International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Healthcare Prediction Based on Big Data Management in Industrial IoT Using Optimized Deep Convolutional Neural Network

A.N.Satyanarayana[1], S.P.V. Subba Rao[2], L.V.R Chaitanya Prasad[3], Shruti Bhargava Choubey[4], Abhishek Choubey[5]

Submitted: 16/01/2024 **Revised**: 24/02/2024 **Accepted**: 02/03/2024

Abstract: Presently, Heart Disease (HD) stands as the foremost global cause of mortality, and the anticipation of cardiovascular ailments demands sophisticated expertise and experience. In recent times, healthcare establishments have turned to Internet of Things (IoT) technology to amass sensor data for detecting and prognosticating heart disease. The rapid generation of big data analysis is a formidable task of gathering and analysing large volumes pose a challenge to prompt action and uncovering latent value during critical situations. However, disease prediction remains challenging due to feature dimension. So the problem is non related feature analysis leads poor accuracy in precision and recall rate. To address this issue, we introduced an Optimized Deep Convolutional Neural Network (ODCNN) for accurate heart disease prediction. Furthermore, accuracy can be achieved by utilizing a dataset on heart disease obtained from Kaggle. Moreover, we implemented a Decision Tree (DT) method to estimate the impact ratio of HD prediction. Similarly, the Decision Function-Based Chaotic Salp Swarm (DFCSS) algorithms is used to select features based on their ranking to achieve an optimal feature set. Finally, the ODCNN based classifier is used to predict the heart disease more accurately. Furthermore, the proposed framework can illustrate precision, recall, true positive rate, and F-measure as performance evaluation parameters. The simulation results indicate that our approach attains a classification accuracy of 94.47% on the heart disease dataset.

Keywords: *Big data analysis, HD, industrial IoTs, ODCNN, DFCSS, and healthcare prediction.*

1. Introduction

Heart disease is one of the leading causes of death worldwide and a major socioeconomic problem. Moreover, well-defined types of HD include arterial, cerebrovascular disease, radial artery disease, and fetal. Further, an estimated 17.9 million people worldwide die each year from HD and its consequences, accounting for more than 80% of deaths from HD and stroke. Similarly, heart attacks and strokes occur when the blood supply to the heart is blocked [1].

Moreover, these health disorders impact individuals of all ages due to lifestyle changes and are the primary cause of various symptoms. As well, wearable sensors in computers are utilized to gather vital signs. Furthermore, they can be integrated with data from clinical databases to facilitate practical analysis and prediction. Consequently, the healthcare industry has presented that IoT and Machine Learning (ML) are highly relevant [2]. Similarly, big data analytics play an essential role in predicting future health conditions, which can lead to better health outcomes for people. ML techniques can be manipulated to find the best outcomes through multiple predictive analyses. Likewise, Big Data Analytics predicts future health based on health parameters and provides optimal results [3].

However, medical records have become one of the most essential means of retrieving relevant patient health records in medical servers and databases. These address the potential privacy and security issues of IoT. Therefore, due to the complexity of HD, it is necessary to use drugs with caution. Furthermore, transitioning to these medications may jeopardize the heart or result in early death. Nevertheless, it had little information regarding the goals or specifications for utilizing IoT [4].

Nevertheless, severe HD can also impair heart function and result in consequences like blood vessel dysfunction and coronary artery infections. However, these automated methods can be laborious because they can't analyze vast volumes of data or require large amounts of data. Age, smoking, high blood pressure, cholesterol, low blood pressure, and poor hygiene are risk factors for heart disease [5].

The contributions in this section propose that ODCNN provides accurate predictions for heart disease patients by processing cardiac disease datasets. Afterward, the DT method is suggested for estimating the cardiovascular disease prediction influence ratio. Further, the DFCSS algorithm can be employed for selecting optimal features and attaining the best feature set based on ranking. By leveraging big data management and industrial IoT architecture, ODCNN can be used for accurate assessment of heart disease.

[1,2,3,4,5]*Department of Electronics and Communication Engineering, Sreenidhi Institute of Science & Technology, Hyderabad-501301, Telangana, India,*
[1]*Email:satyanarayana.a@sreenidhi.edu.in* ,
[2]*Email:spvsubbarao@sreenidhi.edu.in,*
[3]*Email:lvrchaitanya@sreenidhi.edu.in*
,[4]*Email:shrutibhargava@sreenidhi.edu.in*
,[5]*Email:abhishek@sreenidhi.edu.in*

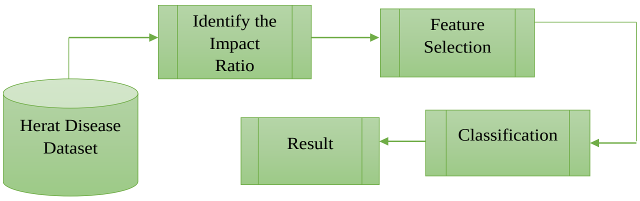

Fig. 1. The Architecture Diagram for Big Data Management Analysis Based on Heart Disease Prediction

Figure 1 describes the big data management analysis and architecture diagram of IOT for predicting HD impact rate, feature selection, and classification accuracy based on heart disease prediction. The Optimized Deep Convolutional Neural Network (ODCNN) model. To train and evaluate the ODCNN, to utilize a comprehensive heart disease dataset obtained from Kaggle. Additionally, we employ a Decision Tree (DT) method to assess the impact ratio of heart disease prediction. Furthermore, we incorporate the Decision Function-Based Chaotic Salp Swarm (DFCSS) algorithm for feature selection, which ranks features based on their importance and selects an optimal feature set. The selected features are then fed into the ODCNN classifier for heart disease prediction. The performance of the proposed framework is evaluated using various metrics, including precision, recall, true positive rate, and F-measure. Extensive simulation results demonstrate the effectiveness of our approach, achieving a remarkable classification accuracy of 94.47% on the heart disease dataset.

2. Literature Survey

IoT-driven Big Data Analytics (BDA) facilities and Demand-Side Management (DSM) engines can collaborate smartly with consumers in trusted businesses [6].

According to that, smooth edge L1 normalized Support Vector Machine (SVM) classifiers are designed to choose salient features and manage extensive dimensional data. However, ensuring protection is crucial to encourage participation and adherence to data privacy [7].Therefore, they suggest DL-based frameworks and improved systems to predict HD. Similarly, selective Modified Deep Long-Short-Term Memory (MDLSTM) can identify normal and abnormal cardiac diagnostic features [8].

