International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-9 Issue-6, March 2021

An Enhanced Framework To Secure Big Data Based on Hybrid Machine Learning Technique:ANN-PSO

Salim Raza Qureshi

***Abstract:** With the advancement of smart devices and cloud computing, more and more public health data can be collected from various sources and analyzed in unprecedented ways. The enormous social and academic impact of this development has led to a global buzz for bigdata. Moreover, due to the massive data source, the security of big data in the cloud is becoming an important issue. In these days, various issues have arisen in the field of big data security, such as Infrastructure security, data confidentiality, data management and data integrity. In this paper, we propose a novel technique based on Artificial Neural Network-and Particle Swarm Optimization Algorithm (ANN-PSO) for enabling a highly secured framework. The ANN-PSO method was created to predict health status from a database and its functions were selected from these data sets. The particle swarm optimization algorithm matches the ANN for better results by reducing errors. The results show the potential of the ANN-PSO-based methodology for satisfactory health prediction results. This proposed approach will be tested using large medical data in a Hadoop environment. The proposed work will be carried out in the JAVA work phase.*

***Keywords:** ANN-PSO, Accuracy, Classifier, Error, GOA, Health condition.*

I. INTRODUCTION

Lately, great strides have been made in the innovation of data and correspondence that has changed the world. The world is gradually turning into a small area. These successes include distributed computing, remote use (3G / 4G / 5G), and a targeted mobile phone industry [1]. With the rapid advancement of data development including distributed computing, informal communities, various businesses, and the Internet of Things, information is evolving rapidly and the emergence of an information mine has caused many discoveries and difficulties in various fields of research. The term "big information" basically refers to the amount of huge and confusing information that can be linked [2]. After all, finding large amounts of information requires advanced tools and systems to store, process, and analyze large amounts of information. A large amount of information consists of a large amount of unstructured information that requires constant careful investigation [3].

Manuscript received on February 11, 2021.
Revised Manuscript received on February 20, 2021.
Manuscript published on March 30, 2021.

Assoc. Prof. Salim Raza Qureshi, Department of Computer Science and Engineering, Model Institute of Engineering and Technology, Jammu, India.

Big data and applicable advancements can provide committees with critical information and research mechanisms to reduce healthcare costs and improve clinical design and waste. Huge information aimed at concentrating the stimulus on information that has four properties. Volume, range, speed, and reliability [4]. An important informative study continues to concern medical services.

Today, social service structures rapidly collect clinical information, rapidly expanding the reach of electronically available health records [5]. A thorough background check can be viewed as a strategy to find unusual types of information. Therefore, it is now possible to use a number of traditional information search strategies to find detailed information. An in-depth information search can be divided into ongoing reviews and investigations. The current revision is mainly used in e-business and invoicing. Since the information always shows signs of change, it is necessary to review it quickly and to receive the survey results with a small delay. Forbidden search is generally used for applications that do not require a long response time [6]. A large amount of information is associated with diagnosis in medical services in order to identify patient groups, diseases, and future expectations using various artificial intelligence devices [7, 8]. In educational health services [9], information is broken down into parts and constantly used within the framework of knowledge that patients must take into account. During this process, patient information is combined with clinical reports to provide better suggestions and choices. An additional disease prognosis is important and significant for patients with persistent diseases. Many models of waiting for infection have been proposed in the past. Different types of false nervous system (ANS) strategies are used to predict fatigue. Counterfeit Neural Networks (ANNs) are a subdomain of artificial intelligence (AI) structures. Their ability to link information and compare yield data with vector mapping has proven to be of great benefit in a variety of applications [10]. Either way, ANN puts more effort into model preparation due to the increased stress associated with each shift. In fact, every small change in information gathering affects the model, leading to an unstable result [11]. The prediction system contains information on EHR with hazard components to correctly predict osteoporosis and fractures. In [12], the Creator predicts the collapse of cardiac disappointment by thinking of the patient's physiological information. In both cases, the latent symptoms are not taken into account in the actual anticipation models. The different strategies for analyzing big health data to predict future health status are explored in the next segment of the report.

The remainder of the document serves the following purposes: After this part of the presentation, relevant work in this area will be reviewed and addressed in section 2. Segment 3 shows the problem that distinguishes evidence from existing methods, and section 4 provides an overview of ANN to achieve better results by reducing errors Segment 5 includes the architecture of our proposed health prediction model and also provides a brief explanation of GOA and ANN-PSO. Section 6 presents the experiments carried out, the results obtained, and the graphs.
Finally, section 7 concludes the research part of the conclusion

