's data has to be encrypted and the data has to be preserved in the encrypted manner. DNN is used to correlate the complex input layers into



the correlated set of output. Deep learning models correlates and improves the performance by attaining the accurate set of refined correlated results. In recent years, deep learning and big data have been used in the healthcare industry to understand the reducing complexity and its advantages in reducing the time and cost. These identify the data imaging and diagnosis the disease in the earliest manner by providing the medical professionals with the accurate range of results. It gives the personalized set of care by delivering the early intervention in the patient's health condition.

#### 4. Construction

##### Data processing

In the big data processing, the raw data which are attained from the healthcare organizations are cleansed and processed for the integration purpose. These are used to extract the information which supports the decisions over the programmers. These models give the large set of data which extracts the relevant set of information using the following techniques. Analyzing the data standardization is by integrating the data.

To enhance and enrich the data, data quality is increased to further processing and integration purpose. In the analysis stage, the discovery of the data is prepared by a structured set of platforms. The data is processed using the metadata stage. The processing of the data requires the standardization of the data. This standardization of the data improves the quality over the data.

##### Apache Hive Framework

Apache Hive is a tool which analyses and uses a query method to derive the databases. The file system is integrated by using the Hadoop architecture. This hive-based framework provides the SQL type interface which queries along with the Hadoop defined databases. These queries are transformed into the Hive

platform using the Map reduce API, which tends to be Hive Query. These tend to provide the abstract layer which integrates the SQL queries. These clusters up the nodes and provide the parallel computing processed database. Hive is considered as one of the top structures of the Hadoop. In the hive platform, different techniques are used to provide the migration in the data. Using the SQL scripting, the diverse operations are dynamically attained. Even Apache pig can be used to analyze the data warehouse.

Apache Hive framework is considered as the data warehouse to provide the analysis of the data. This hive platform provides the SQL type queries to implement in java API. Hive queries are used to analyze the datasets using the HDFS methodology. In the healthcare sector, the relational database of the big data of the healthcare is analyzed and monitored. In the big data, the data are analyzed in the SQL based relational database using the Apache hive framework. Hive Server checks up the data of the healthcare and attains the predicted results. It analyses the test data and attains the patient's good and bad health by determining the results. Hive server stores the data permanently and periodically checks the updated source of health status of the p[atient's health. Using the map reduce in the Apache hive framework, the measure value over the spread of the scores within the variable are derived and calculated using as follows,  $\sigma = \sqrt{\sum (Y_i - \mu)^2 / N}$ . The variance and standard deviation are derived to attain the sampled out using the SQL. The square over the population standard deviation is termed as population variance.  $\sigma^2 = \sum (Y_i - \mu)^2 / N$ .

#### 5. Experimental Results

##### DNN Metrics

The deep Neural Approach trains up the data multiple which interferes with the shared resource computing. Grain up the data can



improve the latency period and reduces the bandwidth state. Since the latency gets decreased, the computational gets high. The use of the DNN is to reduce the complexity of the data sets where it optimizes the data. Deep learning is used with the edge devices to raise the computation speed high. The layer of flows is high and gets higher speed. DNN optimizes and the data are computationally high in speed. These allocate the space in the memory along with the processing capability. Some of the embedded devices are based upon the resource constraints. In the different technologies such as IPV6, they tend to give interoperability by giving a large range of flexibility in the datasets. It derives the protocol state which tends to be based upon their network requirements. Resource constrained datasets are handled in the dataset using the DNN. DNN trains the model by minimizing the latency and improves the granularity in the trained set of datasets. This DNN layer creates the subset in the gateway over the devices. In this, DNN partitions up the data which uses ANN to correlate the multiple layers of correlated input and output. DNN can be able to boost up the performance and are used in most of the applications such as in healthcare, industry and defense organizations to store up the datasets. Particular set of neurons are attained in the vector form where DNN analyses the medical images by diagnosing the medical issues and health issues of the patients. The hidden layer of functions can cascade the network and gives a certain capability in their medical diagnosed big data.