Furthermore, the impact evaluation programs that use EHR-based ML algorithms can use advanced SVM classifiers to enhance parameter tuning and performance accuracy [9]. Accordingly, surface Data Mining (DM) procedures can be used to identify IoT devices in the healthcare industry. Further, the big data technology of DM appears from the perspective of processing large volumes of complex data [10].

Table 1. Healthcare Big Data Analysis in IoT

Author/Ref.No	Year	Methods	Drawback
A A, Dahan [11]	2023	Cuckoo Search Algorithm (CSA)	However, there are various difficulties in rapid monitoring and diagnosis when using these methods.
Sulagna Mohapatra [12]	2024	Artificial Neural Network (ANN)	Manually analyzing massive and heterogeneous patient data is complex and error-prone.
Khan M. F [13]	2021	SVM	However, technology is essential for existing healthcare systems to provide real-time patient data and transform patient care.
Ma W [14]	2023	Big Data-Particle Swarm Optimization (BD-PSO)	However, managing big data effectively and capably is becoming increasingly difficult.
Ed-daoudy [15]	2019	Decision Tree (DT)	The rapid rate of data generation makes collecting and analyzing large amounts of data difficult.

As described in Table 1, the authors clearly explain the techniques and pitfalls that arise when using big data management for healthcare analysis in IoT.

Accordingly, a big data prediction analysis, microwave antenna design technology, and microwave communication for IoTs can be combined in various aspects. Furthermore, the function of all elements of IoT in DM, communication, and networking are determined [16].

Similarly, based on their resolution IIoT sensors can use ML to manipulate detailed data and deploy healthcare systems. Furthermore, a parallel program designed for controlled real-time and efficient processing can be distributed [17].

Accordingly, smart environments with large-scale and small-scale IoT applications in distributed efficient processing can be developed and evaluated by manipulating various DM techniques. An overview of large systems in IoT environments can be presented to emphasize the importance of DM [18].

According to their emphasis, database technologies such as storage, structure, management, and application processing methods, including DM technologies, can be used to analyze clinical data [19].

Moreover, mentioned that analyzing scientific literature maps trends in the healthcare industry within the IoT-BDA paradigm. Afterward, the IoT-BDA paradigm is required for organizing, impacting, and implementing IoT-based designs in healthcare services [20]

Table 2. Prior Heart Disease (HD) Prediction Using Big Data Analysis

Ref. No	Year	Technique	Method	Achieved Result
21	2023	Deep learning (DL)	Levy Flight – Convolutional Neural Network (LV-CNN)	91%
22	2023	DL	Squirrel Search-Optimized Gradient Boosted Decision Tree (SS-GBDT)	93%
23	2022	DL	Deep Graph Convolutional Network (DG-ConvoNet)	89%
24	2020	ML	Deep Neural Networks (DNN)	90%
25	2023	ML	K-Nearest Neighbour,	91.26%

Table 2 illustrates the techniques, methods, and precision results used for predicting heart disease through big data analysis [34].

It is possible to integrate clinics, patients, and investors to shift from clinic-focused care to healthcare that is focused on the patient. Inappropriately, due to insufficient and ineffective healthcare services, managing the healthcare needs of a growing population with chronic illnesses is becoming more and more difficult [26].

According to that, it provided a deterministic Internet of Things framework to manage healthcare [35] in an increasingly complex manner, which can evaluate HD data with a specially Modified Deep Neural Network (ODCNN). The acquired sensor data can be classified as normal or pathological using the ODCNN approach [27].

The Chaotic Biogeography-Based Optimisation Information Entropy (CBO-IE) method is a data clustering utilized for organizing healthcare IoT datasets[36]. By leveraging data entropy, the CBO-IE method effectively and accurately distributes data points within a dataset, leading to improved clustering outcomes [28].

Similarly, monitoring medication therapy in cardiac patients using Deep Learning-Modified Neural Networks (DLMNN) can offer insights into IoT-centric diagnosis. However, diagnosing HD is challenging due to its diverse symptoms and characteristics [29].

After analyzing their resolutions, it is possible to predict HD and prevent its occurrence based on current HD data evaluation. However, it causes various problems such as cholesterol, diabetes, heart and other organ damage [30].

3. Proposed Methodology

The content of this section discusses the application of ODCNN for accurate prediction of heart disease in patients. By processing heart disease datasets, ODCNN can provide accurate and reliable predictions for individuals with heart disease. Furthermore, the DT method is recommended to estimate the impact ratio of HD prediction. They enable big data management to understand the factors that contribute most to the prognosis of heart disease. After that, the DFCSS algorithm selects the optimal features and finds the best feature set based on the ranking. These are also the most relevant factors to improve the accuracy of predictions. Thus, ODCNN can be effectively used for accurate prediction and management of HD, leading to better patient outcomes and improved healthcare management. The proposed framework for heart disease prediction using the ODCNN model consists of the following steps:

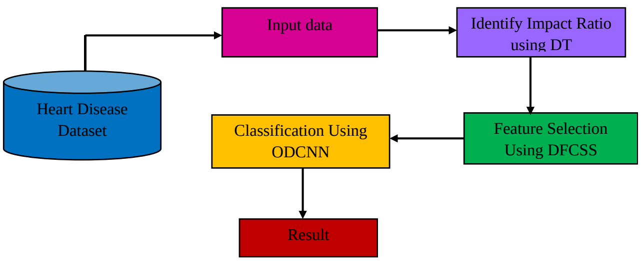

Fig. 2. The Proposed Architecture Diagram for ODCNN

The design's proposed architecture diagram can be assessed using techniques like DD, DFCSS, and ODCNN as shown in Figure 2. Further, these techniques may facilitate understanding big data management to classify the most significant factors in predicting heart disease prediction and provide accuracy. However, the application of DCNNs in heart disease prediction using non-image data, such as clinical records, has not been extensively studied.