II. RELATED WORKS

Sudha Ram et al [13] have proposed a technique for utilizing different information hotspots for foreseeing the quantity of asthma-related crisis division (ED) visits in a particular territory. Twitter information, Google search interests, and natural sensor information were gathered for this reason. The model can foresee the quantity of asthma ED visits dependent on close constant natural and internet-based life information with around 70% accuracy. The outcomes can be useful for general wellbeing observation, crisis office readiness, and, directed patient intercession. Trang Pham et al [14] have presented Deep Care, a start to finish profound powerful neural system that peruses therapeutic records, stores past ailment history, gathers current sickness states and predicts future restorative results. At the information level, Deep Care speaks to mind scenes as vectors and model's patient wellbeing state directions by the memory of chronicled records. Based on Long Short-Term Memory (LSTM), Deep Care acquaints techniques with handle sporadically coordinated occasions by directing the overlooking and combination of memory. Profound Care additionally unequivocally models medicinal intercessions that change the course of ailment and shape future therapeutic hazard. Climbing to the wellbeing state level, verifiable and present wellbeing states are then accumulated through multiscale fleeting pooling, before going through a neural system that appraisals future result. The technique improves forecast exactness incredibly. Fan Zhang et al [15] have proposed an errand level versatile MapReduce system. The system broadens the conventional Map Reduce engineering by planning each Map and Reduce task as a steady running circle daemon. The technique was equipped for not just scaling here and there continuously, yet in addition prompting successful utilization of process assets in cloud server farm. So as to improve the system, two gushing information outstanding task at hand expectation techniques are connected, for example, smoothing and Kalman channel, to appraise the obscure remaining burden qualities. The structure plans the Map and Reduce assignments in all respects effectively, as the gushing information changes its landing rate. ANNs have been widely applied in various pattern recognition and classification applications. Traditionally, ANNs are employed to deal with a small volume of data. With the emergence of big data, ANNs have become computationally intensive for data intensive applications which limits their wide applications. Rizwan et al. [21] employed a neural network on global solar energy estimation. They considered the research as a big task, as traditional approaches are based on extreme simplicity of the parameterizations. Yuan and Yu [22] employed cloud computing mainly for exchange of privacy data in a BPNN implementation in processing ciphered text classification tasks..

III. PROBLEM DEFINITION

Simplicity is a major health management issue due to a large number of patients in the same hospital.

Under traditional social security, patient data is collected, stored, and examined in the usual way, which makes it difficult to identify difficult comfort conditions.

Another difficult task is the method of combining and combining information after separating different layers.
The above issues and the lack of solution motivated me to do research in this area.

IV. PROPOSED METHODOLOGY

A. Overview of the Proposed Technique

In this work, we provide the effective platform for dealing with a big data security with the utilization of hybrid technique. Here, we utilized the dataset from UCI hospital datasets, which are collected from the database and its feature gets selected from those datasets. Initially, we execute the optimal feature selection with the utilization of (GOA) Grasshopper Optimization Algorithm. In the preprocessing step, we categorize the dataset into two types such as training and testing parts. These two databases consist of initialization, fitness calculation, updating and optimal features. It helps to select the important features alone to improve the prediction accuracy more efficient. Afterwards, the prediction system will analyze the disease acquired with severity or not. For that, Artificial Neural Network (ANN) classifier is utilized to determine health condition in future by means of training and testing phases. Furthermore, Particle Swarm Optimization algorithm is integrated with ANN for providing better result by reducing the errors. The structure of presented method is demonstrated in figure.1. Proposed work is simplified in the processing stage of JAVA.

B. Pre processing

In this step, we execute the preprocessing process based on data extraction, data classification, and Repositories development. With these steps, we can make the dataset to prepare for further processes. Once the preprocessing is completed, the processed data are taken for next step based on utilizing different machine learning approaches for effective prediction process. For our work, we prepare the tables with the utilization of Grasshopper optimization algorithm for obtaining the optimal result by selecting the features for disease prediction. With the utilized features, the current nature of specific patient data can be predicted by the hybrid optimization based on PSO and ANN.

C. Feature Selection by Grasshopper Optimization Algorithm

The Feature Selection (FS) for order issues is a troublesome and computationally exorbitant strategy, especially when overseeing high dimensional informational collections. It diminishes capacity use and preparing time and addresses the issue of dimensionality. So as to counteract highlight for expectation issue, Grasshopper Optimization Algorithm (GOA) is proposed in our article.

Retrieval Number: 100.1/ijrte.F5385039621
DOI:10.35940/ijrte.F5385.039621
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication

Grasshoppers examine search space by aversion, and they experience promising zones by interest

