Optimal Artificial Neural Network based Data Mining Technique for Stress Prediction in Working Employees

Anitha S (✉ nathan.anitha@gmail.com)
  Alagappa University   https://orcid.org/0000-0002-5563-1833
Vanitha M
  Alagappa University

Optimal Artificial Neural Network based Data Mining Technique for Stress Prediction in Working Employees

*Ms.S.Anitha[1] Dr.M.Vanitha[2]

[1]*Ph.D Research Scholar, Department Of Computer Applications, Alagappa University, Karaikudi, India*

[2]*Assistant Professor, Department Of Computer Applications, Alagappa University, Karaikudi, India*

[1]*nathan.anitha@gmail.com*, [2]*mvanitharavi@gmail.com*

Abstract

Depression has become a common issue among IT industry professionals today. Lifestyle changes and new work culture increase the risk of depression among employees. Various companies and organizations offer mental health plans and try to pacify the work environment. However, the problem is already out of control. This research paper proposes an effective deep learning model for stress prediction among working employees with the help of lion optimization-based Optimal Artificial Neural Network (OANN) model. Here, the features are selected using optimal ANN technique and the diseases are predicted using lion optimization method. ANN technique eliminates inappropriate and unnecessary attributes in a significant manner, once the information on calculated characteristics and weight is disseminated to lion optimization classifier. The test results inferred that the Artificial Neural Network is highly efficient than the current OANN algorithm method, based on lion optimization. The study evaluated the data and found that the performance of employees working under normal conditions was higher when compared to the performance of employees who work under stress. Furthermore, attitude-coping efforts may be a cognitive-behavioral mechanism, which explains how workload is related to courage and work performance of employees with high stress level.

Keywords: Data Mining, Stress Prediction, Disease prediction, Employees, Artificial Neural Network, Lion Optimization, Rule Generation.

1. Introduction

Work stress is defined as a negative, physical and mental reaction to situations that overwhelm workers. Construction is known as one of the most stressful industries, due to the requirement of both physical and mental activities in hazardous work environments [1]. As per the study conducted earlier [2], excess amount of work stress increases the risk of errors, accidents, injuries and health problems and is associated with stagnation / low productivity. These problems impact the construction sector. It has also been reported that 68% of the workers in construction sector undergo high levels of stress [3, 4]. Competition has intensified since modern technology offers various advantages such as advanced technologies, modern lifestyles and many other services. Politicians, managers and subsidiaries are increasingly interested in gaining a competitive edge within companies. Therefore, employees work hard to gain that competitive advantage and move ahead in the company hierarchy. On one hand, this race for the profit of competition among different organizations led to progress in modern structures and development. While on the other hand, a stressful environment is created at work [5, 6]. These effects impact people which can be understood through reduced job satisfaction, decreasing motivation, increasing attendance, increasing employee income, reducing job quality and productivity, and increasing smoking and consumption. Work stress also affects the overall performance of the company likewise [7].

Several predictors of the disease have been identified in earlier investigations. Psychological problems, unfavorable work characteristics and poor health behaviors have been proved to have association with high risk of disease. By combining these information, efforts have been taken to develop bacterial prognosis models in the absence of disease [8]. Data collection is generally poor while the features are selected based on the specific condition and infection of the patient. In the current study, the number of neurons are 34 and the data set target has 7 attributes. A total of 41 attributes such as diabetes, blood pressure, headache, mental illness, heart disease, gastritis, stroke and the environment is focused. With the development of big data analytics, technology started focusing on disease prognosis from the perspectives of big data analysis. Various studies have been conducted to automatically select the properties from multiple data, in order to improve the accuracy of the risk classification, rather than previously-selected characteristics [9]. However, these existing works primarily considered the structured data. With regards to unstructured data, using synthetic neural networks (ANN) for automatic extraction of text properties has already attracted a lot of attention since it produced excellent results [10]. The main purpose of ANN

organization is to solve problems alike human brain. Neural networks are highly structured mainly due to problems with visual and image recognition. This is called 'data mining'. Furthermore, the disease characteristics vary from one region to another mainly due to climatic and lifestyle differences in the specific region [11].

So, in the past decades, a number of researchers heavily focused Logistic Regression (LR), Convolutional Neural Network (CNN), and machine on machine learning-based disease prognosis (ML), Support Vector Machine (SVM) [12] etc. At the time of testing, patient data is classified as either normal or abnormal. In addition, these programs often have less functionality and disadvantages [13, 14]. The best facilities, created by Machine Learning methods, help the doctors diagnose and treat disease. Few experiments were carried out in the past to predict the dynamics of the disease and to develop techniques for its classification. In addition, due to climatic and habitat changes in the region, the disease characteristics too vary from one region to another. By integrating clinical decision support into pre-determined employee files, one can minimize the medical errors and unnecessary variations while patient safety and patient outcomes can be improved [15]. A data modeling and analysis tool, e.g. the data effect has the potential to create a knowledge-based environment that can significantly improve the quality of clinical outcomes. Finally, the authors propose a new OANN disease prognosis algorithm based on lion optimization for structured and unstructured features. Disease prediction model is developed by combining structured and unstructured traits. Based on the experience of the authors, it is concluded that OANN-based lion optimization performance is better than other existing methods.

2. Literature survey

With the development of big data in biomedical networks and service networks, top-down research about the clinical information contributes to the early diagnosis of diseases, silence and network management. However, when the type of clinical information is poor, the accuracy of the tests gets reduced. In addition, some areas benefit from certain local conditions which can affect the visual appearance of the wounds.

Chen *et al*. [16] proposed a machine learning method to effectively predict the chronic outbreaks in sick communities. A Convolutional Neural Network based Multimodal Disease Risk Prediction (CNN-MDRP) was proposed in this study and it can be modified using structured and unstructured hospital data. To the best of researcher's knowledge, none of the existing works focused on both

types of data in clinical big data analysis. Bone disease prognosis and information risk factor selection were performed by Hugh Lee *et al*. [17]. In this study, electronic patient medical records (EHR) were shared by healthcare organizations and specialists. They analyzed all the EHR data and structured it in order to develop an integrated representation of the characteristics of all the risk factors, available in EHR data. A sample pair was created for two different partners (for example, sick and non-sick patients) so as to analyze their characteristics and identify better informed risk factors.

Ping Luo *et al*. [18] implemented the Protein-Protein Interaction (PPI) disease gene prediction algorithm (called dgSeq) by combining clinical data on RNA-Seq data, and Online Mendelian Inheritance in Man (OMIN). According to the study, dgSeq was estimated to build different networks based on the changed information calculated from clinical data of the RNA opening. A genetic disease network was constructed based on OMIM data to select the stable groups of non-pathogenic genes (negative samples). After separating the functionality of the PPI networks and various other networks, the logistic delay classifiers were created.

A collective prognosis of miRNAs -related disease based on abduction training was developed by Jiao Luo *et al*. [19]. By linking disease similarities and miRNA similarities to the established miRNA disease association, a miRNA disease network was created to predict the association between miRNA and the disease. The CPDL then calculated the compatibility value and updated the network configuration until the integrated criteria were met. The corresponding node values, including miRNA and disease, were calculated by examining the homogeneity based on its neighbors. The network system was updated with an appropriate rating which in turn increased the weight of critical connections.

Nazmi *et al*. [20] developed a phase and phase classification system based on electromyographic signals. In this research, the electromyographic signal templates were extracted from time domain features and inserted it into a classifier for Artificial Neural Networks. In addition, a different number of input functions of the Artificial Neural Network and two important learning mechanisms was identified. The system was further tested using electromyographic signals based on the studied and unknown data to determine the position and phases of vibration and to record the heel impacts and feet timing.

Rudy Division *et al*. [21] considered FERNN as a quick way to separate rules from the trained neurological networks without the need to retrain the network. In the context of a fully integrated network, FERNN first identifies the associated hidden units by calculating their information gain. The corresponding relationships between input units and their hidden units are identified by checking their weights. Finally, FERNN creates rules on the basis of hidden units and weights. Lions Pride Optimization Algorithm (LPOA) was developed by Kaveh *et al.* [22] for optimal design of two-tier cylindrical vaults. To demonstrate the effectiveness of LPOA, three large-scale optimization projects for two-tier wellbore storage were found to be optimal. The results were compared with some measurements in the literature. The second goal of this article was to evaluate these examples using three robust metaheuristic mechanisms i.e., Artificial Bee Colony (ABC), Cuckoo Search (CS) and Particle Swarm Optimization (PSO).

