DIMENSIONALITY REDUCTION OF MEDICAL BIG DATA USING ADAPTIVE HONEY BADGER AND ADAPTIVE DEEP NEURAL NETWORK

Dr.J.S.T.M Poovarasi

Assistant Professor, Department of Computer Science, Agurchand Manmull Jain College

Abstract

Every day, enormous amounts of intricate information are created in numerous fields. Data sets that are so vast that traditional database administration and data analysis technologies are unable to handle them are referred to as complex information. Medical big data management and analysis involve numeral experiments through the organization, storage, and analysis of the data. To avoid the problem, in this paper adaptive honey badger (AHB) algorithm with the selected features is presented for dimensionality reduction. The proposed approach consists of three phases such as pre-processing, feature selection and classification. At first, the medical data are fed to the pre-processing phase to eliminate the missing values and redundant data. Then, to minimize computational complexity, time consumption and storage, the optimal features are selected using the AHB algorithm. Then, the selected features are fed to the adaptive deep neural network (ADNN) classifier to classify data as normal or diseased data. The experimental results show that our presented model is attained an average accuracy of 96.8% for the cervical cancer dataset and 98% for the Wisconsin cancer dataset.

Keywords: Cervical cancer, adaptive honey bee, ADNN, LSTM, CNN, attention mechanism, dimension reduction and big data.

1. Introduction

Digital medical technology has advanced due to information technology, and medicinal information is expanding at an exponential proportion [1]. A biomedical investigation has also transformed into characteristic data-intensive learning [2], creating the issue known as "big data." Information has evolved into a new strategic resource and a key driver of innovation in the age of big data, and it is changing how biomedical research is conducted as well as how people live and think [3]. Through the interoperability assessment and technical specifications of big data in the medical service sector for future advancement and implementation [4, 5]. It also offers a strong theoretical with a practical foundation for the creation and application of big data in the field of medicine and health [6]. The project study outcomes can enhance the theoretical and request organization in the pitch of medicinal fitness big data research by supplying the main technologies and data models of that field [7]. The need for early prevention of disease has never been greater than it is now due to the rise in healthcare costs [8]. Particularly responsible for this are the growing hazards posed by disease variations and bioterrorism, as well as recent advancements in data collecting and processing technologies [9].

The need to create an effective, sensitive and affordable solution for disease prevention grows as healthcare data volumes rise [10]. Traditional preventive efforts lack the tools to process

massive volumes of data and instead primarily concentrate on promoting the benefits of healthcare [11]. Improved health promotion and illness prevention can result from using IT to advance healthcare quality [12]. It is a truly interdisciplinary task that calls for a variety of types of skills in many academic fields as well as enormously huge quantities of information. Due to its improved performance in iterative and incremental processes is known as Apache Spark [13] has recently gained increasing traction in the big data space. The most straightforward and efficient method for supervised learning is regarded to be lazy learning, also known as instance-based learning [14]. In comparison to their competitors [15], these approaches appear to have a classification phase that moves much more slowly. Additionally, those who are considered lazy learners prefer to compile examples from data streams that use information pertaining to out-of-date concepts in order to make decisions. Since the predicted distribution of data cannot be acquired for a variety of reasons [16], particularly in the case of cost-sensitive business circumstances has baffled the majority of academics.

Additionally, virtual illustrations are created to stabilize the circulation of information, improving the minority class's recall rate at the expense of the classification model's precision. Therefore, using big data and deep learning for this prediction is more successful because equally of these domains are expanding quickly [17]. Big data refers to complex huge datasets that can affect traditional data warehousing's famed storage, security, manageability, manipulation, and other processes.

The main objective is to decrease the dimension of information by predicting the diseased data. To achieve this concept, two stages are developed such as AHB algorithm-based dimension reduction and ADNN-based prediction. The proposed AHB is a mixture of the honey badger algorithm and the OBL approach. The OBL strategy is used to increase the searching ability of HB algorithm. Similarly, the suggested ADNN is a grouping of LSTM, CNN and attention mechanism. The main objective of the method is given below;

- To decrease the dimension of the big data, an efficient feature selection method is suggested. The feature selection method is a mixture of the HB algorithm with the OBL approach called as AHB algorithm.
- For the classification process, an ADNN classifier is proposed which is a hybridization of CNN, LSTM and attention mechanism. The presented ADNN classifier avoids the problem present in the individual CNN and LSTM classifiers.
- The performance of the proposed approach is evaluated using two sets of the medical dataset and different metrics.

2. Literature review

Lot of researchers had developed dimensionality reduction with medical data classification. Among them, few works are analyzed here; In 2020 Ismail, A., et al, [18] have evaluated the impact of applying machine learning algorithms using electronic health records. It is used to create a powerful suggestion platform that uses real-time sensor readings, past user profile data, streamed medical data, and knowledge databases to create consumers with the best recommendations and alerts. The technique attained an accuracy of 90.6%. In 2019 Venkatesh,

R., et al, [19] have developed disease prediction the using Naive Bayes Technique. The heart disease data through the UCI machine learning repository is trained using the Naive Bayes approach. Then, it began predicting the classification depending on the assessment consequences. The findings demonstrate that the BPA-NB scheme offers improved accuracy of 97.12% in predicting the illness rate. In 2021 Kim, J. and Lim, J., [20] have analyzed the Korea National Health and Nutrition Examination Survey was used to gather data on 7031 participants over the age of 65 for the prediction of dementia using deep neural networks and information on the use of medical services. The technique produces a superior result to previous algorithms, with 85.5% of the area under the curve (AUC).

In 2021 Safa, M. and Pandian, A., [21] has developed a great information investigation of heart disease. In order to predict cardiac disease, a multi-level fuzzy rule generation estimated with Cardiac Disease Infection Transmission Analysis (CDITA) weight is used. High performance is produced by the system. In 2018 Kiral-Kornek, I., et al, [22] have predicted epileptic seizures with big data and deep learning. The implementation of the predictive model onto an ultra-low influence neuromorphic processor for autonomous processes on a wearable expedient is demonstrated as being feasible. The prediction method considerably outperformed an equal accidental forecaster for all patients by 42%, with a mean sensitivity of 69% and a mean time in warning of 27%.