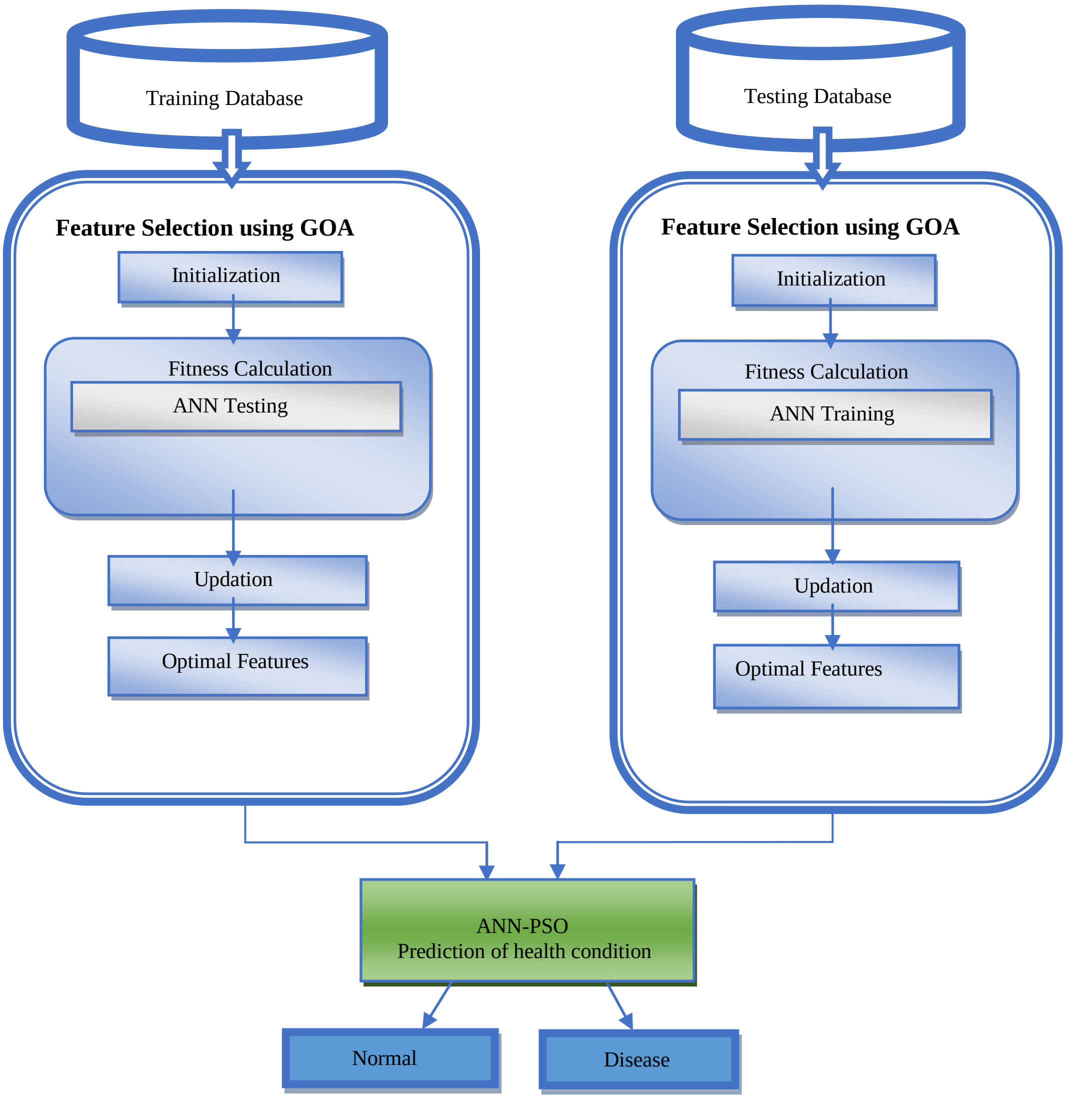

Figure 1: Overview of the Proposed Architecture

The following processing steps are utilized to choose the features:

1. Initialization

To upgrade the highlights, GOA calculation at first makes a subjective populace of the arrangement. Arrangement creation is a significant advance of improvement calculation that recognizes the ideal arrangement rapidly. Each picture having the quantity of highlights among them we chose ideal features. In GOA, at first, we randomly initialize position of all grasshoppers. Here, each individual in the swarm is considered as grasshopper in a dimensional D search region. Among total features we select the important features. The underlying arrangement initial solution is given in Table 1. In Table 1, the value "1" denotes the comparing highlight is chosen and "0" speaks to the relating highlight isn't chosen.

An initial solution generation is given below; the solution $P(s_{11}, s_{12}, \ldots, s_{1n})$ is calculated based on the equation,

$$S_{ij} = L_i + U_1 - S_{ij} \tag{1}$$

Where; Li speaks to the lower bound coefficient, Ui speaks to the upper bound coefficient, Sij speaks to the underlying arrangement.

2. Evaluation of Fitness

In the wake of creating the underlying arrangement Sij, the wellness of the arrangement is assessed. The determination of the wellness is a significant part of GOA calculation. It is utilized to assess the inclination (goodness) of hopeful arrangements. Here, arrangement precision is the primary criteria used to structure a wellness work. The wellness calculation is executed for every arrangement. For every emphasis, the wellness is determined utilizing condition (2),

$$\text{Fitness} = \frac{\text{TrNe} + \text{TrPo}}{(\text{TrNe} + \text{TrPo} + \text{FaNe} + \text{FaPo})} \tag{2}$$

$TrPo$ → True positive, TN → True negative,

TABLE I. INITIAL FEATURE SELECTION OF GOA

		G^1	G^2	G^3	G^4	……….	G^{28}	G^{29}	G^{30}	…….	G^{83}	G^{84}	G^{85}	……..	G^{100}
Sol_i	L_1	1	0	1	0	……….	1	0	1	…….				……..	0
	L_2	0	1	0	1	……….	0	1	0	…….				……..	1
	L_3	1	0	1	0	……….	1	0	1	…….				…….	0
							⋮								
	L_n	0	1	0	1	……….	0	1	0	…….				…….	1

FP → False positive, FN → False negative

3. Updation

Subsequent to, figuring the fitness esteem, we update the arrangement dependent on grasshopper optimization algorithm.