1. **Data Preprocessing:** The heart disease dataset is preprocessed to handle missing values, outliers, and data normalization.
2. **Feature Selection:** The Decision Function-Based Chaotic Salp Swarm (DFCSS) algorithm is employed for feature selection. The DFCSS algorithm ranks features based on their importance and selects an optimal feature set.
3. **ODCNN Model Architecture:** The ODCNN model is constructed using a series of convolutional layers, pooling layers, and fully connected layers. The convolutional layers extract features from the input data, while the pooling layers reduce the dimensionality of the feature maps. The fully connected layers are used for classification.
4. **Training and Optimization:** The ODCNN model is trained using the Adam optimizer and the cross-entropy loss function. The model is trained on the training set and evaluated on the validation set.
5. **Performance Evaluation:** The performance of the ODCNN model is evaluated using various metrics, including precision, recall, true positive rate, and F-measure

Deep Convolutional Neural Networks (DCNNs) have achieved state-of-the-art results in various image recognition tasks. Their ability to extract hierarchical features from images makes them suitable for medical image analysis, including heart disease diagnosis

3.1 Dataset

Different mathematical analyses, statistical variables, and multivariate numerical data are recommended for handling HD prediction in this context. Furthermore, the database comprises 76 features, utilizing a subset of 14 of the characteristics published in all studies. The primary purpose of this HD dataset is to predict whether an individual has HD based on their characteristics and to provide a diagnostic task.

Table 3. Heart Disease Prediction Dataset

id	age	sex	dataset	cp	trestbps	chol	fbs	restecg	thalch	exang	oldpeak	slope	ca	thal	num
1	63	Male	Cleveland	typical an	145	233	TRUE	lv hypertr	150	FALSE	2.3	downslop	0	fixed defe	0
2	67	Male	Cleveland	asympton	160	286	FALSE	lv hypertr	108	TRUE	1.5	flat	3	normal	2
3	67	Male	Cleveland	asympton	120	229	FALSE	lv hypertr	129	TRUE	2.6	flat	2	reversabl	1
4	37	Male	Cleveland	non-angin	130	250	FALSE	normal	187	FALSE	3.5	downslop	0	normal	0
5	41	Female	Cleveland	atypical ar	130	204	FALSE	lv hypertr	172	FALSE	1.4	upsloping	0	normal	0
6	56	Male	Cleveland	atypical ar	120	236	FALSE	normal	178	FALSE	0.8	upsloping	0	normal	0
7	62	Female	Cleveland	asympton	140	268	FALSE	lv hypertr	160	FALSE	3.6	downslop	2	normal	3
8	57	Female	Cleveland	asympton	120	354	FALSE	normal	163	TRUE	0.6	upsloping	0	normal	0
9	63	Male	Cleveland	asympton	130	254	FALSE	lv hypertr	147	FALSE	1.4	flat	1	reversabl	2
10	53	Male	Cleveland	asympton	140	203	TRUE	lv hypertr	155	TRUE	3.1	downslop	0	reversabl	1
11	57	Male	Cleveland	asympton	140	192	FALSE	normal	148	FALSE	0.4	flat	0	fixed defe	0
12	56	Female	Cleveland	atypical ar	140	294	FALSE	lv hypertr	153	FALSE	1.3	flat	0	normal	0
13	56	Male	Cleveland	non-angin	130	256	TRUE	lv hypertr	142	TRUE	0.6	flat	1	fixed defe	2
14	44	Male	Cleveland	atypical ar	120	263	FALSE	normal	173	FALSE	0	upsloping	0	reversabl	0
15	52	Male	Cleveland	non-angin	172	199	TRUE	normal	162	FALSE	0.5	upsloping	0	reversabl	0
16	57	Male	Cleveland	non-angin	150	168	FALSE	normal	174	FALSE	1.6	upsloping	0	normal	0
17	48	Male	Cleveland	atypical ar	110	229	FALSE	normal	168	FALSE	1	downslop	0	reversabl	1
18	54	Male	Cleveland	asympton	140	239	FALSE	normal	160	FALSE	1.2	upsloping	0	normal	0
19	48	Female	Cleveland	non-angin	130	275	FALSE	normal	139	FALSE	0.2	upsloping	0	normal	0
20	49	Male	Cleveland	atypical ar	130	266	FALSE	normal	171	FALSE	0.6	upsloping	0	normal	0
21	64	Male	Cleveland	typical an	110	211	FALSE	lv hypertr	144	TRUE	1.8	flat	0	normal	0
22	58	Female	Cleveland	typical an	150	283	TRUE	lv hypertr	162	FALSE	1	upsloping	0	normal	0

In Table 3, the HD prognostic dataset aids in predicting and diagnosing HD prognosis. It encompasses features such as CP-chest pain type, chol-cholesterol in mg/dl, FBS-fasting blood sugar, thal-maximal heart rate achieved, ca-number of great vessels stained by fluoroscopy (0-3), and thal-normal.

3.2 Decision Tree (DT)

This section uses the DT algorithm to assess the HD impact rate of evaluation parameters such as performance and processing time. Similarly, the binary split-initiated partitioning process continues until no further split is possible. Further, the DT algorithm divides each node of the recursive partition with or without a hierarchical generation process. Their separation is determined by parameters such as gini impurity and entropy. Likewise, a consistent measure of node-level label measures the impact ratio for HD based on node purity. Moreover, a tree-structured DT algorithm can be created to determine the characteristics of the internal node dataset. The tree branches can generate the development of each leaf node by defining decision rules.

Algorithm 1. DT

Input: $Training_{dataset}$W, attribute R

Output: $Decision_{tree}(W)$

Start

If W isN^{ull}, then

Re_{turn} Defeat

End if

If R is N then

$Re_{turn}\ R^h(W)$

End if

If $|R| = 1$

Return $W_{S_n}(R)$

End if

Step 1: Compute the set tree {}

For each $x \in R$ do

$$Se_p(x, W) \& Sp_p(x, W) = 0 \quad (1)$$

Step 2: Compute the entropy

For each $\mathbb{v} \in Va_{lues}(x, W)$ do

Step 3: Compute the set as a subset tree attribute.

$$I^o(x, w) += \frac{w_{x,u}}{w_x} C_{(x_u)} \quad (2)$$

$$Sp_p(x, W) += \frac{w_{x,u}}{W_x} log \frac{w_{x,u}}{w_x} \quad (3)$$

End for each

$$m_{(x,w)} = C(x) - I^o(x, w) \quad (4)$$

$$IM_S(x, w) = \frac{Im_S(x, w)}{Sp_{p(x,w)}} \quad (5)$$

End for each

Step 4: Established $X_{best} = x_m(IM_S(x, w))$

Step 5: Assign $A_{best}(T)$

For each $\mathbb{v} \in Val_{ues}(X_{best}, W)$ do

End for each

Return W

End

The impact rate of node contamination can be estimated by assessing the integrity of node-level labels. Let's assume N-null value, Se_p −set information, W_{S_n} −single node, Sp_p −split information, I^o −information, M −gain, IM_S −impact gain ratio, $C(x)$ −entropy, R-entropy of attribute, q-class label, e-number of class, x-entropy, W-set tree, u-value.