In 2020 Ramani, R., et al, [23] developed the prediction of chronic disease using an artificial neural network. In circumstances of retrieving large amounts of data, it boosts throughput and redundancy. Therefore, combining a modified ANN classifier with a simplified framework can help to get better results. The empirical findings on the chronic diabetes dataset demonstrate the artificial neural network's ability to forecast the precision, sensitivity, and modified specificity levels. In 2019 Vijayakumar, D.R., et al, [24] have developed the outcomes of many machine learning approaches that can help forecast diabetes for a patient with a higher degree of accuracy. Diabetes can also be predicted using two separate supervised machine learning techniques, including SVM for classification and K-means for clustering. The algorithmic application's predicted prediction accuracy is 94.9%, and its convergence speed.

3. Proposed high-dimensional medical data classification

The aim of this study is to diminution the dimension of information to predict the diseased data. Handling high-dimensional data are time-consuming and takes large storage places. So, dimension reduction is introduced in this paper. For, dimension decrease development, in this paper AHB algorithm is presented. The overall structure of the proposed methodology is presented in figure 1. The suggested method consists of three phases such as pre-processing, feature extraction and prediction. In pre-processing, the missing values and redundant data are removed. Then, the pre-processed data is fed to the input of the feature selection stage. For feature selection, in this paper AHB algorithm is presented. In this stage, the high dimensional data sizes are reduced. Then, the reduced dataset is given to the input of the prediction process. For prediction, the ADNN classifier is presented. The proposed classifier effectively predicts the diseased data.
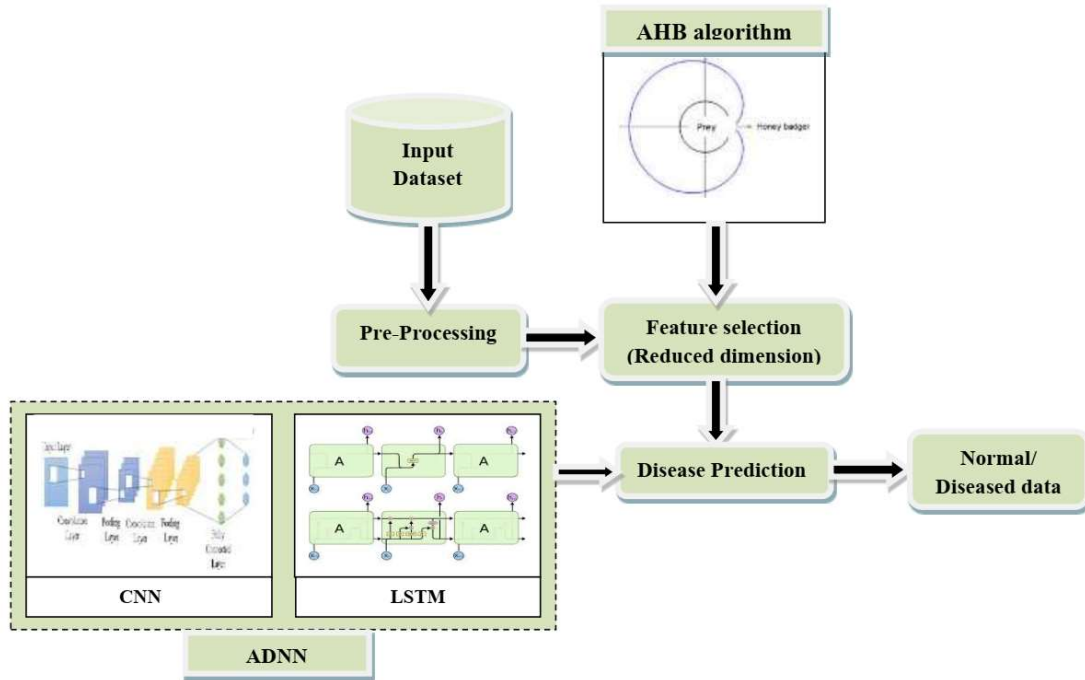
Figure 1: Overall structure of proposed methodology

3.1 Pre-processing

For medical data classification, in this paper different types of diseased datasets are utilized. The dataset has **m** characteristics spread across **n** records, making up the entire number of records in the dataset. The dataset records in this could contain duplicates, erroneous data, or both. The accuracy of classification will be impacted by this. Therefore, the incoming data are pre-processed before the categorization process begins. The step-by-step explanation of the pre-processing procedure is given below;

Step 1: Let the dataset D which contain of **n** number of archives and **m** number of characterizes. Originally, the input feature values are transformed into arithmetical standards.

Step 2: After that, the arithmetical standards are standardized. The standardization function is given in equation (1).

$$A^{ij} = \frac{A^{ij} - M^i}{(Max - Min)\,value\,of\,feature\,j}$$

(1)

Where;

$A^{ij} \square \ i^{th}$ row j^{th} column attribute value

$M^i \square$ Mean value of the ith column attribute

Once the normalization is accomplished, the pre-processed information is given to the dimension reduction process.

4.2 Feature selection using adaptive honey Badger algorithm

For the feature selection process, in this paper, the AHB algorithm is employed. The AHB is a mixture of a honey badger and oppositional-based learning (OBL) strategy. HBA follows the honey badger's behavior to capture its prey. This process is done through a set of steps named digging and honey. In the digging phase, prey is determined by the honey badger's scent, while in the honey stage, the honey badger follows the honey bird to locate the hive. The HB method possesses the advantages of dynamic search with the drawback of becoming stuck in local optima as a result of population diversity loss, particularly when attempting to solve a challenging optimization problem. In this regard, the goal of this research is to enhance the initial HBA by preserving population variety throughout the search procedure. By preserving the diversity of the badger population, the suggested AHBA improves the search method. Better converging towards the global optimum results from this. The key advantage of the basic HBA it is converging, resulting in the OBL approach being used to preserve the variety of potential solutions. The following lists the procedures involved in creating the feature section;

Step 1: Solution encoding: Initially, the candidate solutions are arbitrarily dispersed in the search space. All the optimization algorithms are starts with this candidate initialization process. In this paper, selected features are considered as the candidate solution. The initial population is given in equation (2).

$$Initial\ population\ P_i = \begin{bmatrix} S_1 \\ S_2 \\ ... \\ S_N \end{bmatrix} \tag{2}$$

$$S_1 = [F_1, F_2, F_3, ..., F_N] \tag{3}$$

Where, F_i represent the feature value, P_i represent the total population and S_i represent the solution (Honey). Sample solution encoding format is given in table 1.