Updated equation is given in equation (3).

$$X_i = R_i + S_i + T_i \tag{3}$$

Where Yi speaks to the situation of the ith grasshopper, Ri is the social association, Si is the gravity power on the ith grasshopper and Ti demonstrates the breeze shift in weather conditions.

$$R_i = \sum_{\substack{k=1 \\ k \neq i}}^{N} s(d_{ik})\hat{d}_{ik} \tag{4}$$

$$d_{ik} = |Y_k - Y_i| \tag{5}$$

$$\hat{d}_{ik} = \frac{Y_k - Y_i}{d_{ik}} \tag{6}$$

where dik is the separation between the ith and the kth grasshopper and s is a quality of social power. The gravity force (G_i) is calculated using equation (7).

$$G_i = -g\hat{e}_g \tag{7}$$

Where S is the gravitational consistent and demonstrates a solidarity vector towards the focal point of the earth. wind advection T_i is evaluated using equation (8).

$$A_i = u\hat{e}_w \tag{8}$$

Where u is the steady float and is a unit vector toward the breeze. The substituting esteems R, S and T in (9).

$$Y_i = \sum_{\substack{k=1 \\ k \neq 1}}^{N} R(|Y_k - Y_i|)\frac{Y_k - Y_i}{d_{ik}} - se_s + ce_w \tag{9}$$

$R(r) = f\, ex^{\frac{-r}{l}} - e^{-r}$ and the number of grasshoppers are denoted as N. Using equation (9), we can update the solution.

4. Termination Criteria

The calculation stops its execution just if a greatest number of cycles is achieved and the arrangement which is containing the best wellness worth is picked utilizing GOA and it is given as a best answer for characterization.

V. PREDICTION OF HEALTH CONDITION USING ANN-PSO

The health prediction system will analyze the disease acquired with severity or not. For that, ANN classifier is utilized to find the health condition in future by means of training and testing phases. After the selection of optimal features, we have applied Particle Swarm Optimization algorithm is integrated with ANN for providing better result by reducing the errors. Here, ANN-PSO is explained in a detail manner. PSO algorithm includes simplicity, simplicity of execution, high caliber of arrangement, quicker assembly towards ideal arrangement and less parameters.

A. ANN training by PSO

At first, the particles are dispersed haphazardly in the arrangement space. Each molecule P in the swarm S is addressed as {P, Q} where P = {p1, p2, p3… pn} addresses the situation of the particles and Q = {q1, q2, q3… qn} speaks to the weight of the particle. In each cycle, the particles gain from one another and update their insight with respect to the whereabouts of a best solution. Each molecule screens its best arrangement with its contrasting position in pbest and the swarm's best position is followed in gbest.

B. ANN-PSO

1. Initialization

Adjusting the PSO calculation to prepare the ANN includes the accompanying advances. Since the loads of the ANN should be advanced, they should be followed as the situation of the particles in the PSO calculation.

The issue space contains the blends of all conceivable weight esteems for the ANN. In the preparation of the ANN by the PSO, the portrayal of the association weight of the ith molecule is given underneath,

$$Q = q_i^1, q_i^2, q_i^3 .... q_i^n \quad (10)$$

Where, q_i^1 q_i^2 ...... q_i^n denoted as weight of the between input i[th]particle, hidden layers, between hidden layers , output respectively.

Weight position of the previous best fitness value of particle is given below,

$$P = p_i^1, p_i^2, p_i^3 .... p_i^n \quad (11)$$

Where, p_i^1 p_i^2 ...... p_i^n represents the position of the input between i[th]particle, hidden layers and between hidden layers, output respectively.

2. Fitness Evaluation

The fitness of the i[th] molecule is communicated as far as a yield mean-squared blunder of the neural systems as pursues,

$$fit = \min\ (error \quad function) \quad (12)$$

Where, $fit$is represented as fitness value, min represented as minimum error values. These procedures are rehashed for a predefined number of emphases or until fitness is come to.

3. Updation

After the calculation of fitness, we update the best solution. In the event that the fitness is the best so far for the molecule it will be taken as its own best and in the event that it is the best so far for the swarm, it would be considered as worldwide best. The worldwide best position after an ideal number of cycles yield the streamlined loads for the ANN.

Pbest position among all the particles between input and output is denoted as,

$$P_{best} = p_{best}^1, p_{best}^2, p_{best}^3 .... p_{best}^n \quad (13)$$

The weights and position of manipulated particles are,

$$Q_i(t+1) = Wq_i(t) + r_1 z_1 (pbest - x_i(t)) + r_2 z_2 (gbest - x_i(t)) \quad (14)$$

$$P_i(t+1) = p_i(t) + q_i(1+t) \quad (15)$$

W denoted as inertia weight, $q_i(t)$ represents particles present weight and $q_i(t+1)$ represents the updated weight of the particle. r_1 and r_2 represents the randomly distributed inputs i.e. [0,1] and z_1 and z_2 are the constants and $p_i(t)$ denoted as present position of particles, $p_i(t+1)$ denoted as updated position of particles. Condition (14) is utilized to register the new weight of the molecule dependent on its past weight and the separations of its present position from the best encounters both in its very own and as a gathering. Refreshed weight must be inside the predefined extend. On the off chance that it disregards the points of confinement, it is set to a well-known worth. The underlying loads of the underlying particles were produced arbitrarily in the scope of (0, 1).