#### **Extraction of features**

The extraction of the feature involves the reduction in the datasets computational speed. As the features are utilized to classify the model based upon their classification. In the DNN approach, feature extraction extracts the assigned datasets based upon the number of

occurrences. These reduce the number of assets where it defines the huge set of information. It enhances the performance analysis by detecting the complicated data. It uses the sum of variables using the convolutional theorem. These validate the amount of memory with high computational power in range. It checks the number of entities in the schema for the identifying the size. A feature generates the generic entities for analyzing the pattern type generation. Feature extraction precedes the analysis over the disease risks and their existence prediction. Using the AUC model, the feature which is extracted is attained and predicted. These features represent each in the small form with a small number of features between the raw data and its knowledge. To attain the high level of accurate prediction over the data and to attain the diagnosed medical status, these frameworks are initiated. It targets the risk estimation using the paradigm shift which identifies the medical status of reports.

#### **Data Preservation**

Data privacy is used to preserve the data which is shared and accessed for later purpose. These data have to be secured to avoid from the intruder. A massive amount of data is mishandled by the attacker without the user acknowledging it. To avoid such circumstances, the encryption of the data is handled. The homomorphic encryption is used to encrypt the data where DL training and tested. These improve the privacy protection of the data. There are different computational encryption algorithms to achieve cryptographic in the medical data. This homomorphic encryption can be able to preserve the data in the encrypted form. It is highly flexible and gives unique techniques in the standardized encryption. This homomorphic encryption encrypts the data in the server in the cloud where the practitioner/physician gives the request to access the information in the



encrypted mode. A unique set of secret keys is used to encrypt and decrypt the data. Different mathematical processes are performed to avoid the intruder attack over the medical

data. The following table represents the health parameters which are characterized for a unique id, i.e., the patient's Id.

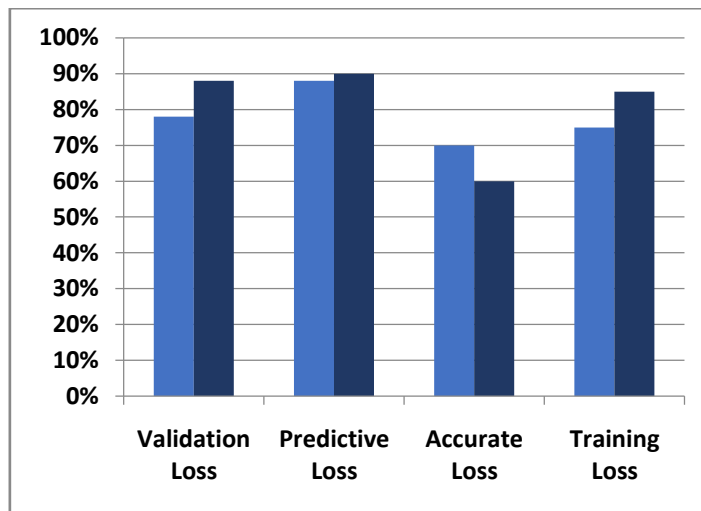
Parameters	Unique Id	Low	Medium	High	Encrypted Uid	Normal/Abnormal
Vital Signs (BP, Body Temperature, Pulse Rate)	3045	Yes	No	No	30**	Normal State
Cardiac Test	1236	No	Yes	No	12**	Normal State
Diseases (Detected)	3455	No	No	Yes	34**	Abnormal State

**Table 1: Parameters in Healthcare**

Dataset	Validation Loss	Predictive Loss	Accurate Loss	Training Loss
Plain Dataset	78%	88%	70%	75%
Encrypted Dataset	88%	90%	60%	85%

**Table 2: Privacy Preservation**

The above table represents the data compared both in plain data and encrypted data. From this, preservation of the data can increase the performance metrics and reduces the losses in the data metrics comparatively.