	F_1	F_2	F_3	F_4		F_V
S_1	0	1	1	1	……	1
S_2	0	1	1	0	……	1
S_3	1	1	0	1	…..	0
S_4	1	0	0	1	……	0
S_5	1	1	1	0	……	1

Table 1: Sample solution format

Table 1 shows that "1" indicates that the related feature is enabled and "0" indicates that it is not.

Step 2: Opposite solution generation: After initialization process, the opposite solutions are created. The opposite solutions are utilized to increases the solution with better diversity and also this strategy increases the search space. The opposite solutions are produced based on the conflicting direction of candidate solution S_i. The opposite solution calculated using below equation.

$$\bar{S}_i = \vec{LB} + \vec{UB} - \vec{S}_i \qquad (4)$$

Where, LB and UB represent the upper and lower bands respectively. The generated opposite solution and initial solutions are given to the fitness calculation process.

Step 3: Fitness calculation: It is utilized to find the best feature set. In this classification approach, maximum accuracy (A) is measured as the fitness value. The fitness calculation is given in equation (6).

$$Accuracy = \frac{True^{Pos} + True^{Neg}}{True^{Pos} + True^{Neg} + False^{Pos} + False^{Neg}} \times 100 \qquad (5)$$

$$Fitness = Max(A) \qquad (6)$$

Step 3: Described the intensity (I): Intensity can be related with the concentration of the prey in addition distance among the honey badger. The intensity function of HBA is presented as follows,

$$I_I = R_2 \times \frac{X}{4\pi D_I^2} \qquad (7)$$

$$X = (S_I = S_{I+1})^2 \qquad (8)$$

$$D_I = S_{\Pr ey} - S_I \qquad (9)$$

Where, R_2 represent the Random variable among 0 to 1, D_I represent the Distance among prey and i^{th} badger and X represent the concentration strength or source strength

Step 4: Density factor updating: In the HBA, the time changeable randomization is measured by the density factor to empower a smooth transition from exploitation to exploration. Decreasing the density factor, which reduces the iteration and reduces the randomization with respect to the time based on the below equation,

$$a = C * \exp\left(\frac{-T}{T^{\max}}\right) \qquad (10)$$

Where is represented as a maximum number of iterations and C represented as the constant parameter which is taken as 2.

Step 5: Local optimum condition: In this step, local optimum conditions are checked. With this algorithm, the Flag FL can be altered in the search way intended aimed at achieving height opportunities aimed at the agent to scan the exploration interplanetary thoroughly.

Step 6: Agent position update: The updating process is split into two sections such as digging phase (exploration) and honey phase (exploitation).

<u>Digging phase:</u> During the digging phase, the cardioid motion can be computed as follows,

$$S_{NEW} = S_{\Pr ey} + FL * \beta * I * S_{\Pr ey} + FL * R_3 * \alpha * D_I * \left| \cos(2\pi R_4) * \left[1 - \cos(2\pi R_5) \right] \right| \qquad (11)$$

Here, FL represent the flag which change the search direction, R_3, R_4, R_5 can be described as the random variables among 0 and 1, D_I can be described as the distance among the prey and ith honey badger, β can be described as the ability of the honey badger to achieve the food and $S_{\Pr ey}$ can be described as the position of the prey.

The flag operation can be formulated as follows,

$$FL = \begin{cases} 1 & if \ R_6 \leq 0.5 \\ -1 & else \end{cases} \qquad (12)$$

Where, R_6 represented the haphazard number among 0 and 1. During the digging stage, time changing search influence factor, distance among the prey and badger and a honey badger can be depends on smell intensity. Additionally, throughout digging behaviour, a badger might get slightly trouble that permits it towards compute smooth optimal prey site.

Honey phase (exploitation): The honey badger is following the honey guide bird towards to achieve the optimal results which formulated as follows,

$$S_{new} = S_{prey} + FL * R_1 * \alpha * D_i \qquad (13)$$

Where, R_1 represent the random number among 0 to 1, D_i represent the distance information, $X_{\Pr ey}$ represent the honey badger prey location and X_{new} represent the honey badger new position.

Step 7: Termination criteria: When the highest level of fitness is attained, the procedure is over. For further evaluation, the solution with the highest fitness rating is chosen. The ADNN classifier receives the chosen characteristic as input. The Pseudocode of the AHBA is given in table 2.

Algorithm 1: Pseudocode of the AHB algorithm
Input: pre-processed dataset, parameters of AHB algorithm
Output: Reduced dataset (optimal features)
Start
Generate the initial solution (refer table 1)
Compute the opposite solution using equation (4)
Compute fitness function using (6)
Save the optimal value
While $T \leq T^{max}$ do
Update the decreasing factor
For I=1 to N do
Compute the intensity
If $T < 0.5$ then
Update the position (11)
Else
Update the position (13)
End if

End for
End while stop condition compensated
Return
Output: Reduced dataset

<div align="center">Table 2: Algorithm of the AHB algorithm based feature selection</div>