Chan *et al*. [23] implemented Artificial Neural Network model. The two methods are referred to Piece-Wise Linear Artificial Neural Network (PWL-ANN) and the advanced PWL-ANN algorithm. The PWL-ANN algorithm is the rule extraction algorithm for degenerating Artificial Neural Networks (ANNs). The improved PWL-ANN algorithm gets enhanced on the PWL-ANN algorithm and derives more linear regression equations from the ANN model. It was formed by accessing the functions of the latent sigmoid function using linear N equations. The algorithm then provides understandable models from the first-trained ANN opacity models.

Higgins *et al*. [24] used a fuzzy logic model to predict a number function from the models of a function and its dependent variables. This method used a theoretical approach based on the previous work with single value data. The learned rules can be used to predict the functional value in a neural network based on dependent variables.

Ottajima *et al*. [25] presented a GRG (Rule Generation) method to create rules for the classification of data from unique properties. The greedy algorithm sought the best destiny over and over again. The criteria for the best rule were inclusive of number of samples, number of subsections within it and the number of properties of the rule. This method was used to separate the rules of trained and disconnected neural networks to solve classification problems. Classification rules are extracted from neural networks using standard decay approach. Neural networks are trained with hidden layer and the proposed GRG algorithm is used for customized activation values of the hidden device.

Tasaki *et al*. [26] studied the fuzzy rules using neural networks by following two-pronged method of planned landing structure. The neural network consists of a three-phase input phase and a packet system with a hidden level from the output level. In the hidden layer, the neurons are arranged in a rod so that every neuron is assigned a rod. Each neuron in the hidden layer gives fuzzy proposition which creates a fuzzy rule. Structural examination is performed with the generation / destruction of network neurons. Once they understand the rules, they can take simple, fuzzy production rules from the internet.

Fukumi *et al*. [27] evaluated the example of rules based on neural network. The Neural Network (NN) is generally trained using a Genetic Algorithm (GA) to diagnose viral infection. This method used GA modular structure. Each module examines different neural network structures for instance, sigmoid high-sequence neural networks. This chromosomal information is then passed on to other modules through viral infection. High-volume modules are connected to the output module or hidden modules.

Human *et al*. [28] introduced built-in purchasing rules using memorable fuzzy neural networks. For starters who use T. Kohonen's self-regulating computation, they developed a basic routine neurological system with pre-defined smooth notes and less significant recording capabilities. Therefore, Mr. Ishilcawa followed flexible guidelines based on the calculation of ambiguity of the main trend, in which the numbers are defined in the next section. The basic qualities and multiplication of the ability to participate in the last section are partially ignored. Thus, the guidelines for the implementation of a simple general policy is formulated in this study through the elimination of recurring factors.

Shi *et al*. [29] presented a fuzzy planning system in which the rules depend on the procedures of neurological system and computer calculations. In a step-by-step approach, the main limits of fuzzy rules are usually plotted with the help of fuzzy cluster calculation. So flexible guidelines for clearly mischievous in nature since it use neuro-fuzzy learning calculation. With this method, the training time gets reduced and the fuzzy standards generated are relevant to a sensitive and recognized structure. Fan *et al*. [30] devised an approach to develop ambiguous and fuzzy learning rules using neurological systems. After all, the information about the data becomes a neurological system. By disconnecting loads from the nervous system, a network of value records gets structured. At this point, the gentle and thoughtful creative rules are removed from the nervous

system. In order to accurately reflect the information included in neurological system, the calculation for relevant thought was developed. Victor *et al.* [31] developed an ANN training in which two or more ANNSER students coexist in a multi-agent learning environment. The ANNSER learners collaborate with each other using someone else's high quality rules to create new learning experiences. In this way, the generalization of the ANN gets improved which results in a set of high quality rules that describe the knowledge included in the resultant ANN. The article is divided into five sections. Section 1 provides a brief introduction of the system. Section 2 details the records of existing neighboring clubs. Section 3 shows the performance of the system. Section 4 contains results such as discussion and comparison measures. Finally, the paper is summarized with future scope for future researchers.

3. Proposed methodology

The study included a large amount of clinical database for 1197 employees with various illnesses such as Diabetes, Blood Pressure, Headache, Mental Illness, Heart Disease, Gastritis, Stroke and environment. The dataset created by the authors contained 523 employee records with a total of 41 attributes including Diabetes, Blood Pressure, Headache, Mental Illness, Heart Disease, Gastritis, Stroke and environment etc. It is difficult for specialists to predict the illnesses because it is a complex process that requires experience and expertise. The aim of the proposed study is to determine whether an employee is sick or normal due to stress. Using data mining techniques, the authors decide the best optimum model that can provide best percentage of data collection. Figure 1 shows the disease prediction working flowchart based on employee clinical database.
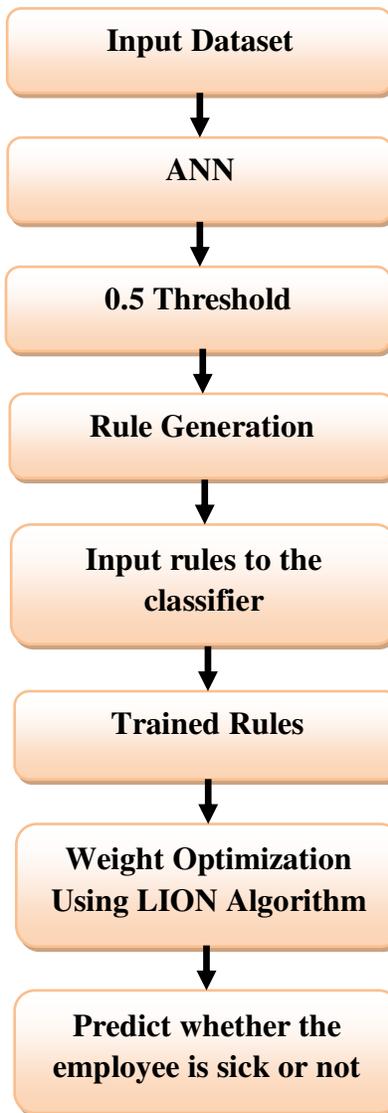
Figure 1: Proposed employee clinical database-based disease prediction

The employee file information was classified and tested using ANN classifier. During classification phase, 34 attributes were fed as input to the ANN classifier to determine the risk of various diseases. In addition, weighted ANN training rules were framed using lion optimization algorithm. The proposed system is intended to provide doctors with a tool that helps in the prediction of disease more effectively. The classifier's performance was compared with the records collected by 1,197 employees. In this article, Optimal Artificial Neural Network (OANN)-based lion optimization method was used to create rules, remove relevant actions from the database, compare generated rules and database details, and predict whether an employee is sick or not. The algorithm was used to obtain accurate results with high efficiency and fast processing.

3.1. Data representation

It is important to present the concentration of input and output characteristics from database comparison elements using OANN-based lion optimization algorithm since it influences the conclusions. Information gathering is generally a mix of different digital, representative, printed and lost properties. Before production, the information is introduced in a structure suitable for the preparation of RNA. At this stage, it eliminates the lost qualities and converts all the properties of the composite into digital properties. The information is collected in the form of works that present input and performance properties as required.

3.2. OANN based lion optimization algorithm

This algorithm is useful in the creation of forecast models and offers a variety of options for viewing and understanding information. OANN-based lion optimization algorithm strategy is an adjustment method, if the characteristics of a particular category are high. This algorithm is simple yet can perform better than other forecasting methods. The OANN-based lion optimization algorithm separates whether the patient is sick or not. In the unpredictable state of a disease, this algorithm finds the possibility of information functionality.

A. OANN-based lion optimization algorithm yields the best results in the following circumstances

1. OANN-based lion optimization algorithm yields the best results, when there is a high probability for the number of properties to be high.

2. If the specified properties are not linked, then this calculation is the best algorithm

3. If the results are accurate, then it is important to use contrast and other optimal methods

3.2.1. Rule extraction

The rule extraction phase provides a range of data including RNA information neurons, h envelope neurons, n -yield neurons, and selected several models T, originally ordered from the training dataset. The standard extraction procedure is as follows.

1) The neural system is designed using special skills with $\{E_1, E_2, E_3, E_4, E_5\}$ as input units, HU_a hidden units and ages f as the output unit.

2) The estimate of the polarization function, proposed for the information level, is owned by equation 1 provided herewith.

$$X = \beta + \sum_{n'=0}^{H_{NH}-1} w_{(n')} E_1(n') + w_{(n')} E_2(n') + w_{(n')} E_3(n') + ... + w_{(n')} E_5(n') \quad (1)$$

The additional equation 2 evaluates the activation capacity for low output conditions.

$$Active\,(X) = \frac{1}{1+e^{-X}} \qquad (2)$$

3) The learning error can be displayed as follows.

$$L = \frac{1}{H_{NH}} \sum_{n'=0}^{N_{NH}-1} D_{n'} - A_{n'} \qquad (3)$$

where, L - learning rate of FFBNN.