At that point the new position of every molecule is assessed as aggregate of its past position and comparing refreshed weight utilizing (14). Halting guideline (greatest emphases have completed, or the best wellness worth is accomplished) or there is stagnation in the ideal arrangement, i.e., it has not changed for quite a while. In the event that none of the ceasing standard is met, at that point go to step (ii) until any halting foundation is met.

ANN-PSO ALGORITHM

Input: $Q = q_i^1, q_i^2, q_i^3 .... q_i^n$ (Weight)
Output: Optimized weight

1. Characterize the ANN engineering – number of information, covered up and yield neurons.
2. Distinguish the wellness work which returns the blunder as distinction of real and anticipated yield for the ANN
3. Start a swarm of 'p' particles with irregular loads of 'n' measurement where n is the absolute number of loads that should be upgraded for the ANN
4. For every cycle do this to the 'p' particles

Discover the fitness of every particle as characterized in Step 2

```
If   fitness >pbest
     {
          Update pbest
     }
If  fitness >gbest
     {
          Update gbest
     }
Update speed and position
```

Do the means till the cycles are finished

gbest has the best loads for the ANN which yields the most noteworthy expectation precision.

In figure 2, we provide the flowchart for ANN & PSO process and it is how effective to deal with the optimization solution. The flowchart starts with the initialization of ANN for effective feature selection. After that we divide the dataset into two modules. One module takes care of the training purpose and the second module takes care of the Testing module. The process starts with the initialization of the multiple ANNs and particle swarm. The endpoint of the flowchart obtain the optimal position of the particle swarm.

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \times 100 \quad (17)$$

5. F-measures

The combination of precision and recall measures are comprises i=a mean value of harmonic is called as F-measure. It is represented as follows,

$$\text{f - measures} = 2 \times \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right) \quad (18)$$

B. Execution investigation of the introduced method

The performance analysis of introduced method is appeared in below area. Here the table demonstrates the performance measures for the proposed ANN-PSO method. Here we are thinking about the performance measures as precision, recall, f-measures, time and memory esteem. The performance analysis of proposed system by shifting the of information is explained in Table 2.

The graphical representation of the proposed investigation of precision by various training data for our proposed method is exposed in fig 3.

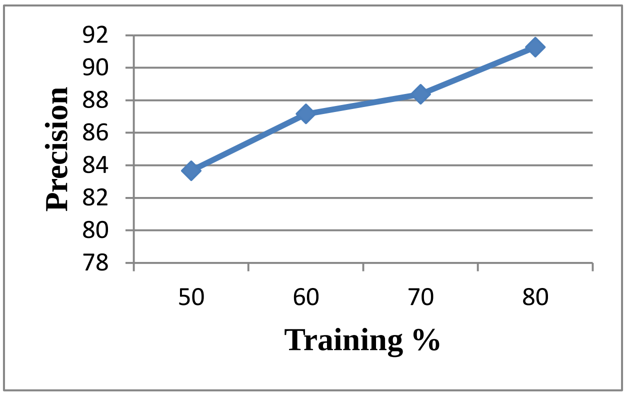

Figure 3: Precision of Proposed Method for different Training Data

Fig.3 shows the proposed precision for various training data. For the training data, 50% obtained 83.654 precision and for the training data, 60%achieves precision 87.155, the precision 88.36 and 91.258 is for the training data 70% and 80%. The graphical representation of recall for proposed ANN-PSO method is shown in figure 4.

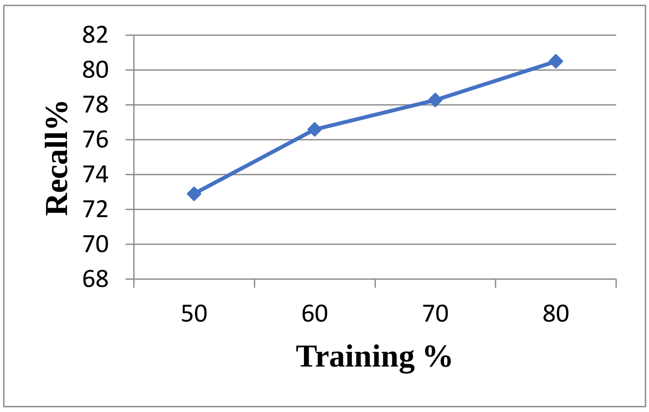

Figure 4: Recall of Proposed method for different Training Data

The recall measure of our proposed method is varied based on the different input data sets. The input data values are 50%, 60%, 70% and 80% obtained recall values i.e. 72.89127, 76.58645, 78.27195 and 80.49227. The incoming data is increased similarly the recall values also increased. The maximum value of recall is 80.49227 and the minimum value is 72.89127. The graphical representation of f-measure is demonstrated below, the f-measures are increased based on the input data. The input data 50% got 77.90266, 60% got 81.52965 f-measure, 70% obtains 83.01061 and the 805 got highest value of f-measure is 85.53772. The lowest value of f-measure is obtained for the 50% of input data.