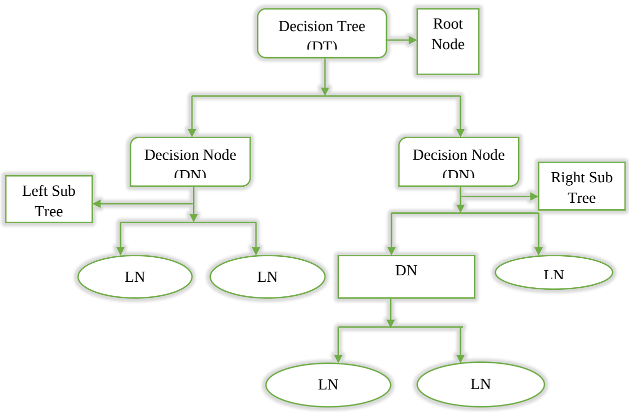

Fig. 3. Flow Chart for Decision Tree

Figure 3 shows that this DT model can be utilized to organize data and constructed using a recursive section. Similarly, dataset members can be split into subgroups based on some element, and these split subgroups are further subdivided until a certain depth is achieved.

3.3 Decision Function-Based Chaotic Salp Swarm (DFCSS)

In this section, the best feature range can be selected using the DFCSS algorithm based on the feature selection process. Moreover, the learning process can be accelerated and made more effective by utilizing the DFCSS method to identify and eliminate numerous irrelevant features. Similarly, the chaos map used in the DFCSS method can reduce the dimensionality of the data and accommodate random and chaotic variables. Furthermore, utilizing the DFCSS method for feature selection helps specify the swarm behavior of program layers and estimate appropriate weighting factors.

Improve the position of the leader of the remaining salps, as depicted in Equation 6. Let's assume b_p^1 −leader position, p-dimension, v_{y^p} &O_{y^p} −lower and upper boundary value, d_p −denote food position, s_1, s_2, s_3 −random number.

$$b_p^1 = W \begin{cases} d_p + s_1([v_{y^p} - O_{y^p}]s^2 + o_{y^p}) & S_3 \geq 0 \\ d_p - S_1([v_{y^p} - \mathbb{O}_{y^p}]s^2 + o_{y^p}) & S_3 < 0 \end{cases} \quad (6)$$

The remaining salps move ahead of the leader chain and become updated followers. Further, the balance between mining and exploration can be estimated by optimizing the remaining salps to select the leader's position, as shown in Equation 7. Where w-present iteration, W-extreme amount of iteration, l-time, α and β_0 −initial speed denoted,

$$\begin{cases} S_1 = 2c^{-(4w/W)^2} \\ B_p^q = \frac{1}{2}\alpha O^2 + \beta_0(q) \\ \mathcal{B}_p^q = \frac{1}{2}(b_p^q + \mathbb{B}_{\mathbb{P}}^{q-1}) \end{cases} \quad (7)$$

Equation 18 shows the coefficient improved chaos map estimate. R1 decreases linearly over iterations, and r3 determines if the next level should go to negative infinity or positive infinity. The r1 and r2 are crucial for enhancing salp posture. These advantages can enhance the effectiveness of swarm algorithms.

$$s^2 = \Psi_2 \quad (8)$$

The updated salp level is estimated from the chaotic map shown in Equation 9. Furthermore, it can also be predicted that if the variable is 1, select the equivalent feature, and if the value is equal to 0, there is no set function. Each solution has different properties and lengths, where B is assumed to be a number between 0 and 1. Where Ψ^w −choatic map, B-slap.

$$B_1(p) = \begin{cases} d_p + s_1([v_{y^p} - O_{y^p}]\,\Psi^t + o_{y^p}) & \Psi^w \geq 0 \\ F_i - r_1([v_{y^p} - o_{y^p}]\,\Psi^t + o_{y^p}) & \Psi^w < 0 \end{cases} \quad (9)$$

Equation 12 evaluates each agent's move from one continuous, binary location to another. Let's assume a Y-arbitrary number, R-solution.

$$B_p^w = \begin{cases} 1 & if\ (R(b_p^s) \geq y) \\ 0 & Otherwise \end{cases} \quad (10)$$

The solution with the lowest fitness value is chosen as the best solution. Optimum feature selection enhances accuracy and minimizes the necessary features. The best solution is determined by ranking these features to create an ideal feature set and selecting the fitness value. The fitness function in Equation 11 can also calculate all solutions and feature weighting factors. Let's assume the D_d −fitness function, $\mathcal{T}_d$ −weight factor, L-length, A_{cc} −feature accuracy, l_t −total number of features, l_f −select feature length, T-weight factor, y-bias value, h-number of features.

$$D_d(a) = \left(\frac{m_{ax}\left(\mathcal{T}_d \times \mathcal{A}_{cc} + (1 - T_d) \times \left(1 - \frac{o_d}{O_d}\right)\right)}{\sum_{p=1}^{h} T_p a_p + y} \right) \quad (11)$$

The model salp position is chosen and updated once each salp's fitness has been evaluated. This procedure is iterated until the best possible resolution is achieved.

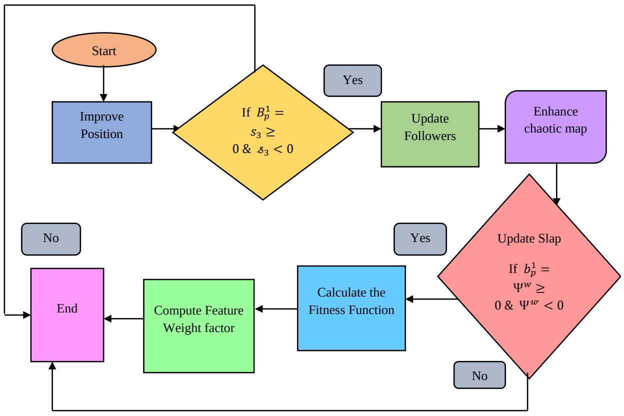

Fig. 4. Flow Chart Diagram for DFCSS

As illustrated in Figure 4, the flowchart diagram of DFCSS allows optimizing the followers by updating the activation state and calculating the fitness function to estimate the weighting coefficients of features

3.4 Optimized Deep Convolutional Neural Network (ODCNN)

In this context, the proposed ODCNN in IoT architecture could be utilized to assess HD accurately. Each chosen feature could also be inputted into the ODCNN classifier for processing. The ODCNN method can be implemented to assign approximate values to the weights associated per input. The weight vectors of all linked input nodes can be multiplied by the input value to determine the basis for adding nodes to the hidden layer. Furthermore, the backpropagation process can be enhanced at the base of the ODCNN method to obtain the results of random weight values. Activation functions are used to transform the output of this layer to their specified states.