$D_{n'}$ - Desired outputs.

$A_{n'}$ - Actual outputs.

In this neural network strategy, a back propagation technique is used as shown in the next section.

❖ **Learning Algorithm – Back Propagation Algorithm**

In Feed Forward Neural System, the back propagation algorithm calculation is actually used as a learning algorithm. In terms of redistribution, it is necessary to ensure that this is a controlled learning method and a factor in the decline of delta law. Many contributions require basic profitability information to update the acquisition. Clearly, back propagation algorithm is suitable for Feed Forward Networks whereas if the calculation needs to be learnt, then the delivery targets used by neurons are unique.

4.5. *Back Propagation Algorithm*

1) The neural charges of the hidden layer and charge exit layer must be chosen arbitrarily. However, the passage level has a constant weight.

2) The standard bias and execution works should be evaluation using equations, (4) and (5).

3) Back propagation error behind each node should be evaluated and then the weight is updated according to the corresponding equation 4.

$$O_{(n')} = O_{(n')} + \Delta O_{(n')}$$

(4)

4) The weight $\Delta O_{(n')}$ is compared to the condition, with which, it becomes equation 5.

$$\Delta O_{(n')} = \delta . X_{(n')} . F^{(BP)}$$

(5)

Where - δ the learning rate which is always between 0.2 and 0.5.

$F^{(BP)}$ - BP Error.

5) The procedure is repeated using the methods provided in equations (4) and (5) until the BP error is limited i.e. $F^{(BP)} < 0.1$.

6) When the baseline is reached, the back propagation of returning to screening phase appears.

As a result, lion optimization effectively qualifies and verifies the association principles using attributes. The final phase is the optimization phase during when the weight optimization based on the OANN algorithm is actually used to guarantee an accurate prediction.
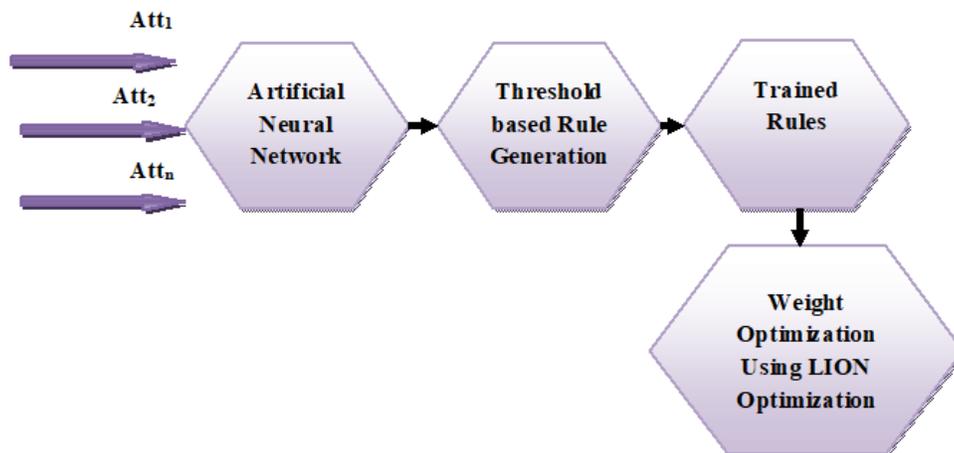

Figure 2: Proposed OANN prediction process

Lion Optimization

Let us consider the objective function in this equation (6).

$$A^{optimal} = \arg\min_{a_i \in (a_i^{min}, a_i^{max})} f(a_1, a_2, \ldots, a_n); \quad n \geq 1 \tag{6}$$

In Eq. (6), f (.) is a continuous unimodal or multimodal function, in which the space of one solution remains the size of the other $\mathscr{R}^n$. Here $\mathscr{R}$ represents the real numbers, a_i·i = 1, 2, ..., n is the ith vector dimension of the ninth variable solution respectively. a_i^{min} and a_i^{max} are the minimum and maximum limits of the ith variable. $a^{optimal}$ is the optimization / destination solution derived from the optimization algorithm as represented in the Eq. (8). The size of the solution region $f(.)$ can be determined as follows:

$$\mathscr{R}^n = \prod_{i=1}^{n}(a_i^{max} - a_i^{min}) \tag{7}$$

$$x^{optimal} = x : f(x) < f(a'|a' \neq a; a_i' \in (a_i^{max}, a_i^{min})) \tag{8}$$

Where A is the vector of the representation arrangement $A = [a_1, a_2, \ldots a_n]$. Equation (6) shows the objective function which can be found as a minimization process. It can sometimes be a retraining work. So a rational calculation requires the use of an appropriate determination process.

Pride generation

According to the sense of pride and condition (6), Pride is presented with local lion A^{male}, its lioness A^{female} whereas A^{normal} is the incorrect example of a lion. Migratory lion is not a person out of pride and it is said so when pride is created. The representation of the lion is equal to that of the vector representation of the arrangement. If the vector elements of A^{male}, A^{female} and A^{normal}, i.e., a_j^{male}, a_j^{female} and a_j^{normal} are self-confirmed integers at the base, the largest limit points are n>1 (search with the original code). Here l = 1, 2, ..., L and L indicates the lion's length, which can be resolved as follows,

$$L = q \begin{cases} p; & n > 1 (general\ case) \\ m; & otherwise\ (Special\ Case) \end{cases} \tag{9}$$

Where, *p* and *q* are integers to determine the lion's length. In the case of $n = 1$, the calculation must be performed using a parallel code lion. In this way, the vector elements are generated as either 1 or 0 based on the constraints indicated in equations (5) and (6) when taken into consideration.

$$h(a_J) \in (a_J^{\min}, a_J^{\max}) \ (10)$$

$$m\%2 = 0$$

Where

$$h(a_J) = \sum_{J=1}^{L} a_J 2^{(\frac{L}{2}-1)} \ (11)$$

There is a claim that the binary lion, created by the terms (10) and (11), is in control area and in the equation (11). The number of binary bits guaranteed is equal to the decimal point. Since different things are being attempted, only through a true calculation of the coded lion, we shall confine ourselves from dealing with the complexities of a double coded lion. All conversations from this stage are clearly a true calculation of the coded lion. Two traveling lions are expected with attempts to enter the area, as the exorcism created A^{normal} which affects one of the two positions of the migratory lions. It introduces other migratory lions during local conservation. Hence, the second sediments show A^{normal} will be $A^{normal}1$ with zero position.

3.3 Fertility evaluation

In the continuous process of lion algorithm, every region and lion start aging or sometimes decompose. This delays the survival or territorial acquisition of the lion. If A^{male} and A^{female} got saturated in their physical form, they would have achieved a better global or regional optimum, from which they could not lead us to better solutions. Fertility assessment can help you skip better local solutions. In this process, A^{male} is calculated as a delay and its laggardness rate is increased by L_r. If A^{male} is greater than f^{ref}, this is nothing but the physical reference form. If L_r exceeds the maximum limit i.e., $L_r^{\max}$, local resistance occurs. The fertility rate A^{female} of the peeling is guaranteed by increasing the S_r infertility rate after the cross. If the tolerance S_r gets exceeded, A^{female} is subjected to update as given in the equation (13). If the updated $A^{female+}$ woman is hired as A^{female} because of her improvement, the pairing process can be implemented. On the contrary,

the update continues until the number of female count g_c reaches g_c^{max} generations. If $A^{female+}$ is not there to replace A^{female} during the upgradation process, A^{female} can always decide that it is better to produce better puppies [11].

$$A_{\int}^{female+} = \begin{cases} a_k^{female+}; & if \int = k \\ a_{\int}^{female}; & otherwise \end{cases} \tag{12}$$

$$A_k^{female+} = min[a_k^{max}, max(a_k^{min}, \nabla_k)] \tag{13}$$

$$\nabla_k = [a_k^{female} + (0.1r_2 - 0.05)a_k^{female} - a_k^{female}] \tag{14}$$

Where $a_{\int}^{female}$ and $a_k^{female+}$ are the $\int th$ and $k th$ vector elements of $a^{female+}$, k is a random number generated in segment *[1, L]*, ∇ is function of the female feminine update and r_1 and r_2 are arbitrary numbers in the array *[0, 1]*.