4.3 Disease prediction using ADNN classifier

The reduced dataset is sent to the ADNN classifier to determine if a piece of data is normal or abnormal once the features have been chosen. The proposed ADNN classifier combines attention mechanisms with DNN and LSTM. Figure 2 shows the suggested illness prediction model's organizational structure.
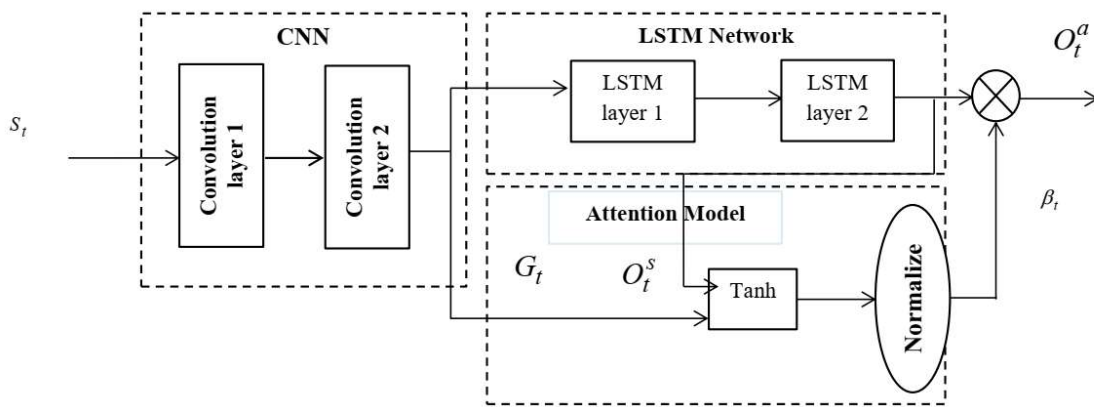
<div align="center">Figure 2: Structure of proposed ADNN (CNN-LSTM) with attention layer</div>

4.3.1 Convolution neural network (CNN)

The CNN classifier originally provided the chosen characteristics S_t . There are two convolution layers in the CNN. A chosen collection of features is used as the CNN's input. At each time step t, the 1D convolution technique is applied to the input characteristics. The immediate perceptional region is acquired by a mobile filter expending a 1-D convolution kernel filter. The convolution kernel filter method is described in the sentences that follow.

$$Y_t = \sigma\left(W_s * S_t + b_s\right) \qquad (14)$$

Where, W_s denotes the weight value of the filter, b_s represent the input data, * represent the convolution operation, σ represent the activation function and Y_t represent the output of the convolution layer. The downsampling process is carried out by the pooling layer after the convolution layer. The attribute map's dimension can be reduced and overuse can be avoided thanks to the pooling layer. In order to define the non-linear gathering of data recovered by the maximum combining to construct the future consequence, FC levels are often generated in the previous phases of the CNN architecture. Because the size of the input data is not very vast, we just used one convolution layer in this research instead of a pooling layer.

4.3.2 Long short term memory (LSTM)

The LSTM classifier receives the CNN's output as input. A better form of the RNN that can store data for a long period is the LSTM. The section of the LSTM that is pertinent is cellular orientation. The cell configuration's content can only be modified or deleted via the entrances, and it is encrypted using the sigmoid value. Three gates—a forget gate, an input gate, and an output gate—combine to form the LSTM. In figure 3, the LSTM design is shown.
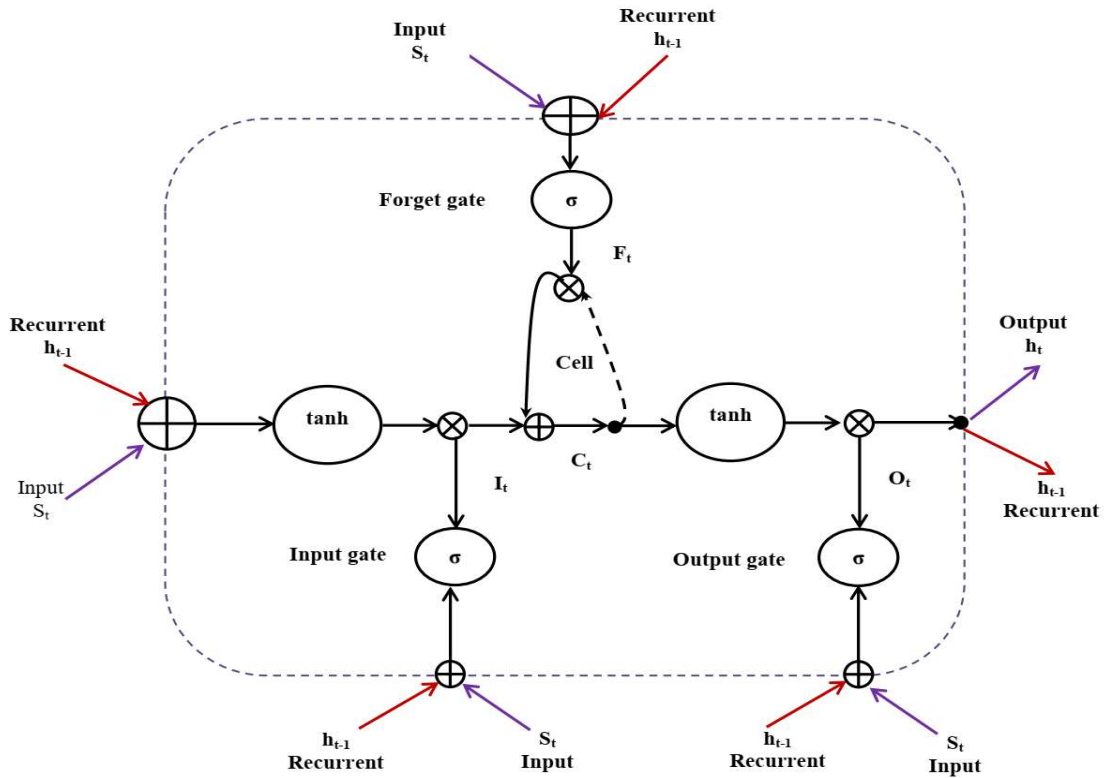
Figure 3: Structure of LSTM classifier

Forget gate is depending on the preceding covered complaint, F_t contributions the LSTM in defining the information those requirements to be mixed laterally by the cell situation h_{t-1}, as well as information input S_t.

$$F_t = \sigma\left(W_F \cdot [h_{t-1}, S_t] + B_F\right) \tag{15}$$

Where, h_{t-1} represent the previous state, S_t denotes the input information at t, $\sigma(\cdot)$ denotes the sigmoid activation function, W_F denotes the weight environments and B_F denotes the bias vector.

Input gate I_t determines which data would be rescued on the different applicant call state, $\tilde{C}_t$;

$$\tilde{C}_t = \tanh\left(W_c \cdot [h_{t-1}, S_t] + B_c\right) \tag{16}$$

$$I_t = \sigma\left(W_i \cdot [h_{t-1}, S_t] + B_i\right) \tag{17}$$

Where, $\tanh(.)$ denotes the hyperbolic tangent function in equation (16).