The assortment of the input can be estimated after choosing an equal weighting of the eigenvalues, as shown in Equation 12. Furthermore, the activation function can be evaluated with weight factors. Let's assume, the D-Input value, T-weight value, p-value, h-chosen features weight, and G-summed value. E_p −Specifies the exponential value, x_s^p −activation function,

$$d(a) = \begin{cases} DT_p = \{D_{t^1}, \mathcal{D}_{t^2}, \mathbb{D}_{t^3}, \ldots, \mathfrak{D}_{t^h}\} \\ G = \sum_{p-1}^{h} D^p T^{\mathscr{p}} \\ \begin{cases} x_s^p = E_p\left(\sum_{p=1}^{h} D^{\mathcal{P}} T^{\mathscr{p}}\right) \\ \mathcal{E}_{\mathcal{P}} = \mathcal{C}^{-\mathcal{D}_{\mathscr{p}}^2} \end{cases} \end{cases} \quad (12)$$

As stated in Equations 13 and 14, the output layer returns the value of the output layer neuron obtained by summing the weights of all input signals and calculating the value of the output unit to determine the value of the output hidden layer neuron. Where S_p −output unit, $L^{\mathcal{P}}$ −output layer, $\mathcal{T}^{\mathrm{p}}$ −weight of hidden layer, y^p −bias value, T^p −imout weight, b_p −hidden layer, C^s −error signal, F_p −output target.

$$D = \begin{cases} b_p = y^{\mathscr{p}} + \sum \mathcal{E}_p \mathcal{T}^{\mathbb{P}} \\ S_p = y_p + \sum L^{\mathcal{P}} \mathcal{T}^{\mathrm{p}} \end{cases} \quad (13)$$

$$C^s = F_p - s_p \quad (14)$$

The output unit and target values are compared to calculate the relative error. Using these errors, the values can be computed to send output errors to all other units in the network. The weight correction can be achieved through the backpropagation algorithm described in Equations 15 and 16. Where δ_p −error value, α -denotes momentum, T_{e^p} −weight correlation, δ_p −error network.

$$\delta_p = [\mathcal{C}^s(\mathcal{d}[\mathcal{S}_{\mathscr{p}}])] \quad (15)$$

$$T_{e^p} = \alpha \delta_p [D_{\mathcal{P}}] \quad (16)$$

Similarly, it transmits output errors to all other network units. Moreover, the activation function can efficiently transfer the weight vector from the input node to the output layer.

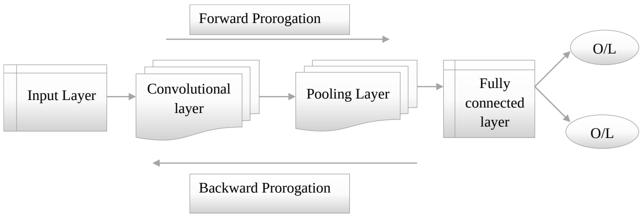

Fig. 5. The Architecture Diagram for DCNN

The input layer values are transferred to the edges of the hidden layer, where they are then multiplied by weights and biases, as illustrated in Figure 5. Furthermore, these values are propagated to the output layer.

4. Result and Discussion

The performance analysis of the algorithm can be evaluated using different data sets and compared with the results obtained using existing optimization algorithms. Furthermore, by analyzing these with the proposed ODCNN methods, their implementation can be tested and trained on standard datasets obtained from Kaggle. Moreover, https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data/code can further test and train this database to learn more about the HD patient dataset. Similarly, the proposed ODCNN method can be effectively predicted by measures such as true positive rate, precision, accuracy, recall, and F1, and various metrics can be selected to demonstrate the improvement of the model.

Table 4. Simulation Parameter

Simulation	**Variable**
Dataset Name	UCI Heart Disease data
No of Records	921
Tool	jupyter
Language	Python
Training	753
Testing	168

The simulation parameters can be utilized to construct an HD prediction model using the Python language in a Jupyter notebook and assess its accuracy with training and test data from the datasets outlined in Table 4.

a. True Positive Rate (TPR)

The TPR can be analyzed and assessed based on the percentage of accurate predictions for the trust class.

$$\text{TPR} = \frac{T_P(R)}{T_P(R) + F_N(R)} \quad (17)$$

b. Accuracy

Compute the accuracy of event categorization by counting the pertinent negative events and dividing them by the entire number of instances.

$$\text{Accuracy} = \frac{(TP) + (TN)}{(\mathcal{TP}) + (\mathcal{TN}) + (\mathcal{FP}) + (\mathcal{FN})} \quad (18)$$

c. Precision

The high accuracy of the application can be achieved to suggest the value of the required data relative to the incorrect data.

$$\text{Pre}=\frac{T_P^R}{T_P^R+F_P^R} \tag{19}$$

d. Recall

Several related profiles involved in cases are being recalled because of their recovery.

$$\text{Rec}=\frac{T^P(R)}{T^P(R)+F^N(R)} \tag{20}$$

e. F-Measure

The F-measure indicates the test's accuracy classification problem. Furthermore, the optimum F-measure can be obtained by performing the highest precision and recall.

$$\text{F-M}=\frac{2\Delta P_{re}+\Delta R_{ec}}{P_{re}+R_{ec}} \tag{21}$$

Probability of Misclassification Error (PME)

PME should be used if the classifier is not accurately predicted.

$$P_{ME} = \frac{F_p+F_N}{T_p+T_N+F_p+F_N} \tag{22}$$

Disease prevalence (DP)

DP can demonstrate the probability that a person has a disease before clinical examination.

$$\mathcal{D}_{\mathcal{P}} = \frac{x^q+y^m}{x^q+y^m+x^m+y^n} \tag{23}$$

Root-Mean-Square Error (RMSE)

The difference between the true values can be predicted based on the square root, which is RMSE.