3.4 Mating

The calculation report of the continent is in line with the principles and another support. Here, the types of hybrid transformation, sexual grouping and other advances are also mentioned. Some articles talk about the hybrid, potential for change and the neglect of developmental calculations. These administrators clearly convince us and coordinate them in our calculations. A^{male} and A^{female} give birth to young cubs with hybrids and mutations. This is the term for A^{male} and A^{female} parts. The most common contaminants such as four litters in the lion's stomach (mainly) were traced and the hybrid process produced four cubs.

The crossover mask B was modified for each cub, i.e. p thB_p mask was used to get $A^{cubs}(p)$. These four cubs produce four more new cubs. From now on, we use " 'A^{cubs}' to designate cubs from cross and 'A^{new}' to represent mutations. These eight cubs are in the cub pool to complete a^{m_cub} and a^{f_cub}. As a result, a^{m_cub} and a^{f_cub} follow to automatically update the bear growth function [11].

3.5 Lion operators

Local resistance [10, 11] encourages the non-compliance with the examination of storage space on the base. This is because it allows one to calculate the points from an approximate point of view

and to distribute the differences in details of relative similarity. The local opposition was resolved in a moment of union of immigrants [21, 22], of recovery by resistance [23] and of pride and compensation for the individual. The a^{e_normal} reorganizes the nomad coalition with a single trial period [24]. Then a^{e_normal} is selected whether in Eq. (10) - (12) and the process is completed.

$$f(a^{e\_normal}) < (a^{male}) \tag{10}$$

$$f(a^{e\_normal}) < (a^{m\_cub}) \tag{11}$$

$$f(a^{e\_normal}) < (a^{f\_cub}) \tag{12}$$

A^{male} meets the goodness of the day, while wandering and if he meets the legendary $A^{e\_normal,}$ it vanishes. Pride setting is a replaceable i.e., A^{male} by A^{e_normal}. An update of the indivisible A^{normal}, which has E^{normal} solo travel partnership is electronic selection or university membership (see Appendix 2 of the refer theorem) [11]. When xm _cub and xf _cub are fully developed, this is when the dogs are larger than younger cubs, A^{male} and A^{female} if a^{m_cub} and a^{f_cub} must be renewed to receive terrestrial protection [10].

3.6 Termination

The term algorithm is the universe of two criteria associated with, which always gets satisfied.

$$N_g > N_g^{max} \tag{13}$$

$$\left| f(a^{male}) \_ f(a^{optimal}) \right| \le e_T \tag{14}$$

In other words, N_g is the number of generations, part of the zeal and if it is double in memory, N_g^{max} while e_T is the most uniform number of groups of orgies or the margin of error $|\cdot|$ There is only one difference. If the equation is written as Eq. (14), then all these very simple reasons explain the perfection of cable for the optimal $f(a^{optimal})$ (or the most extreme), just like $f(a^{optimal})$ being important to $a^{optimal}$. This Optimal Associative Neural Network Optimization (OANN) method can be used to create rules, remove relevant actions from the database, compare generated rules and

database details, and predict whether an employee is sick or not. The algorithm is used to obtain accurate results with high efficiency and fast processing.

4. Dataset description

The authors developed a new stress prediction dataset with 1,197 records of employees collected from school, bank and universities etc. who belonged to different institutions. This dataset was filtered to have 523 records of employees with various diseases and work pressure. A total of five independent studies with different recruitment strategies was found. The designs were inclusive of 34 input attributes, 7 target attributes i.e., totally 41 attributes.

Table I: Dataset

Attribute	Values
Age	1.20-30,2.31-40,3.41-50,4.Above 50
Sex	1. Male, 2. Female
Marital status	1.Single 2.Married
Highest Qualification	1.Below SSLC, 2.Graduate,3.Post Graduate, 4.Technical Qualification
Working Environment	1. City, 2. Village
Nature of works	1.Physical,2.Mental,3.Both
Working Sector	1.Govt.2.Private,3.Business
Occupation	1.Teacher/Professor, 2.Doctor, 3.Engineer, 4.Bank Employees, 5.Business, 6.Labours, 7.others
Working hours	1.6hrs, 2.8hrs, 3.>8hrs
Overtime hours	1.2hrs, 2.5hrs, 3.>5hrs
Job Satisfaction	0. Yes, 1. No
Working Experience	1.2-5 years,2.6-10 years,3.>10 years
Sufficient Income	0. Yes, 1. No
Work Pressure	0. Yes, 1. No
Addiction	1. Smoking, 2. Alcohol, 3,Tobacco
Food	1. Veg., 2.NV, 3. Both
Sleeping hours	1.<6hrs, 2.6-8hrs, 3.>8hrs
Physical Activities	1. Yes, 2. No
Type of roles	1. Major, 2. Minor
Leisure Time	1.Reading books, 2.Listening Music, 3.Watching TV, 4.Playing chess

Diabetes	0. Yes, 1. No
Blood Pressure	0. Yes, 1. No
Headache	0. Yes, 1. No
Mental Illness(depression)	0. Yes, 1. No
Heart Disease	0. Yes, 1. No
Gastritis(Ulcer)	0. Yes, 1. No
Stroke	0. Yes, 1. No
Exercise regularly	0. Yes, 1. No
Any continuous medication	0. Yes, 1. No
how long suffer	1.6 months,2.1 year,3.>1year
Body Weight	1.<40kg, 2.41-60kg, 3.>60kg
Feel tired or depressed	0. Yes, 1. No
Working conditions	1.satisfactory, 2.disatisfactory, 3.Can't say
better on the job if conditions changed	0. Yes, 1. No
job affect your family	0. Yes, 1. No
Control over life	0. Yes, 1. No
Any argument	0. Yes, 1. No
Activity for stress relief in organization	0. Yes, 1. No
Are underpaid	0. Yes, 1. No
Are undervalued	0. Yes, 1. No
Appreciation for good work	0. Yes, 1. No

When each participant provides data for at least two days, it allows them to explore the differences between themselves in different days. Therefore, the authors included everyone who submitted at least two complete daily entries (i.e., they responded the morning and afternoon surveys within at least two days). There were inconsistencies in the original data. It had to be rejected due to errors and inconsistencies. The data was converted to excel sheets. Thus, the data types were edited and changed. Missing values in the data were cleaned up and compiled before using the function selection in the database. The data was then used to determine the key attributes and obtain the appropriate subset of key attributes to be used in classification exercise. These files were later generated and converted to ARF format compatible with Data Mining Toolkit that is generally used to create the model. The data was split into three data sets. The first dataset was used to train the sample whereas the second and third datasets were used to validate the dataset model. Group

comparisons of employees and disease characteristics between groups showed that there is a high percentage of employees reporting adverse symptoms in the workplace as shown in table I. In addition, the participants of normal and work stress groups reported significantly higher prognosis and a higher degree of disability in all the functional areas in terms of productivity (prognosis of the disease based on working conditions).

5. Result and discussion

Various parameters (Diabetes, Blood Pressure, Headache, Mental Illness, Heart Disease, Gastritis, Stroke) were considered in the prediction of diabetes for patient. Both OANN and time series techniques were compared with manual prediction that evaluates the performance of the current study forecasting system. In this work, the authors considered three standard evaluation parameters such as accuracy, sensitivity and specificity of the classification and are defined as follows:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$$

$$Sensitivity = \frac{TP}{TP+FN}$$

$$Specificity = \frac{TN}{TN+FP}$$

Where TP; FP; TN; FN is True Positive, False Positive, True Negative or False Negative. In this study, it is assumed that accuracy is the best entity.