The preceding cell state C_{t-1} and the new candidate cell state $\widetilde{C}_t$ are collective to inform the new cell state, C_t as follow;

$$C_t = F_t * C_{t-1} + i_t * \widetilde{C}_t \tag{18}$$

The output gate is O_t then created to control the LSTM cell's output. The expected outcome h_t is indicated as shown below: cell state activation C_t and multiplication O_t.

$$O_t = \sigma\left(W_o.[h_{t-1}, X_t] + B_o\right) \tag{19}$$

$$h_t = O_t.\tanh\left(C_t\right) \tag{20}$$

The loss function of the system is ultimately determined and is calculated as follows by using the MSE as the error calculation.

$$Loss = \sum_{t=1}^{N}\left(H_t - T_t\right)^2 \tag{21}$$

Tt stands for the expected outcome, N represent the number of data and H_t represents the obtained output.

4.3.3 Attention mechanism

An attention layer that integrates the CNN and LSTM output is the main part of the ADNN. The attenuation layer in ADNN is initially added to the output of the LSTM. AM is an excellent strategy for increasing the importance of important data depending on the unique visual development. When the pictorial organization studies something in the environment, it often focuses on a specific area as needed rather than watching a scenario from beginning to end. Depending on the situation, the AM chooses to concentrate on some of the most crucial information, discard unnecessary information, and emphasize desired information. AM is used in a variety of fields, including machine translation, seismic forecasting, and image captioning. AM uses a weight distribution system to select the most valuable information by giving it a higher value. As a result, it boosts the effectiveness of traditional.
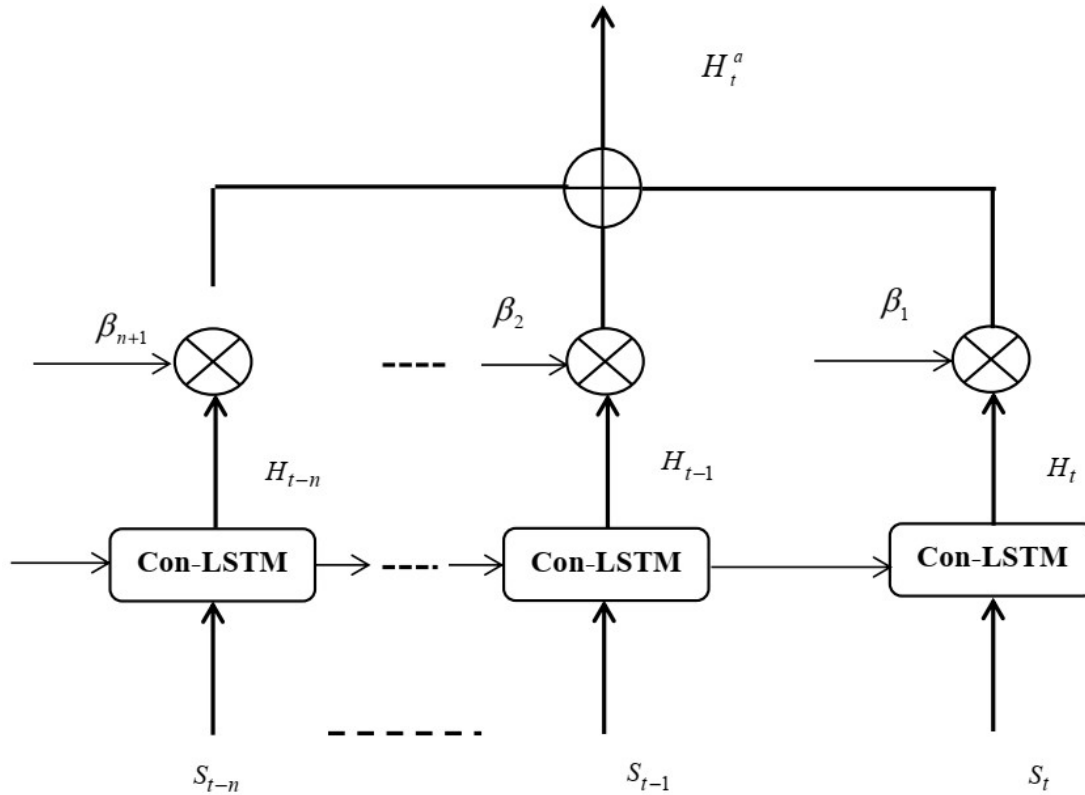
Figure 4: Attention mechanism with Conv-LSTM networks

4. Results and discussion

The effectiveness of suggested organization is analysed in this segment. The suggested technique is implemented using java and the system has the configuration of 4GB Intel core processor. The performance of proposed approach analysed based on three metrics namely, accuracy, sensitivity and specificity. For performance analysis two set of dataset is utilized namely, cervical cancer Dataset and Breast Cancer Wisconsin Dataset.

4.1 Dataset description

For experimental investigation, we utilize two set of dataset namely, ***Cervical cancer Dataset and Breast Cancer Wisconsin Dataset.*** The Cervical cancer Dataset consists of 858 instances and 36 attributes. The information was gathered in Caracas, Venezuela at the "Hospital Universitario de Caracas." The dataset contains information on the activities, prior medical history, and demographics of 858 patients. Some patients decide not to respond to some of the questions out of privacy concerns. This dataset makes use of four classes. 569 cases, 32 attributes, and two classes—malignant and benign—make up the Breast Cancer Wisconsin Dataset. Characteristics are estimated utilizing a digital file of a fine needle aspirate (FNA) of a breast mass.