$$R_{mse} = \sqrt{\frac{1}{p}\sum_{a=1}^{p}(p_i - a_i)^2} \tag{24}$$

Negative Predictive Value (NPV)

The NPV is the predictive value of a negative test result.

$$N_{PV} = \frac{\mathcal{T}^n}{\mathcal{T}_N+\mathcal{P}_N} \tag{25}$$

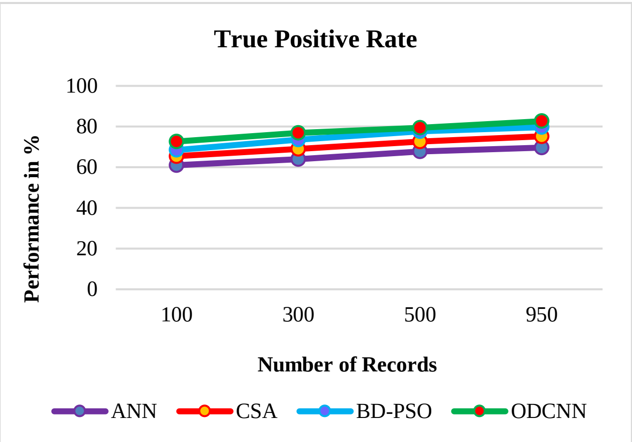

Fig. 6. Analysis of the True Positive Rate

In Figure 6, health data for HD prediction from industrial IoT can be analyzed in big data management to estimate the TPR and determine their average value. The approximation of the TPR using three approaches - BD-PSO, ANN, and CSA - derived from the literature, is as low as 61%. However, the proposed method for big data management has increased the estimate of the TPR to 82.64% compared to existing methods.

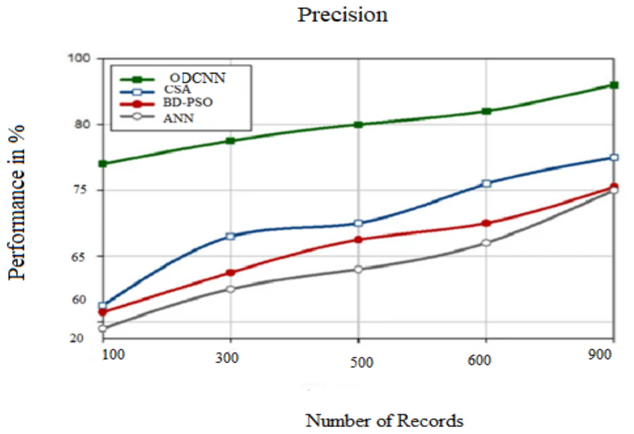

Fig. 7. Analysis of the Precision

As shown in Figure 7, Industrial IoT medical data can be assessed through big data management, analyzing the accuracy and finding the value. Therefore, the data accuracy of approximating HD patients using three methods (CSO, BD-PSO, and ANN) calculated in the literature is 65.4% lower. However, the proposed big data management method improves the accuracy estimation by 84.46% compared to existing methods. An accuracy estimate can be calculated using the entire false profile to determine the percentage of positive values. Furthermore, the application is more effective at representing accurate data than inaccurate data.

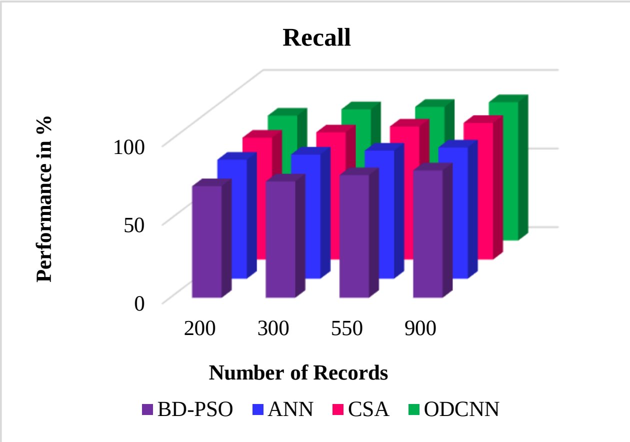

Fig. 8. Analysis of the Recall

Figure 8 illustrates HD data prediction can be determined based on recall analysis by evaluating the health data of industrial IoT in big data management. After that, the recall estimate is as low as 71% when using three approaches such as ANN, CSA, and BD-PSO, which are formalized from the literature to obtain an evaluation of their data for testing and training. However, the proposed method for big data management improved the recall rating to 87.97% compared to the existing methods. Furthermore, the accuracy of the recovery rate can be determined by estimating the number of related events among them. Similarly, the difference in TP and FN rates in identifying negative events can measure the accuracy.

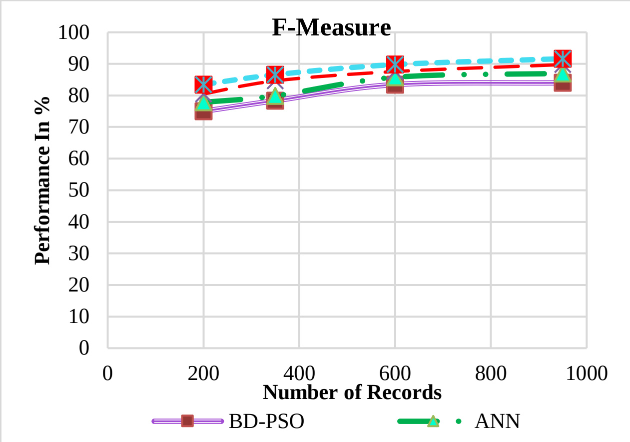

Fig. 9. Analysis of the F-Measure

In Figure 9, the F-measure analysis evaluates health care data of industrial IoT in big data management. Similarly, ANN, CSA, and BD-PSO in HD data prediction using three approaches from the literature evaluated the test and training data with an F-measure of less than 75%. However, the proposed big data management method improved the F-measurement estimate to 91.64% compared to existing methods. Obtaining high precision and recall values can lead to improved accuracy of the F-measure. Moreover, enhances the ability to extract important information from features and provides a more precise representation of computational efficiency.