Table II: Comparison of the proposed and existing sensitivity measures

| Threshold	OANN	ANN
0.6	0.907	0.693
0.5	0.907	0.742
0.4	0.907	0.704
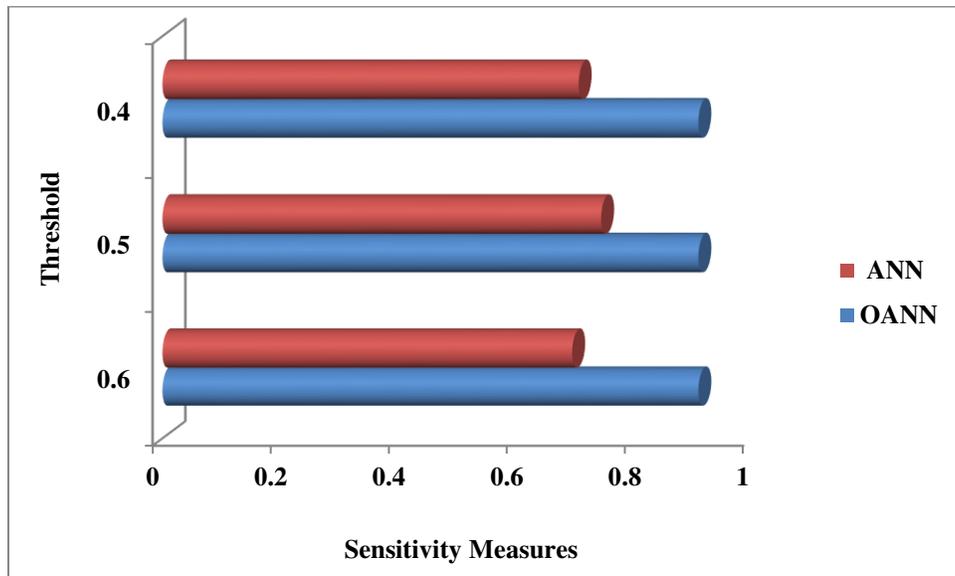
Figure 3: Graphical representation of the proposed and existing sensitivity measures

In table II and figure 3, when the proposed and existing investigations for sensitivity measures were compared, the proposed OANN attained better results.

Table III: Comparison of the proposed and existing specificity measures

Threshold	OANN	ANN
0.6	0.912	0.717
0.5	0.912	0.7637
0.4	0.912	0.786
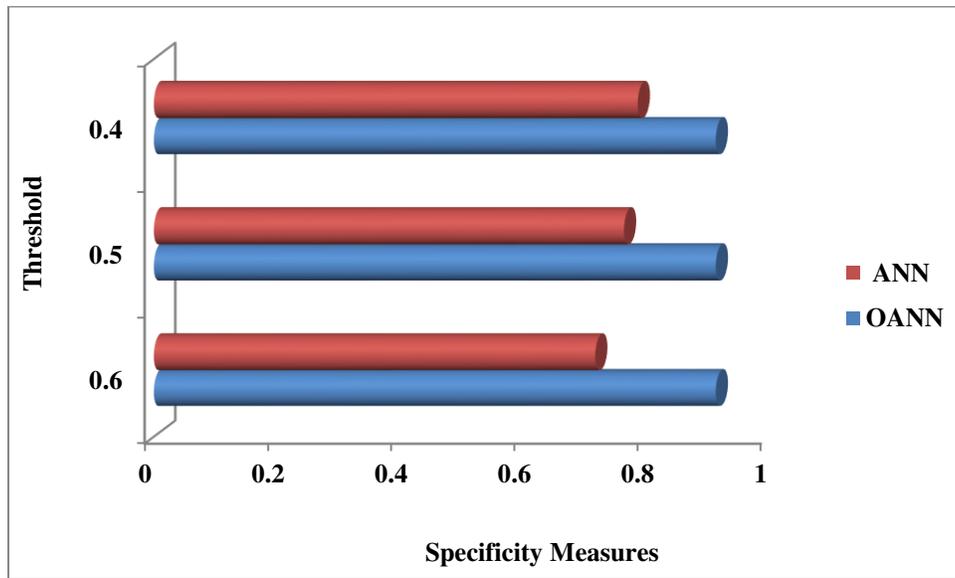
Figure 4: Graphical representation of the proposed and existing specificity measures

In table III and figure 4, the authors explained the specificity computation for which OANN achieved 0.912 % compared to the specificity measures achieved by existing ANN. The proposed method yielded better results.

Table IV: Comparison of the proposed and existing accuracy measures

Threshold	OANN	ANN
0.6	0.909	0.703
0.5	0.909	0.751
0.4	0.909	0.7376
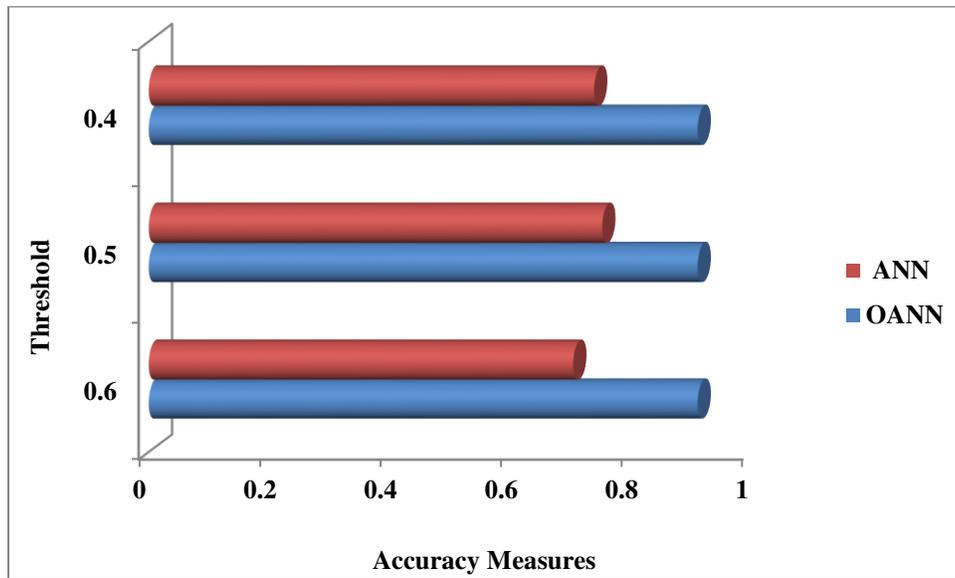
Figure 5: Graphical representation of proposed and existing Accuracy Measures

In table IV and figure 5, the accuracy computation is explained in which OANN achieved 0.909% compared to the accuracy measures achieved by existing ANN. The proposed method attained better results.

Table V: Threshold-based rule generation

Threshold	Number of Rules
0.6	187
0.5	640
0.4	2397
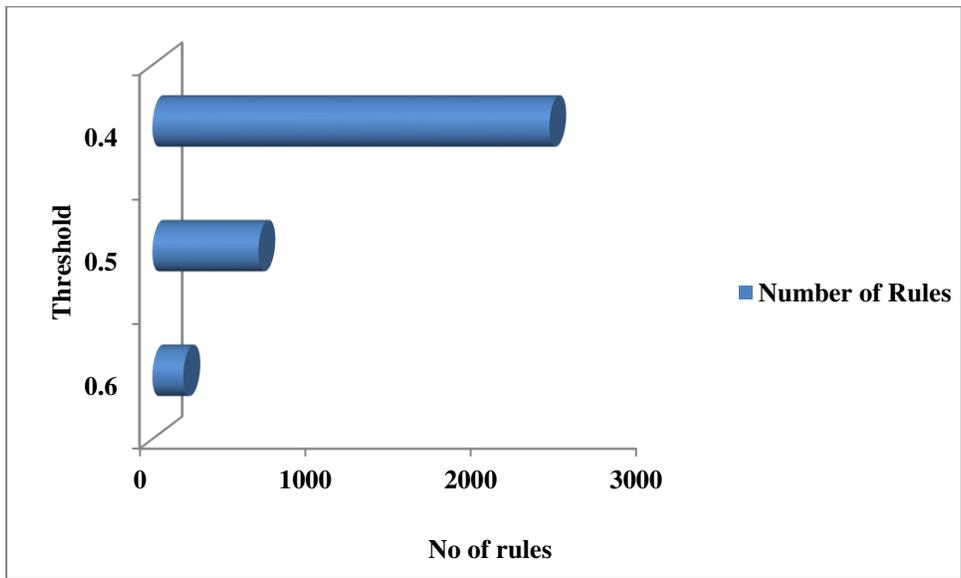
Figure 6: Graphical representation of rule-based graph generation
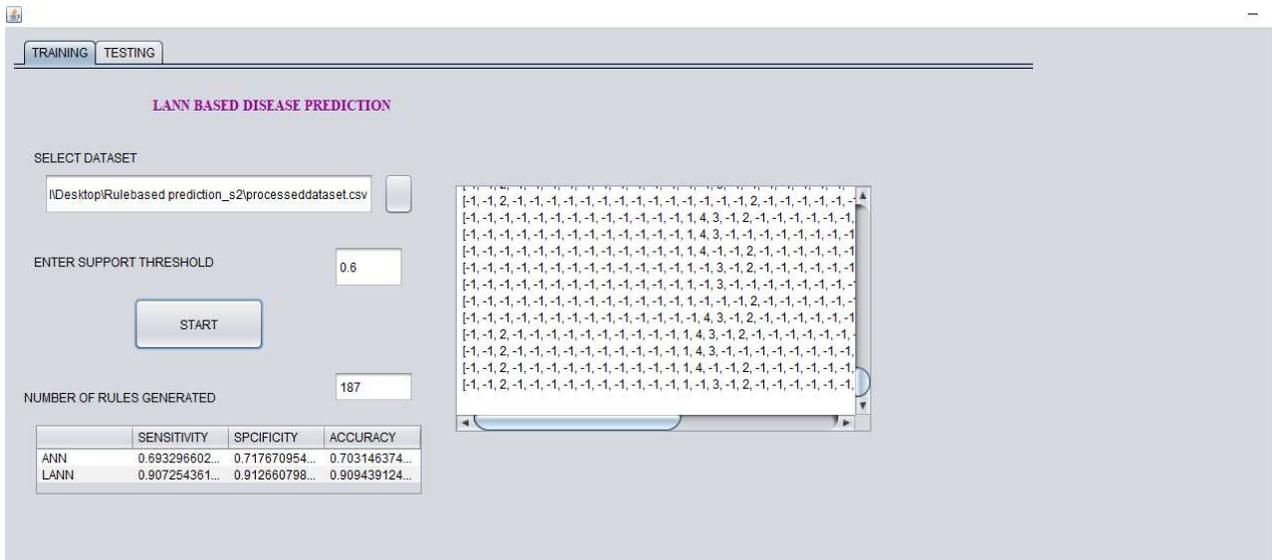
Figure 7: JAVA output of rule-based disease prediction using ANN and OANN techniques