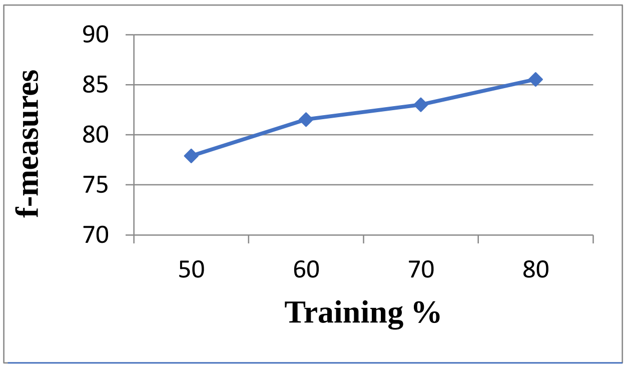

Figure 5: f-measures of proposed method for varied training data

The line chart for proposed method ANN_PSO is shown in figure 6; the running time of proposed method is represented in ms. if the input of ANN-PSO is increased and the running time also increased. For input data, 50%, 60%, 70% and 80% takes 21547ms, 23684ms, 28947ms and 3125 7ms.The maximum time is 31257 ms is taken by the data 80%.

TABLE II. PERFORMANCE MEASURES OF ANN PSO

Train %	Precision	Recall	f-measures	Time (ms)	Memory(bits)
50	83.654	72.89127	77.90266	21547	1120451
60	87.155	76.58645	81.52965	23684	1351475
70	88.36	78.27195	83.01061	28947	1424876
80	91.258	80.49227	85.53772	31257	1635875

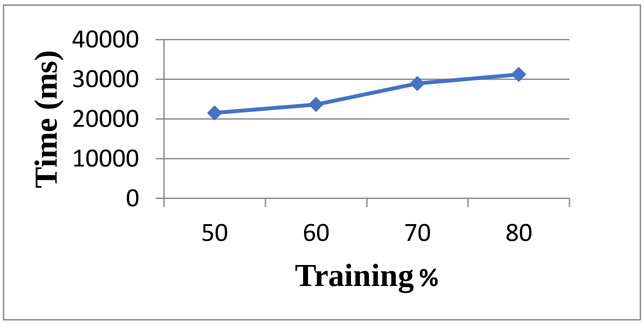

Figure 6: Time of proposed method for varied training data

The proposed ANN-PSO memory graphical representation is demonstrated below, the memory usage of ANN-PSO is presented in bits, input training data in % 50 got 1120451 bits, 60 obtained 1351475 and the maximum memory usage is 1635875 bits and it is taken by data 80%. The memory usage of ANN is increased based on the increased input data.

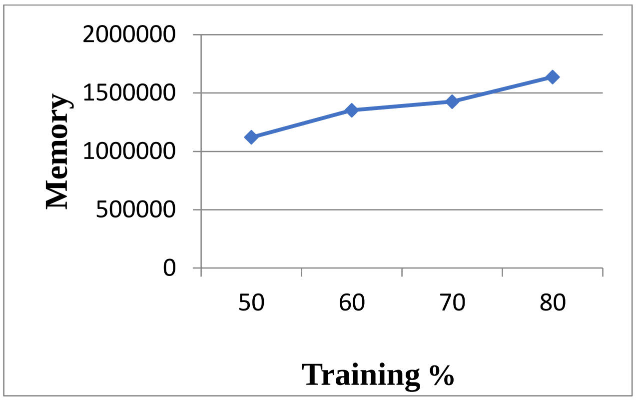

Figure 7: Memory of proposed method for varied training data

The performance evaluation of our existing method is presented in the table 3. Here, the table 3 shows the performance evaluation for the OFF-method ANN. Graphical representations of our OFF-method also presented in the following section.

The graphical representation of our existing method is shown below fig.8, fig.9 and fig.10. The fig.8 is the line chart of precision and recall. The precision of existing method is low compared with proposed ANN-PSO. The precision of ANN is 80.22858 for 50%, 84.01957 for 60%, and 84.91539 for 80%. The maximum value of ANN is 87.65084 but the ANN-PSO got maximum at 91.258. The recall value of ANN is low while compared with proposed ANN-PSO. From this proposed ANN-PSO performs better than the existing ANN.

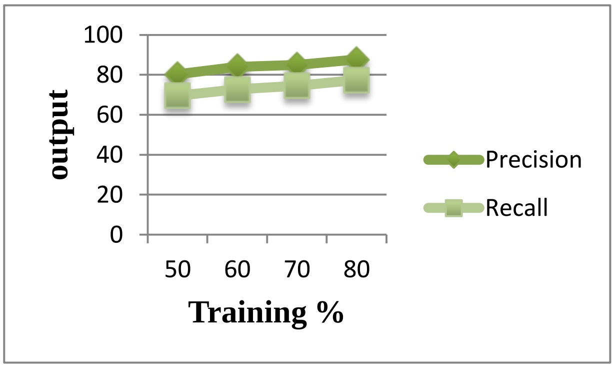

Figure 8: Precision and Recall of existing method for varied training data

The f-measure and utilized time of existing method graph is demonstrated in below. The existing method takes more time for running compared with ANN-PSO. The minimum running time of ANN is 22844ms but the minimum running time of ANN-PSO is 21547ms and in ANN the input data 60% and 70% takes 24278ms and 30167ms time, the maximum running time of ANN is 32048 but our ANN-PSO takes maximum of 31257 ms for running.