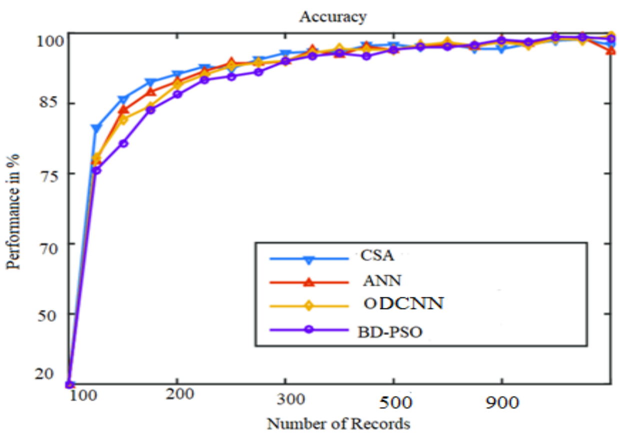

Fig. 10. Analysis of the Accuracy

Figure 10 illustrates that the accuracy of prediction for HD patients can be evaluated in industrial IoT healthcare data analyzed in big data management. Moreover, a systematic approach utilizing three methods - CSA, BD-PSO, and ANN - drawn from the literature can be used to evaluate the drop rate, which has been determined to be as low as 92%. Among these, the proposed method has achieved an increased accuracy of 94.47% compared to existing methods. Efficient classification algorithms can be used to evaluate a classification model and predict its accuracy value.

5. Conclusion

In that sense, HD is an incurable disease, and wearable technology can be used effectively in healthcare. Furthermore, timely intervention monitoring and prediction systems can assist in saving a great deal of lives. These are often most valuable when patients are in remote or underserved areas. In this regard, initially, we collected heart disease data from Kaggle to predict heart disease. Furthermore, we have introduced the DT method to estimate the impact rate of HD prediction. Moreover, we utilized the DFCSS algorithm to select features based on their ranking, resulting in an optimal set of features for the analysis. We proposed using ODCNN in conjunction with big data management and an industrial IoT framework to further enhance the prediction's accuracy. This approach allowed us to assess heart disease with greater precision, resulting in a significant improvement in accuracy by 94.47% compared to existing methods. Furthermore, we employed efficient classification algorithms to predict their accuracy values, ultimately enhancing the overall predictive capabilities of the approach.

Reference

[1] Kareemulla Shaik, Janjhyam Venkata Naga Ramesh, "Big Data Analytics Framework Using Squirrel Search Optimized Gradient Boosted Decision Tree for Heart Disease Diagnosis," Appl. Sci. 2023, 13(9), 5236; https://doi.org/10.3390/app13095236.

[2] Henry Okemiri, Alo Uzoma, "Internet of Things-based Framework for Smart Healthcare Using Hybrid Machine Learning," Posted Date: March 9th, 2022, DOI: https://doi.org/10.21203/rs.3.rs-1428239/v1

[3] Venkatesh R, Balasubramanian C, Kaliappan M. Development of Big Data Predictive Analytics Model for Disease Prediction using Machine Learning Technique. J Med Syst. 2019 Jul 5;43(8):272. doi: 10.1007/s10916-019-1398-y. PMID: 31278468.

[4] Govindaraj Ramkumar, J. Seetha, R. Priyadarshini, M. Gopila, G. Saranya, "IoT-based patient monitoring system for predicting heart disease using deep learning," Volume 218, 15 August 2023, 113235,https://doi.org/10.1016/j.measurement.2023.113235.

[5] Zafer Al-Makhadmeh, Amr Tolba, "Utilizing IoT wearable medical device for heart disease prediction using higher order Boltzmann model: A classification approach," Measurement, Volume 147, 2019, 106815, ISSN 0263-2241, https://doi.org/10.1016/j.measurement. 2019.07. 043

[6] Madupoju Kumar, "Big Data analysis of demand-side management for Industrial IOT applications", 2021, Materials Today: Proceedings by Elsevier Publications, https://doi.org/10.1016/j.matpr.2021.03. 301.

[7] Bebortta S, Tripathy SS, Basheer S, Chowdhary CL. FedEHR: A Federated Learning Approach towards the Prediction of Heart Diseases in IoT-Based Electronic Health Records. Diagnostics (Basel). 2023 Oct 10;13(20):3166. doi: 10.3390/diagnostics 13203166. PMID: 37891987; PMCID: PMC10605926.

[8] G. Rajkumar, T. Gayathri Devi, "Heart disease prediction using IoT based framework and improved deep learning approach: Medical application", Medical Engineering & Physics, Volume 111, January 2023, 103937, https://doi.org/10.1016/j.medengphy.2022.103937.

[9] Ahmed Ismail, Samir Abdlerazek, I. M. El-Henawy, "Big Data Analytics In Heart Diseases Prediction ", Journal of Theoretical and Applied Information Technology, 15th June 2020. Vol.98. No 11.

[10] Rahman, Nayem. "Data Mining Problems Classification and Techniques." IJBDAH vol.3, no.1 2018: pp.38-57. http://doi.org/ 10.4018/IJBDAH.2018010104.

[11] An A, Dahan F, Alroobaea R, Alghamdi WY, Mustafa Khaja Mohammed, Hajjej F, Deema Mohammed Alsekait, Raahemifar K. A smart IoMT based architecture for E-healthcare patient monitoring system using artificial intelligence algorithms. Front Physiol. 2023 Jan 30; 14:1125952. Doi: 10.3389/fphys.2023.1125952. PMID: 36793418; PMCID: PMC9923105.

[12] Sulagna Mohapatra, Prasan Kumar Sahoo, and Suvendu Kumar Mohapatra, "Healthcare Big Data Analysis with Artificial Neural Network for Cardiac Disease Prediction," Electronics 2024, 13(1), 163; https://doi.org/10.3390/electronics13010163.

[13] Khan M. F., Ghazal T. M., Said R. A., Fatima A., Abbas S., Khan M. A., et al. (2021). An IoT-enabled smart healthcare model to monitor elderly people using machine learning techniques. Comput. Intell. Neurosci. 2021, 2487759. 10.1155/2021/2487759

[14] Ma W, Hou X. Big Data Value Calculation Method Based on Particle Swarm Optimization Algorithm. Comput Intell Neurosci. 2022 Jul 1; 2022:5356164. doi: 10.1155/2022/5356164. Retraction in: Comput Intell Neurosci. 2023 Jul 26; 2023:9891753. PMID: 35814581; PMCID: PMC9270169.

[15] Ed-daoudy, A., Maalmi, K. A new Internet of Things architecture for real-time prediction of various diseases using machine learning on big data environment. J Big Data 6, 104 (2019). https://doi.org/10. 1186/s40537-019-0271-7

[16] Ankan Bhattacharya, "Internet of Things and Data Mining for Modern Engineering and Healthcare Applications", February 2022.