In overall threshold-based rule generation, the proposed system generated rules such as 187, 640 and 2397 based on the thresholds 0.4, 0.5 and 0.6 of OANN. Figure 7 shows the performance analyses of JAVA output in rule generation-based ANN and OANN-based lion optimization algorithm technique with disease prediction. Here the number of rules generated i.e., 43 and threshold 0.6 were evaluated. The current study proposed stress predictions using data mining approaches based on the data of employees' who work in the management of schools, banks and universities. In tables II, III and IV, the results of the proposed approaches are summarized without

displacement but with all original data. The results show that the sample weighted and group weighted input approaches performed better than the non-transfer approach. The thresholds and models approach received similar scores and the group approach received better scores. This increased the accuracy by an average of 10%.

6. Conclusion

In this research field, many researchers have focused on predicting different diseases using different work applications. In this manuscript, OANN-based lion optimization technique was evaluated for stress-based disease prediction among working employee. ANN method is used for feature selection which eliminates unnecessary and inappropriate attributes. Lion optimizer classifies various diseases and avoids dimensional complications. The current study compared the classification performance of the proposed method with that of the existing ANN method. The performance of the OANN-based lion optimization system was determined on the basis of various evaluation criteria such as specificity, sensitivity and accuracy. The OANN-based lion optimization method achieved 0.909 with high accuracy for disease prognosis. In future, research activities may be expanded into effective hybrid technique so as to improve the effectiveness of stress prediction based on employees working in different work environments.

Author contributions :

S.Anitha : Conceptualization, Data curation and Formal analysis, Investigation and Methodology, Validation and Visualization, Writing-original draft

Dr.M.Vanitha : Conceptualization, Project administration and Resources, Validation and Visualization, , Writing-review and editing

All the authors have read and agreed to the publication of this manuscript.

Compliance with ethical standards

Conflict of interest: The authors declare that they have no conflict of interest.

Human and animal rights: This article does not contain any studies with animals performed by any of the authors.

References

1.   Han, Sang-Sook, Jeong-Won Han, Young-Suk An, and So-Hee Lim, "Effects of role stress on nurses' turnover intentions: The mediating effects of organizational commitment and burnout", Japan Journal of Nursing Science, vol. 12, no. 4, pp. 287-296, 2015

2. Garcia-Ceja, Enrique, Venet Osmani, and Oscar Mayora, "Automatic stress detection in working environments from smartphones' accelerometer data: a first step", IEEE journal of biomedical and health informatics, vol. 20, no. 4, pp. 1053-1060, 2015

3. Maxhuni, Alban, Pablo Hernandez-Leal, L. Enrique Sucar, Venet Osmani, Eduardo F. Morales, and Oscar Mayora, "Stress modelling and prediction in presence of scarce data." Journal of biomedical informatics, vol. 63, pp. 344-356, 2016.

4. Havermans, Bo M., Evelien PM Brouwers, Rianne JA Hoek, Johannes R. Anema, Allard J. van der Beek, and Cécile RL Boot, "Work stress prevention needs of employees and supervisors", BMC public health, vol. 18, no.1, pp. 642, 2018

5. Fohr, Tiina, Julia Pietila, Elina Helander, Tero Myllymäki, Harri Lindholm, Heikki Rusko, and Urho M. Kujala, "Physical activity, body mass index and heart rate variability-based stress and recovery in 16 275 Finnish employees: a cross-sectional study", BMC public health, vol. 16, no. 1, pp.701, 2016.

6. Hessels, Jolanda, Cornelius A. Rietveld, and Peter van der Zwan, "Self-employment and work-related stress: The mediating role of job control and job demand", Journal of Business Venturing, vol. 32, no. 2, pp. 178-196, 2017

7. Mackay, Michael M., Joseph A. Allen, and Ronald S. Landis, "Investigating the incremental validity of employee engagement in the prediction of employee effectiveness: A meta-analytic path analysis", Human Resource Management Review, vol. 27, no. 1, pp. 108-120, 2017.

8. Mauss, Daniel, Marc N. Jarczok, and Joachim E. Fischer, "A streamlined approach for assessing the Allostatic Load Index in industrial employees", Stress, vol. 18, no. 4, pp. 475-483, 2015.

9. Q. Xiao, J. Luo and J. Dai, "Computational Prediction of Human Disease- Associated circRNAs Based on Manifold Regularization Learning Framework," in IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 6, pp. 2661-2669, Nov. 2019.

10. Hara, Atsushi, and Yoichi Hayashi, "Ensemble neural network rule extraction using Re-RX algorithm", The 2012 International Joint Conference on Neural Networks (IJCNN), IEEE, 2012.

11. Wang, Qi, Bo Yu, and Jie Zhu, "Extract rules from software quality prediction model based on neural network", 16th IEEE International Conference on Tools with Artificial Intelligence, 2004.

12. Jayaprabha, D., and K. Nirmala, "Efficiency stress prediction in BPO industries using hybrid k-means and artificial bee colony algorithm", International Journal of Computers and Applications, vol.42, no. 1, pp. 9-16, 2020.

13. Laijawala, Vidit, et al, "Classification Algorithms based Mental Health Prediction using Data Mining", 2020 5th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2020.

14. Pakarinen, Tomppa, Julia Pietila, and Hannu Nieminen, "Prediction of Self-Perceived Stress and Arousal Based on Electrodermal Activity", 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019.

15. McCarthy, Julie M., Berrin Erdogan, and Talya N. Bauer, "An interpersonal perspective of perceived stress: Examining the prosocial coping response patterns of stressed managers", Journal of Organizational Behavior, vol. 40, no.9-10, pp. 1027-1044, 2019.

16. M. Chen, Y. Hao, K. Hwang, L. Wang and L. Wang, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities," in IEEE Access, vol. 5, pp. 8869-8879, 2017.

17. H. Li, X. Li, M. Ramanathan and A. Zhang, "Prediction and Informative Risk Factor Selection of Bone Diseases," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 12, no. 1, pp. 79-91, 1 Jan.-Feb. 2015.

18. P. Luo, L. Tian, J. Ruan and F. Wu, "Disease Gene Prediction by Integrating PPI Networks, Clinical RNA-Seq Data and OMIM Data," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 16, no. 1, pp. 222-232, 1 Jan.-Feb. 2019.

19. J. Luo, P. Ding, C. Liang, B. Cao and X. Chen, "Collective Prediction of Disease-Associated miRNAs Based on Transduction Learning," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 14, no. 6, pp. 1468-1475, 1 Nov.-Dec. 2017.

20. Nazmi, Nurhazimah, Mohd Azizi Abdul Rahman, Shin-Ichiroh Yamamoto, and Siti Anom Ahmad, "Walking gait event detection based on electromyography signals using artificial neural network", Biomedical Signal Processing and Control, vol. 47, pp. 334-343, 2019.

21. R. Setiono and Wee KhengLeow, "Generating rules from trained network using fast pruning," IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No.99CH36339), Washington, DC, USA, 1999, pp. 4095-4098 vol.6, doi: 10.1109/IJCNN.1999.830817.