TABLE III. PEFORMANCE MEASURES FOR EXISTING METHOD ANN

Train %	Precision	Recall	f-measures	Time (ms)	Memory(bits)
50	80.22858	69.60372	74.53944	22844	1150731
60	84.01957	72.72496	77.96533	24278	1378063
70	84.91539	74.54259	79.39161	30167	1436650
80	87.65084	77.32636	82.16554	32048	1680751

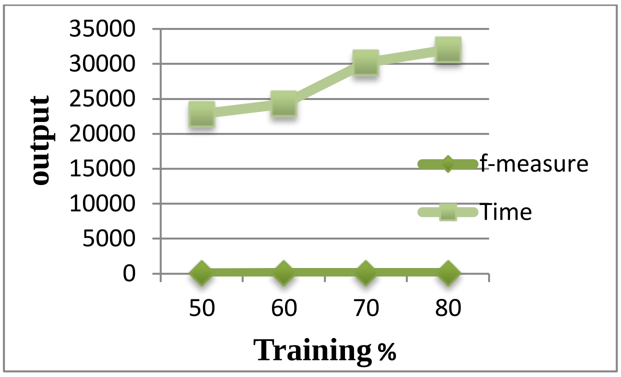

Figure 9: f-measure. time of existing method for varied training data

The graphical representation for memory usage of existing ANN is demonstrated in figure 10; the input data, 50%, 60%, 70% and 80% obtained 1150731, 1378063, 143665 and 1680751. The usage of memory is represented in bits.

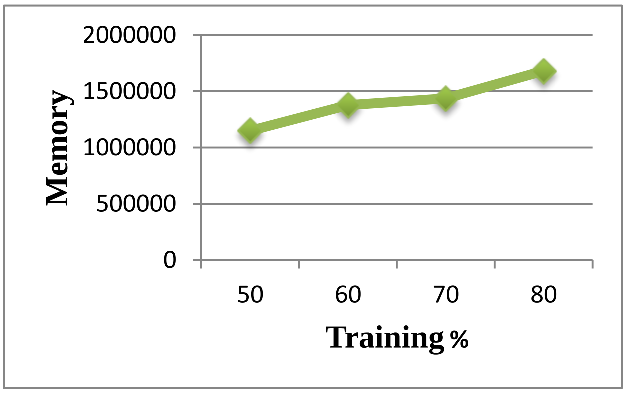

Figure 10: Memory of existing method for varied training data

From the above graphs and tabulation, it clearly shows that our proposed ANN-PSO approach predicts the health condition efficiently while compared with other existing ANN approach. Moreover, it additionally limits the error rate.

VII. CONCLUSION

Neural network-based strategies have been viewed as hopeful method for health condition prediction. In this work, neural system based PSO detection method is proposed. ANN based PSO methodology is created for the determination of health condition of the UCI emergency clinic datasets are gathered from the database and its element gets chose from those datasets. In this paper, the features are chosen optimally by applying Grasshopper Optimization Algorithm and the seriousness of disease is analyzed by applying an ANN (Artificial Neural Network) classifier. Also, Particle Swarm Optimization algorithm is coordinated with ANN for giving better outcome by diminishing the errors. Finally, the proposed methodology is tested in the framework of big data Hadoop. The outcomes exhibit the capacity of the ANN-PSO based methodology for producing sufficiently good health condition determined results.