[17] Jingfeng Zang, Pengxiang You, "An industrial IoT-enabled smart healthcare system using big data mining and machine learning," November 2022Wireless Networks 29(4), DOI:10.1007/s11276-022-03129-z.

[18] Priyank Sunhare, Rameez R. Chowdhary, Manju K. Chattopadhyay, Internet of things and data mining: An application-oriented survey, Journal of King Saud University - Computer and Information Sciences, Volume 34, Issue 6, Part B, 2022, Pages 3569-3590, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2020.07.002.

[19] Yang J, Li Y, Liu Q, Li L, Feng A, Wang T, Zheng S, Xu A, Lyu J. Brief introduction of medical database and data mining technology in the big data era. J Evid Based Med. 2020 Feb;13(1):57-69. doi: 10.1111/jebm.12373. Epub 2020 Feb 22. PMID: 32086994; PMCID: PMC7065247.

[20] Tahereh Saheb, Leila Izadi, "Paradigm of IoT big data analytics in the healthcare industry: A review of scientific literature and mapping of research trends," Telematics and Informatics, Volume 41, 2019, Pages 70-85, ISSN 0736-5853, https://doi.org/10.1016/j.tele.2019. 03.005.

[21] Jain A et al (2023) Optimized levy flight model for heart disease prediction using CNN framework in a big data application. Exp Syst Appl 223:119859.

[22] Kareemulla Shaik, Janjhyam Venkata Naga Ramesh, "Big Data Analytics Framework Using Squirrel Search Optimized Gradient Boosted Decision Tree for Heart Disease Diagnosis", Appl. Sci. 2023, 13(9), 5236; https://doi.org/10.3390/app13095236.

[23] K. Saikumar, V. Rajesh, Gautam Srivastava, Jerry Chun-Wei Lin, "Heart disease detection based on internet of things data using linear quadratic discriminant analysis and a deep graph convolutional neural network," Front. Comput. Neurosci., 07 October 2022 Volume 16 - 2022 | https://doi.org/10.3389/fncom.2022.964686.

[24] G. Joo, Y. Song, H. Im and J. Park, "Clinical Implication of Machine Learning in Predicting the Occurrence of Cardiovascular Disease Using Big Data (Nationwide Cohort Data in Korea)," in IEEE Access, vol. 8, pp. 157643-157653, 2020, doi: 10.1109/ACCESS. 2020.3015757.

[25] E. Narayanan; R. Jayashree, "IoT based heart disease prediction using smote and machine learning techniques, "Volume 2946, Issue 1, 9 November 2023, https://doi.org/10.1063/5.0178157.

[26] Bahar Farahani, Farshad Firouzi, Victor Chang, Mustafa Badaroglu, Nicholas Constant, Kunal Mankodiya, "Towards fog-driven IoT eHealth: Promises and challenges of IoT in medicine and healthcare," Future Generation Computer Systems, Volume 78, Part 2, 2018, Pages 659-676, ISSN 0167-739X, https://doi.org/10.1016/j.future. 2017.04.036.

[27] Mohammad Ayoub, "An IoT Framework for Heart Disease Prediction based on ODCNN Classifier," February 2020IEEE Access PP (99):1-1, DOI:10.1109/ACCESS.2020.2974687.

[28] Manish Kumar Ahirwar, Piyush Kumar Shukla, "CBO-IE: A Data Mining Approach for Healthcare IoT Dataset Using Chaotic Biogeography-Based Optimization and Information Entropy," Volume 2021 | Article ID 8715668 | https://doi.org/10.1155/2021/ 8715668.

[29] Simanta Shekhar Sarmah, "An Efficient IoT-Based Patient Monitoring and Heart Disease Prediction System Using Deep Learning Modified Neural Network", July 2020IEEE Access PP(99):1-1, DOI:10.1109/ACCESS.2020.3007561

[30] Jameel Ahamed, Abdul Manan Koli, Khaleel Ahmad, Mohd. Alam Jamal, B. B. Gupta, "CDPS-IoT: Cardiovascular Disease Prediction System Based on IoT Using Machine Learning", Received 23 January 2021 | Accepted 18 June 2021 | Published 7 September 2021, DOI: 10.9781/ijimai.2021.09.002.

[31] Mohammad Aljanabi. (2023). Safeguarding Connected Health: Leveraging Trustworthy AI Techniques to Harden Intrusion Detection Systems Against Data Poisoning Threats in IoMT Environments. Babylonian Journal of Internet of Things, 2023, 31–37. https://doi.org/10.58496/BJIoT/2023/005

[32] Jamal, B., Alsaedi, M., & Parandkar, P. (2023). Portable Smart Emergency System Using Internet of Things (IOT). Mesopotamian Journal of Big Data, 2023, 75–80. https://doi.org/10.58496/MJBD/2023/011

[33] Tam Sakirin, & Iqra Asif. (2023). Infusing k-means for securing IoT services in edge computing. Mesopotamian Journal of Computer Science, 2023, 42–50. https://doi.org/10.58496/MJCSC/2023/007

[34] Somasekhar, G., Patra, R.K., Srujan Raju, K. (2021). The Research Importance and Possible Problem Domains for NoSQL Databases in Big Data Analysis. In: Jyothi, S., Mamatha, D.M., Zhang, YD., Raju, K.S. (eds) Proceedings of the 2nd International Conference on Computational and Bio Engineering . Lecture Notes in Networks and Systems, vol 215. Springer, Singapore. https://doi.org/10.1007/978-981-16-1941-0_43

[35] Joshi, A., Choudhury, T., Sai Sabitha, A., Srujan Raju, K. (2020). Data Mining in Healthcare and Predicting Obesity. In: Raju, K., Govardhan, A., Rani, B., Sridevi, R., Murty, M. (eds) Proceedings of the Third International Conference on Computational Intelligence and Informatics . Advances in Intelligent Systems and Computing, vol 1090. Springer, Singapore. https://doi.org/10.1007/978-981-15-1480-7_82

[36] Patel, Riyam, et al. "Quality and Performance Measures in Healthcare Systems Using Fog Computing." In Multi-Disciplinary Applications of Fog Computing: Responsiveness in Real-Time, edited by Debi Prasanna Acharjya and Kauser Ahmed P., 95-120. Hershey, PA: IGI Global, 2023. https://doi.org/10.4018/978-1-6684-4466-5.ch006