22. A. Kaveh, S. Mahjoubi, "Optimum Design of Double-layer Barrel Vaults by Lion Pride Optimization Algorithm and a Comparative Study", In Proceedings of ELSEVIER Journal of Centre of Excellence for Fundamental Studies in Structural Engineering, Iran University of Science and Technology, January 2018, pp. 213–229.

23. Chan, Veronica KH, and Christine W. Chan, "Towards explicit representation of an artificial neural network model: Comparison of two artificial neural network rule extraction approaches", Petroleum, 2019.

24. Higgins, C. M., and R. M. Goodman. "Learning fuzzy rule-based neural networks for function approximation." Proc. of IJCNN. Vol. 1. 1992.

25. Odajima, K., Hayashi, Y., Tianxia, G. and Setiono, "Greedy rule generation from discrete data and its use in neural network rule extraction", Neural Networks, vol. 21, no.7, pp. 1020-1028, 2008.

26. Tazaki, Eiichiro, and Norimasa Inoue, "A generation method for fuzzy rules using neural networks with planar lattice architecture", Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94), vol. 3, 1994.

27. Fukumi, Minoru, Yasue Mitsukura, and Norio Akamatsu, "A new genetic approach to universal rule generation from trained neural networks", 2000 TENCON Proceedings. Intelligent Systems and Technologies for the New Millennium (Cat. No. 00CH37119), vol. 1, 2000.

28. Umano, M., Fukunaka, S., Hatono, I. and Tamura, H, "Acquisition of fuzzy rules using fuzzy neural networks with forgetting", Proceedings of International Conference on Neural Networks (ICNN'97), vol. 4, 1997.

29. Shi, Y., M. Mizumoto, and P. Shi. "Fuzzy if-then rule generation based on neural network and clustering algorithm techniques." IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering. TENCOM'02. Proceedings. Vol. 1, 2002.

30. Fan, Tie-gang, Shu-tian Wang, and Jun-min Chen, "Generating weighted fuzzy production rules using neural networks", International Conference on Machine Learning and Cybernetics, 2006.

31. Viktor, H. L., and Ian Cloete, "Improved generalisation using cooperative learning and rule extraction." IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339), vol. 2, 1999.

32. S.Anitha, Dr.M.Vanitha , "Imputation Methods for Missing Data for a Proposed VASA Dataset", International Journal of Innovative Technology and Exploring Engineering (IJITEE),ISSN: 2278-3075,Volume-9 Issue-1,November 2019.

33. S.Anitha, Dr.M.Vanitha,"The Predicting Diseases of Employees with VASA Dataset using Entropy", International Journal of Advanced Science and Technology Vol. 29, No. 4, (2020), pp. 5867 – 5874.

34. S.Anitha, Dr.M.Vanitha,"Selecting Features Related To Work-Pressure And Assessing Possible Diseases In Humans", International Journal Of Scientific & Technology Research Volume 8, Issue 12, December 2019 Issn 2277-8616.

Figures
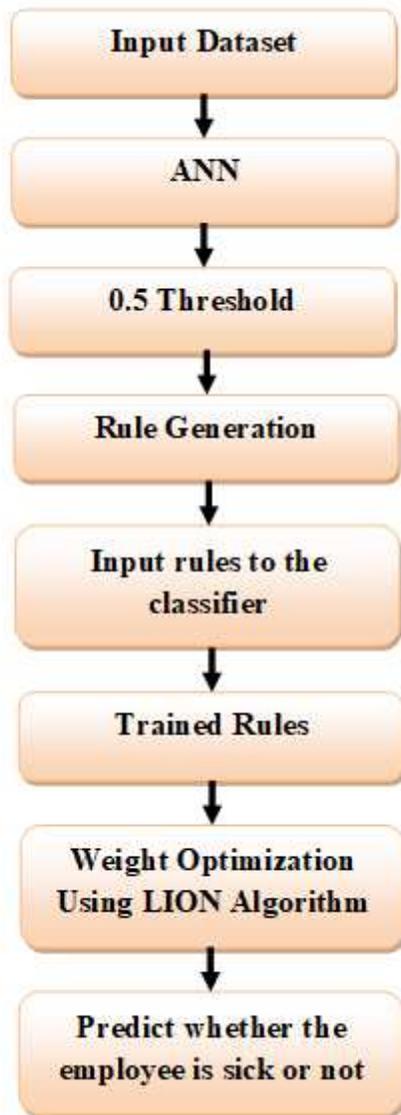
Figure 1

Proposed employee clinical database-based disease prediction
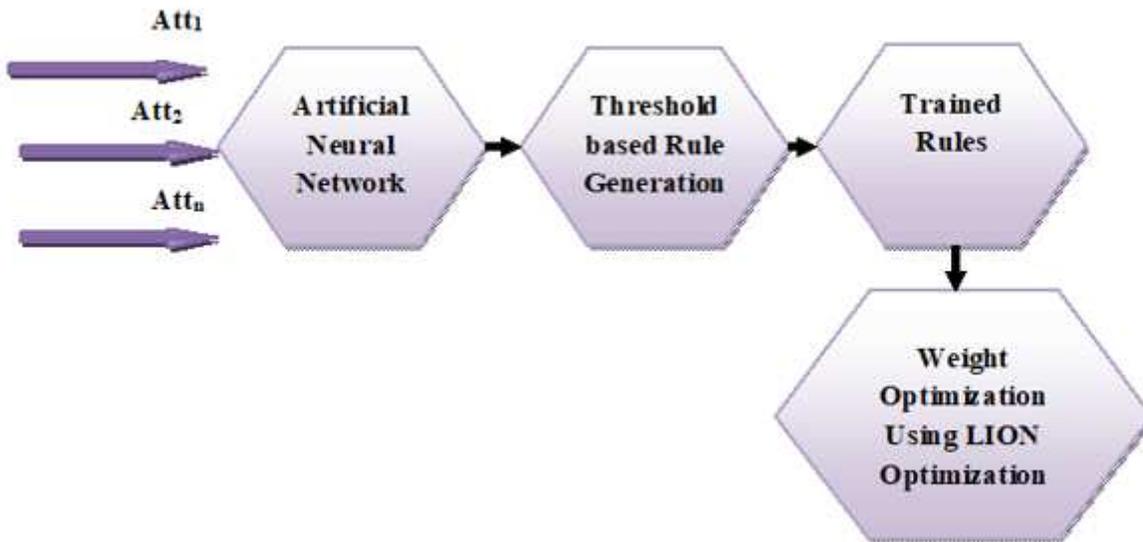
Figure 2

Proposed OANN prediction process
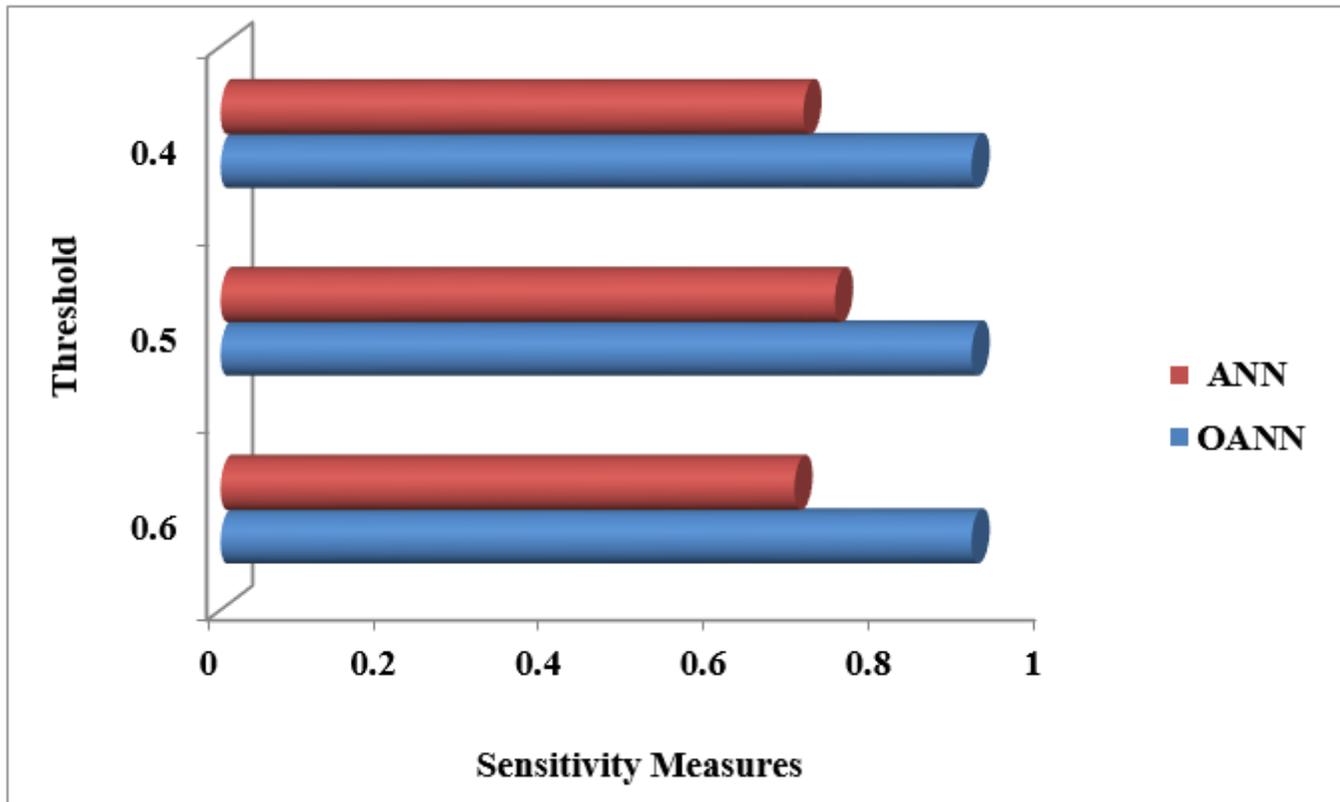
Figure 3