REFERENCES

1. Loai A. Tawalbeh, Rashid Mehmood, Elhadj Benkhlifa, Houbing Song," Mobile Cloud Computing Model and Big Data Analysis for Healthcare Applications", IEEEAccess, Volume: 4,2016
2. DongxiaoGu,Jingjing Li, XingguoLi,Changyong Liang," Visualizing the knowledge structure and evolution of big data research in healthcare informatics", International Journal of Medical Informatics, Volume 98, Pages 22-32, 2017
3. Aisha Siddiqa, Ahmad Karim, Abdullah Gani," Big data storage technologies: a survey", Frontiers of Information Technology & Electronic Engineering, August 2017, Volume 18, Issue 8, pp 1040–1070
4. Min Chen,ShiwenMao, Yunhao Li," Big Data: A Survey", Mobile Networks and Applications, Volume 19, Issue 2, pp 171–209, 2014
5. Lopez, D., Gunasekaran, M., Murugan, B. S., Kaur, H., Abbas, K. M.," Spatial big data analytics of influenza epidemic in Vellore, India", In Big Data (Big Data), IEEE International Conference on (pp. 19–24). IEEE, 2014
6. Ahmed Oussous, Fatima-Zahra Benjelloun, Ayoub AitLahcen, Samir Belfkih," Big Data Technologies: A Survey" Journal of King Saud University - Computer and Information Sciences, 2017
7. S. Wang, X. Chang, X. Li, G. Long, L. Yao, and Q. Z. Sheng, ``Diagnosis code assignment using sparsity-based disease correlation embedding,'' IEEE Trans. Knowl. Data Eng., vol. 28, no. 12, pp. 3191_3202, Dec. 2016.
8. Gang Luo," Predict-ML: a tool for automating machine learning model building with big clinical data", Health Information Science and Systems, 2016
9. V. Tresp, J. M. Overhage, M. Bundschus, S. Rabizadeh, P. A. Fasching, and S. Yu, ``Going digital: A survey on digitalization and large-scale data analytics in healthcare,'' Proc. IEEE, vol. 104, no. 11, pp. 2180_2206, Nov. 2016.
10. Dimitrios H. Mantzaris, George C. Anastassopoulos and Dimitrios K. Lymberopoulos," Medical Disease Prediction Using Artificial Neural Networks", IEEE International Conference on Bioinformatics and BioEngineering,2008
11. S. Gopakumar, T. Tran, T. D. Nguyen, D. Phung, and S. Venkatesh, ``Stabilizing high-dimensional prediction models using feature graphs,'' IEEE J. Biomed. Health Inform., vol. 19, no. 3, pp. 1044_1052, May 2015.
12. H. Li, X. Li, M. Ramanathan, and A. Zhang, ``Prediction and informative risk factor selection of bone diseases,'' IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 12, no. 1, pp. 79_91, Jan./Feb. 2015.
13. Sudha Ram, Wenli Zhang, Max Williams, Yolande Pengetnze," Predicting Asthma-Related Emergency Department Visits Using Big Data", IEEE Journal of Biomedical and Health Informatics, Volume: 19, Issue: 4, 2015
14. Trang Pham,Truyen Tran, Dinh Phung, Svetha Venkatesh," Predicting healthcare trajectories from medical records: A deep learning approach", Journal of Biomedical Informatics, Volume 69, Pages 218-229, 2017
15. Fan Zhang, JunweiCao, Samee U. Khan,KeqinLi, Kai Hwang," A task-level adaptive MapReduce framework for real-time streaming data in healthcare applications", Future Generation Computer Systems, Volumes 43–44, Pages 149-160, 2015
16. Liqiang Nie, Meng Wang, Luming Zhang, Shuicheng Yan, Bo Zhang, Tat-Seng Chua," Disease Inference from Health-Related Questions via Sparse Deep Learning", IEEE Transactions on Knowledge and Data Engineering, Volume: 27, Issue: 8, 2015
17. Shiva Pratap Gopakumar, Truyen Tran, Tu Dinh Nguyen, Dinh Phung, and Svetha Venkatesh," Stabilizing High-Dimensional Prediction Models Using Feature Graphs", IEEE Journal of Biomedical and Health Informatics, Volume: 19, Issue: 3, 2015
18. J. Henriques, P. Carvalho, S. Paredes, T. Rocha, J. Habetha, M. Antunes, J. Morais," Prediction of heart failure decompensation events by trend analysis of telemonitoring data", IEEE Journal of Biomedical and Health Informatics, Volume: 19, Issue: 5, 2015
19. Gunasekaran Manogaran, R. Varatharajan, M. K. Priyan," Hybrid Recommendation System for Heart Disease Diagnosis based on Multiple Kernel Learning with Adaptive Neuro-Fuzzy Inference System", Multimedia Tools and Applications, Volume 77, Issue 4, pp 4379–4399, 2018
20. PrasantKumarSahoo, savedKumarMohapatra, shih-linewoo," Analyzing Healthcare Big Data with Prediction for Future Health Condition", IEEE Access, Volume: 4 ,2016
21. M. Rizwan, M. Jamil, and D. P. Kothari, "Generalized neural network approach for global solar energy estimation in India," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 3, pp. 576–584, 2012.
22. J. Yuan and S. Yu, "Privacy preserving back-propagation neural network learning made practical with cloud computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 1, pp. 212–221, 2014.

AUTHOR PROFILE

Assoc. Prof. Salim Raza Qureshi has over 20 years of experience in academia and academic administration. He did his post-graduation from Vinayaka Mission Research Foundation and currently leading the CSE department as Head of the Department. He is the member of the Board of Studies of Jammu University and is the Student Branch Coordinator of CSI Student chapter at MIET. Recently, he was awarded by CSI for being the longest continuous Student Branch Coordinator in North India. He is the faculty sponsor for the ACM student branch at the college and manages all the activities of the chapter. He has been certified by Mission 10X (WIPRO, DALE CARNEGIE) on High Impact Teaching Skills. His two resource guides on Software Processes and Risk Management are now available in the Mission10X (WIPRO) portal for thousands of users to read, understand and utilize. He has been certified by EMC on Information Storage & Management, Data Science and Big Data Analytics. His research interest areas are Big Data and Data Analytics. He has, to his credit, over 15 research papers published in Conferences of National and International repute. He is currently pursuing his PhD in computer science.