4.2 Experimental results

Experimental consequences achieved through the suggested and existing approach is analyzed. The main objective of suggested method is to decrease the dataset dimension and effectively

classify a medical data. For dimension reduction, AHB algorithm is employed and for classification ADNN is utilized. Both stage performances are analysed in this section.
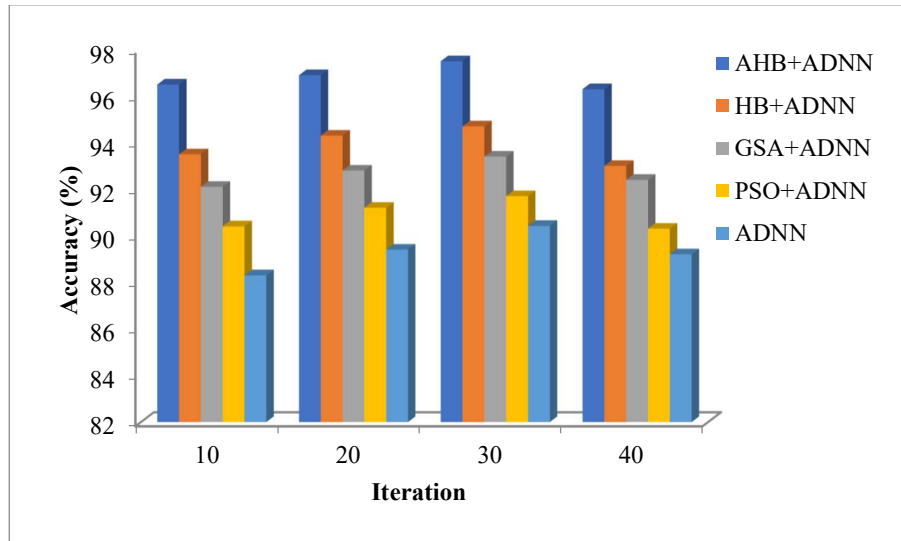

Figure 5: Comparative analysis based on Accuracy using **cervical cancer Dataset**
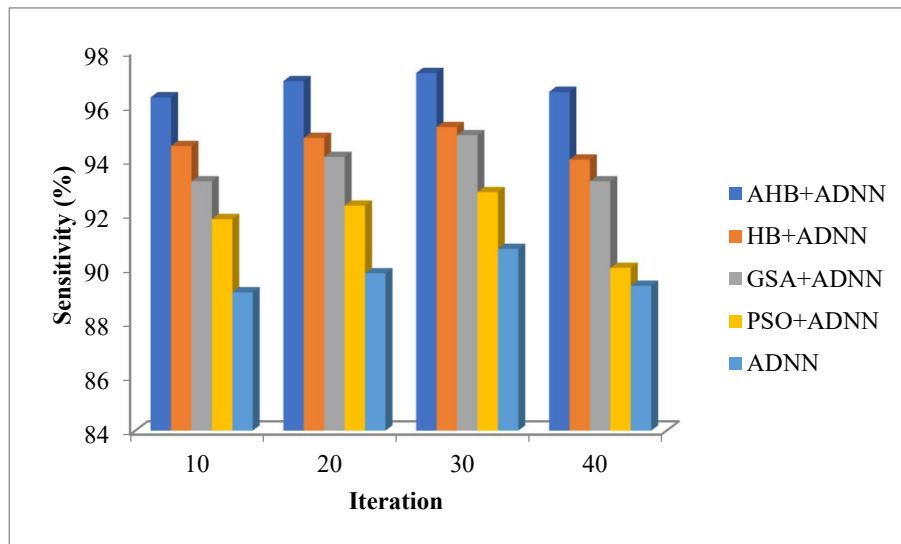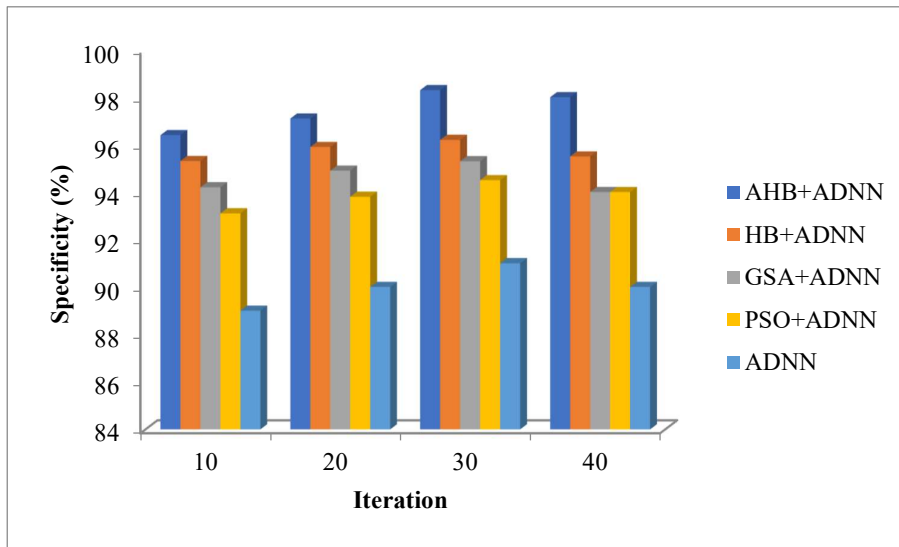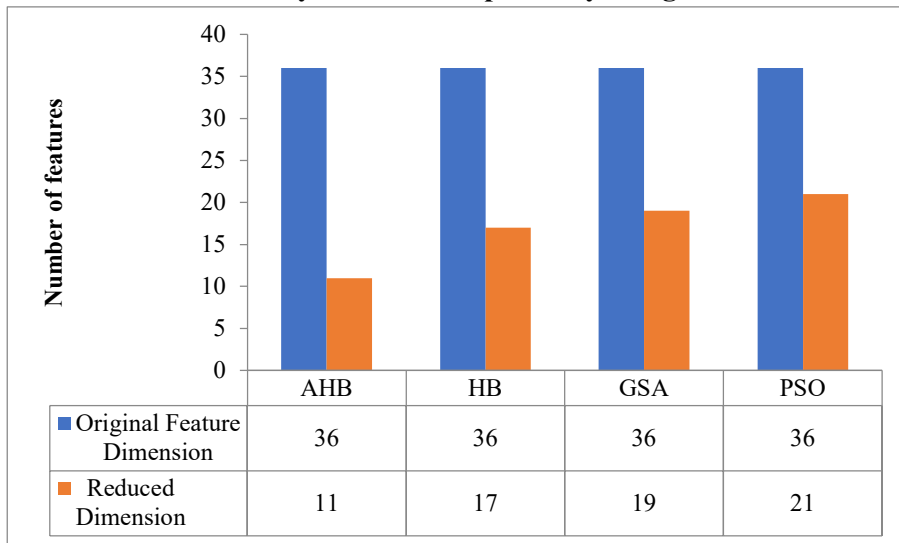