Graphical representation of the proposed and existing sensitivity measures
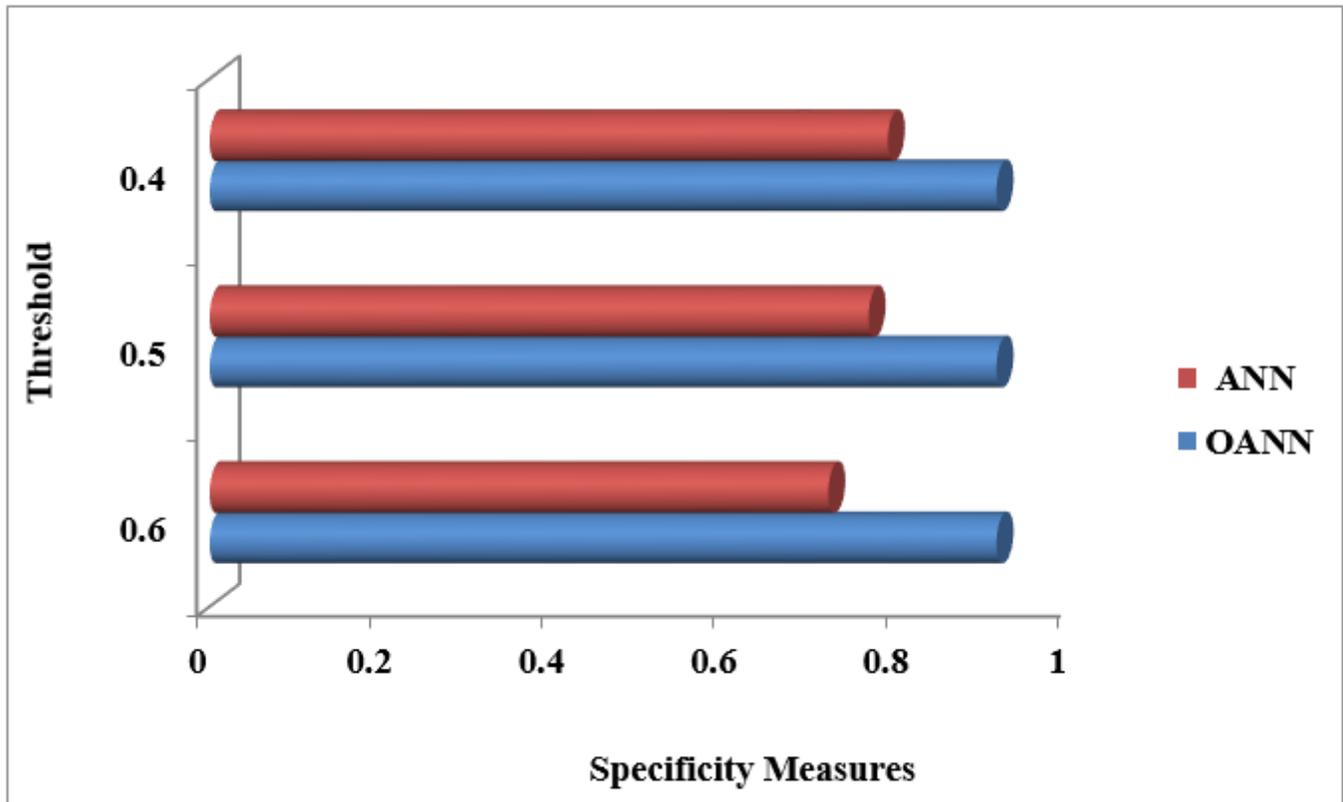
Figure 4

Graphical representation of the proposed and existing specificity measures
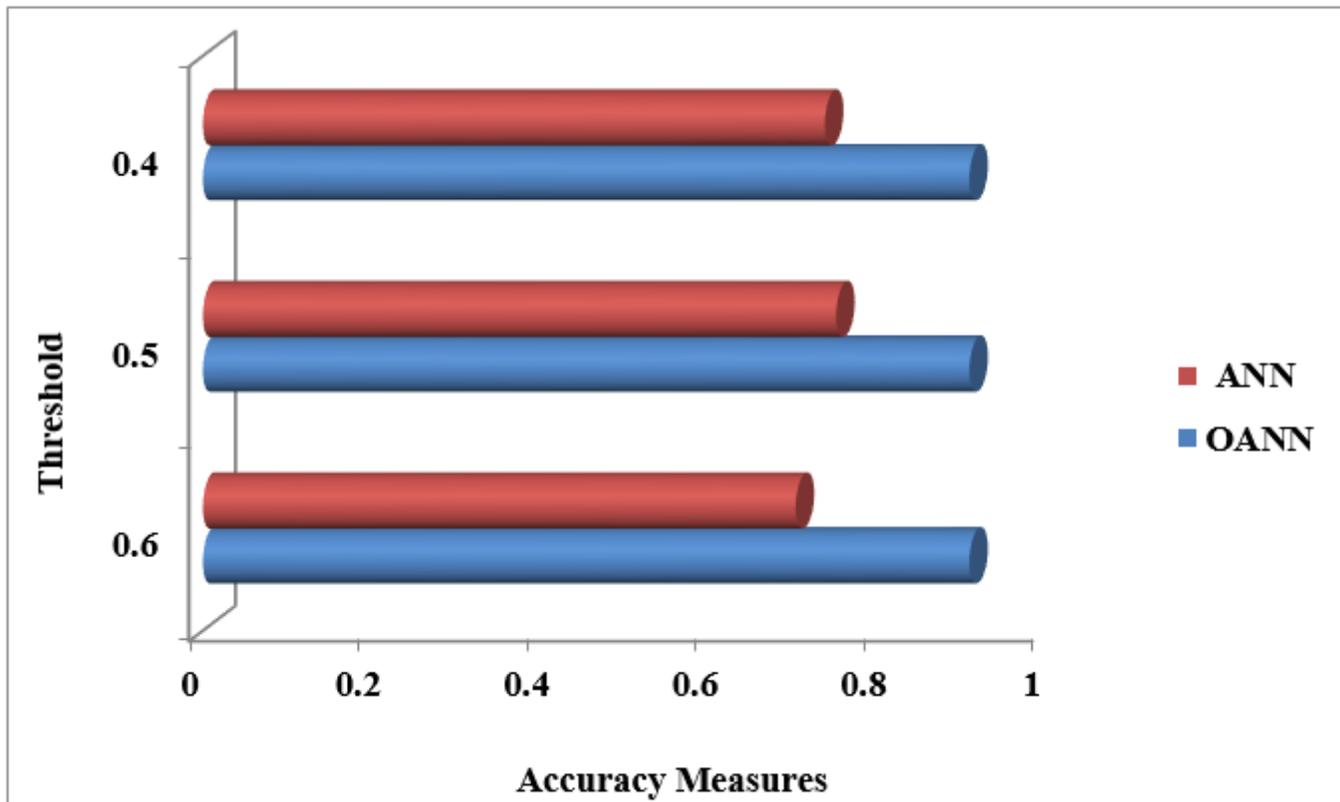
Figure 5

Graphical representation of proposed and existing Accuracy Measures

Figure 6

Graphical representation of rule-based graph generation

Figure 7

JAVA output of rule-based disease prediction using ANN and OANN techniques