**Figure 3: Analytical Performance of Using DNN**

## 6. Conclusion

Apache Hive is used to standardize the health-related massive data which are generated. DNN is used to correlate the complex input layers into the correlated set of output. Deep learning models correlate and improve the performance by attaining the accurate set of refined correlated results. The patient's health condition report is diagnosed to analyze the chronic illness or early diagnosis of health condition of patients. These clinical data are gathered and are further used by the doctor professionals to analyze the patient's health condition. Apache

Hive Query analysis and provides the file distributed system and databases. It integrates the hive SQL which transforms the query database and processes it. These frameworks improve to diagnose the patient's health condition and diagnose the chronic illness by preserving the data. Compare to the other electronic health records, big data reduces the cost-effective and time-consuming by improving the health outcomes. Deep learning analysis is used to diagnosis the medical data analysis and for the visualization. Big data specifically analyze



the statistics of the data by reducing the costs and prevention of onset diseases.

## 7. References

1. Boulemtafes, A., Derhab, A. &Challal, Y. Privacy-preserving deep learning for pervasive health monitoring: a study of environment requirements and existing solutions adequacy. *Health Technol.* 12, 285–304 (2022). <https://doi.org/10.1007/s12553-022-00640-3>
2. Rehman, I. Razzak and G. Xu, "Federated Learning for Privacy Preservation of Healthcare Data from Smartphone-based Side-Channel Attacks," in *IEEE Journal of Biomedical and Health Informatics*, doi: 10.1109/JBHI.2022.3171852.
3. Vadrevu, Pavan&Adusumalli, Sri Krishna &Mangalapalli, Vamsikrishna& Swain, Sangram. (2021). A Review on Privacy Preservation Techniques in Surveillance and Health Care Data Publication.
4. Singh, Shreyansh&Shukla, K. (2021). Privacy-preserving Machine Learning for Medical Image Classification.
5. D. -E. -M. Nisar, R. Amin, N. -U. -H. Shah, M. A. A. Ghamdi, S. H. Almotiri and M. Alruily, "Healthcare Techniques Through Deep Learning: Issues, Challenges and Opportunities," in *IEEE Access*, vol. 9, pp. 98523-98541, 2021, doi: 10.1109/ACCESS.2021.3095312.
6. Alotaibi, Shoayee&Mehmood, Rashid &Katib, Iyad&Rana, Omer &Albeshri, Aiiad. (2020). Sehaa: A Big Data Analytics Tool for Healthcare Symptoms and Diseases Detection Using Twitter, Apache Spark, and Machine Learning. *Applied Sciences*. 10. 1398. 10.3390/app10041398.
7. Baig, Hidayath. (2020). Privacy-Preserving in Big Data Analytics: State of the Art.
8. Abouelmehdi, K., Beni-Hessane, A. &Khaloufi, H. Big healthcare data: preserving security and privacy. *J Big Data* 5, 1 (2018). <https://doi.org/10.1186/s40537-017-0110-7>
9. J. Shi, X. Fan, J. Wu, J. Chen and W. Chen, "DeepDiagnosis: DNN-Based Diagnosis Prediction from Pediatric Big Healthcare Data," 2018 Sixth International Conference on Advanced Cloud and Big Data (CBD), 2018, pp. 287-292, doi: 10.1109/CBD.2018.00058.
10. Zhang, Chuan& Zhu, Liehuang& Xu, Chang & Lu, Rongxing. (2017). PPDP: An efficient and privacy-preserving disease prediction scheme in cloud-based e-Healthcare system. *Future Generation Computer Systems*. 79. 10.1016/j.future.2017.09.002.
11. Ravi, Daniele & Wong, Charence&Deligianni, Fani& Berthelot, Melissa & Andreu-Perez, Javier & Lo, Benny & Yang, Guang-Zhong. (2016). Deep Learning for Health Informatics. *IEEE journal of biomedical and health informatics*. PP. 10.1109/JBHI.2016.2636665.