Figure 6: Performance analysis based on sensitivity using cervical cancer dataset

Figure 7: Performance analysis based on specificity using cervical cancer dataset

	AHB	HB	GSA	PSO
■ Original Feature Dimension	36	36	36	36
■ Reduced Dimension	11	17	19	21

Figure 8: Selected feature for cervical cancer dataset

In figure 5, the efficiency of recommended method is analysed depend on the accuracy. Here, we analyse the performance depend on the feature selection algorithm. To decrease the complexity, we introduce the concept of feature selection. For feature selection AHB algorithm is presented. To establish the efficiency of feature selection technique, we compare our algorithm with different algorithms namely, HB based feature selection, gravitational search algorithm (GSA) based feature selection, particle swarm optimization (PSO) based feature selection and without feature selection based prediction. as per figure 5, our suggested approach achieved the supreme accuracy of 97.5% which is 94.7% for HB+ADNN depend prediction, 93.4% for GSA+ADNN based prediction, 91.7% for PSO+ADNN based prediction and 90.42% for ADNN based prediction. Compared to other algorithms, proposed AHB perform well in feature selection. This is due to OBL strategy. The proposed algorithm effectively reduces the dataset sizes. Figure 6 analyses the presentation of the method depend on sensitivity. The suggested strategy achieved a maximum sensitivity of 97.2%. When

analyzing figure 6, which is high associated to the other approaches. Similar to this, our suggested strategy attained the highest level of specificity. This is because we used a good feature selection strategy. The proposed feature selection approach not easily fall on local optimum and due to multiple search space convergence speed also increased. The total selected feature is given in figure 8. Using our proposed approach among the thirty-six features, eleven features are selected. From the results section, we clearly understand proposed approach reached the better results compared to other methods.
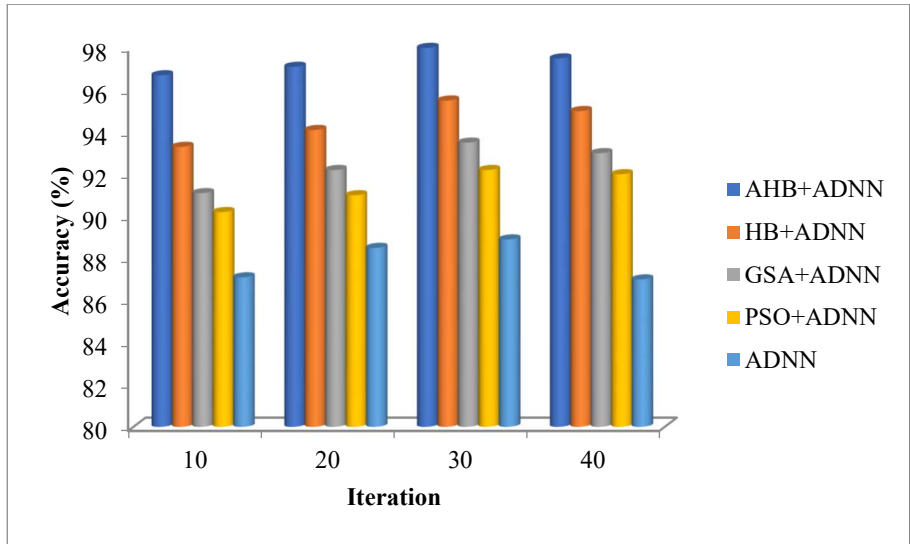

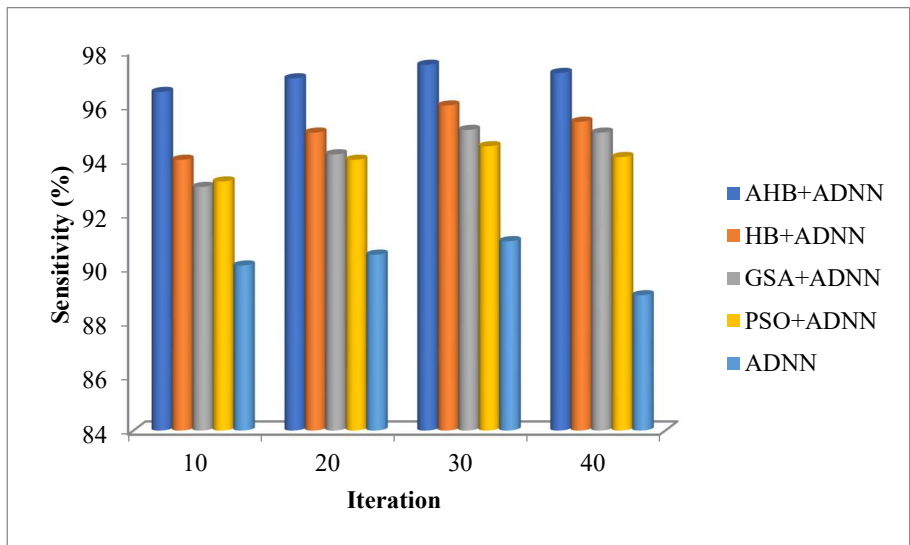Figure 9: **Performance analysis based on accuracy using breast cancer Wisconsin dataset**

Figure 10: Performance analysis based on sensitivity using breast cancer Wisconsin Data Set
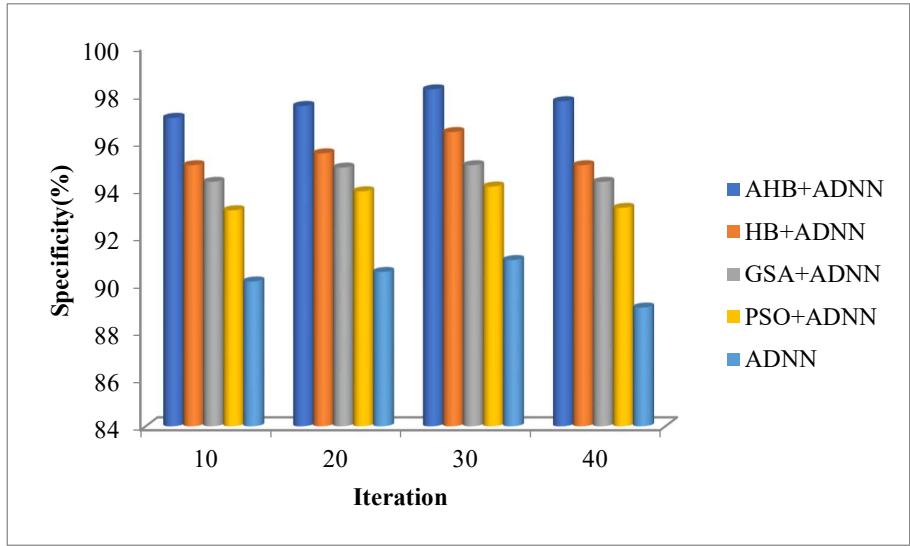
Figure 11: Performance analysis based on specificity using breast cancer Wisconsin dataset

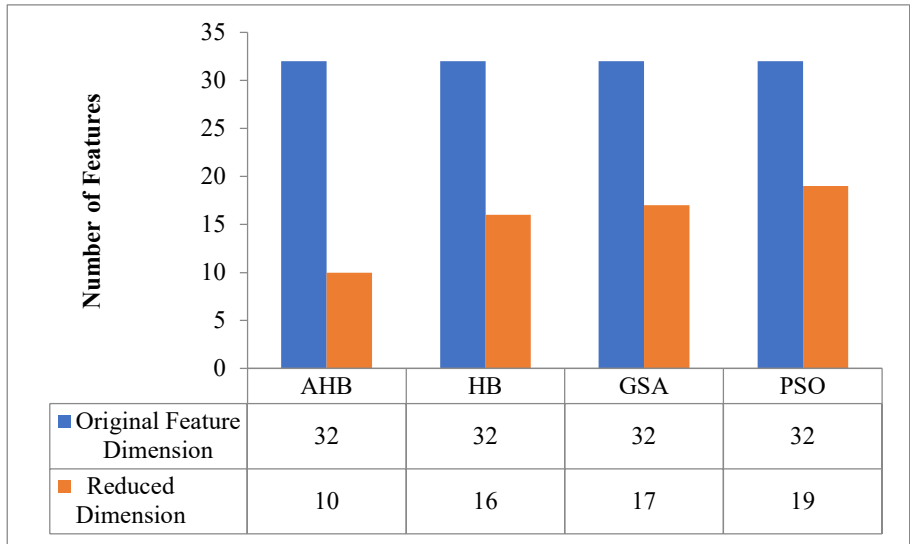|	AHB	HB	GSA	PSO
■ Original Feature Dimension	32	32	32	32
■ Reduced Dimension	10	16	17	19

Figure 12: Selected feature for Wisconsin dataset

The performance of proposed method analysed depend on accuracy by varying iteration using Wisconsin dataset is given in figure 9. When examining figure 9, our suggested method achieved to determined accuracy of 98% which is 2.7% better than HB+ADNN based disease prediction, 4.8% better than GSA+ADNN based disease prediction, 5.78% better than PSO+ADNN based disease prediction and 10.23% better than ADNN based disease prediction. As per the figure 9, we clearly understand proposed method attained the better results related to the additional algorithms. In figure 10, the presentation of suggested method is analysed using sensitivity measure for cancer dataset. According to figure 10, we clearly understand without feature selection based disease prediction approach attained minimum sensitivity compared to other methods. Like that, our proposed approach attained the better specificity. When analysing figure 11, HB+ADNN and GSA+ADNN based prediction methods are attained the slightly similar output. In figure 12, number of selected feature is analysed. From

the results section we clearly understand suggested method accomplished is improved output associated to other method. This is due to efficient feature selection technology.

Measures	AHB+ADNN	AHB+LSTM	AHB+CNN	AHB+ANN	AHB+SVM
Accuracy	96.8	93.2	90.2	88.5	89.2
Sensitivity	96.72	94.4	91.7	89.3	90.2
Specificity	97.45	94.8	92.4	90	90.3

Table 3: Comparative analysis based on different measures for cervical cancer dataset

Measures	AHB+ADNN	AHB+LSTM	AHB+CNN	AHB+ANN	AHB+SVM
Accuracy	98	93.7	91.2	89.5	90.2
Sensitivity	97.5	94.9	92.7	90.3	91.2
Specificity	98.2	95	93.4	90	90.3

Table 4: Comparative analysis based on different measures for Wisconsin dataset

In Table 3, the performance of proposed method is analysed depend on different metrics by varying different classifier using cervical cancer dataset. Here, we analyse the proposed classifier performance with different classifiers namely, LSTM based prediction, CNN based prediction, ANN based prediction and SVM based prediction. For analysis, both the classifier we used same feature selection technique. when analysing table 3, suggested technique accomplished the average accuracy of 96.8 which is 93.2 for AHB+LSTM based cervical cancer prediction, 90.2% for AHB+CNN based cervical cancer prediction, 88.5% for AHB+ANN based cervical cancer prediction and 89.2% for AHB+SVM depend calculation. Correspondingly, our suggested technique accomplished the maximum sensitivity of 96.72% and specificity of 97.45% which is high associated to other approaches. In Table 4, we compare our suggested methodology performance through dissimilar classifier algorithm using Wisconsin dataset. Here, also proposed method attained the better results. As can be seen from Table 3 and Table 4, our presented approach attains better prediction accuracy, which confirms the advantages of our presented approach to some level. The evaluation outcome further shows that our presented approach is effective and has some merits for automatic cancer disease prediction.

5. Conclusion
An efficient dimension reduction based medical data classification system has been presented. To achieve this concept, hybrid dimension reduction approach and hybrid classifier has been used. The proposed dimension reduction algorithm effectively selects the optimal number of features and this algorithm increases the search spaces and reduced the computation

complexity. The proposed ADNN classifier has been used for prediction process. The effectiveness of suggested method has been analysed based on various measurements and our suggested method achieved the average accuracy of 96.8% for cervical cancer dataset and 98% for Wisconsin cancer dataset. These results show that our presented model is effective for automatic cancer diagnosis. In future we will focus on map reduce concept for big data analysis.