

Prediction of Heart Disease using Name Entity Recognition based on Back Propagation and Whale Optimization Algorithms

Velmurugan Thambusamy, Latha Umasankar

Abstract: Objectives/Backgrounds: Nowadays, heart diseases play a very big role in the universe. The Physicians in practice gives various names for heart diseases such as heart attack, cardiac attack, cardiac arrest etc. Among the computerized methods to find the heart disease, Named Entity Recognition (NER) algorithm is used to find the synonyms for the heart disease text to mine the meaning in medical reports and various applications. **Methods/Statistical Analysis:** The Heart disease text input data given by the physician is taken for the preprocessing and changes the input content to the desired format, then that resultant output fed as input for the prediction. This research work uses the NER to find the meanings of the heart disease text data and uses the existing two methods Deep Learning Models and whale optimization are combined and proposed a new method Optimal Deep Neural Network (ODNN) for predicting the disease. **Findings:** For the prediction, weights and ranges of the patient affected data via selected attributes are chosen for the analysis. The result is then classified with the Deep Neural Network to find the accuracy of the algorithms. The performance of ODNN is evaluated by means of classification measures such as precision, recall and f-measure values. **Improvement:** In future, the other classification algorithms or other text data algorithms were used to find for large amount of text data.

Index Terms: Accuracy, Back Propagation, Named entity recognition, Sensitivity, Specificity, Whale optimization Algorithm.

I. INTRODUCTION

The biomedical text mining is revised regularly. The special survey journal indicates the general-purpose text algorithm and data mining tools are not well defined for the biomedical domain because it is highly specialized. The information tells from [5,18] it's about away to explore, query, analyze and manage the underutilized information. Lot of bio medical research informs the rich deposit of knowledge for bio-medical research. Text mining gives the knowledge from billions of text and practical in biomedical research. It has many computational techniques such as machine learning, natural language processing to find the unstructured biomedical text. To define the useful text mining tasks for the specific goals of biologists, biomedical text mining and clinicians are better positioned [22].

Revised Manuscript Received on July 20, 2019.

Dr. T. Velmurugan, Associate Professor, PG and Research Department of Computer Science, D. G. Vaishnav College, Arumbakkam, Chennai-600106, India.

U. Latha, Assistant Professor, PG and Research Department of Computer Science, D. G. Vaishnav College, Arumbakkam, Chennai-600106, India.

In the biomedical science, the extraordinary growth in both the experimental and literature which is difficult for bio-medical research people to gain novel knowledge plus information, this leads in loss of hypotheses into billions of text data. This became one of the biggest topics in Biomedical Text mining. To find new systematic knowledge or hypothesis among the biomedical data, have core tasks of bio-medical Text Mining are to recognize the biomedical named entities expose the interconnected relations of these bio-medical data [25]. To change the unstructured data into structured data, all type of the text data info despite of the format can be altered to numbers indexed for every of documents used in text data mining [4]. Nowadays text mining plays a big role in the disciplines like machine learning, data mining, information retrieval and statistics etc., and which is used to classify, cluster, summarize, correlation, distribution for a large number of dataset [24].

Text mining called as knowledge discovered form text and it is a branch of data mining. It is defined as computational procedure of taking out the important information from the big amount of unstructured text data [8]. Text mining is also used in variety of applications like cyber security such as scam detection and interruption detection etc. and to resolve the purchaser related data such as customer attainment, market basket study etc. [11]. Text mining has the very strong robustness to detect the noise and irregular structure in text data [21]. NER is also used to find the vocabulary like CoPub in liver pathology terms [7].

To recognize disease and human genes text and extract the genes there is a open-source text mining software and it is flexible technology to apply in various tasks in medical and biology [23]. Computational behavior of options, sentiment and bias of text used in Sentiment Analysis (SA) and it is current area in text data mining, also needed for many other natural languages [14]. Nowadays, it is one of the important same kinds of major issue is pattern finding or clustering in text mining [3]. It is also used for getting predictive outcomes, intelligent alert systems and sustaining the medical in the decision-making process [17]. NLP on top of separate datasets to mine the features which represents the common info. It is combined with machine learning techniques to get better the categorization of social medium text containing medical text data [20].

An unsupervised machine learning model for discover hidden communicable disease using social media text data is discussed by Sunghoon Lim et al. [12] and it has presented a method called unsupervised machine learning model to find out hidden communicable diseases without given info, such that the name of the diseases and their symptom. In that journal, a national public health institution and communication with the general public gives a hidden communicable disease was defined as infectious disease that has not yet been dignified. A research study was done by Vinay Kumar Jain et al. [9] in their work. An advanced framework detector used for users in Multilanguage text data using emotion theories that deals with linguistics and psychology.

Emergency department (ED) congestion is a severe issue for hospital. A research work by Filipe R. Luciniet al. [6] tells to predict future hospitalizations and discharges, a text data mining method to procedure information from early on crisis patient account using the SOAP framework. They are tried a range of approach for preprocessing of text data account and to expect hospitalization. Binary demonstration, term occurrence, and term occurrence-inverse document occurrence were used to receiving a set of words. An analysis of Ebola on US soil trigger extensive fear. In reply, the Centers for Disease Control and Prevention held a live Twitter converse to speak to public concern. Another work carried out by Allison J. Lazardet al. [1] in their research work. A text analysis process to disclose insight from those tweets that can tell communiqué strategy. User-generated tweets were composed, sort, and analyze is main theme of their article.

A research work carried out by Jitendra Jonnagaddala et al. [10] tells from unstructured electronic health data, a method called clinical text mining to extract Framingham risk factors are used. That is also used for the calculating the diabetic patients for 10-year coronary artery illness risk score. With the help of gender, age, HDL-C, total cholesterol, Smoking history, diabetes history and blood pressure, they will find risk factor. Melissa Ailem *et al.* [15] has presented a general structure which we used to exemplify relations among 10 genes report linked with asthma by a preceding GWAS. The purpose is to influence unverified text data mining technique using text-based cos resemblance comparison and cluster applied to candidate and arbitrary gene vectors, in order to expand the GWAS outcome.

Another research work carried out by M. A. Jabbar *et al.* [13] in their research work. A method is there to discover organization policy in medical text data to find heart infection for Andhra Pradesh. That comes up to be expected to help physicians to make precise choice. Mortality info from the registrar general of India shows that the coronary heart disease (CHD) were a main reason of death in India. P K. Anooj [16] have planned a biased fuzzy rule-based clinical decision support system (CDSS) was obtainable for the analysis of heart disease, mechanically obtain knowledge from the patient's medical data. Section II discusses about materials and methods used for this research work, that is traditional Neural Network, Deep Neural Network and Optimal Deep Neural Network (ODNN) to find accurateness. The experimental results are illustrated in the section III. Finally, this research work concludes with its innovative information in section IV.

II. MATERIALS AND METHODS

The problem definition of this research work is discussed in this section. The major issue in medical text data removal jobs is organization assorted nature of formless text in the medical data. The precise discovery of illness rank from medical text requires accepting of pattern and keywords in a subject's clinical history, which can differ extensively. The accessibility of gigantic amount of clinical text data lead to require for influential data investigation tackles to take out useful information. The dataset consists of redundant data, missing data and irrelevant attributes is preprocessed by means of name entity recognition and resultant data is stored in a text file named sentiwordnet and the cleaned heart disease dataset given in an Optimal Deep Neural Network(ODNN) to predict which patient is affected heavily and mildly with the help of weights that are taken from the usage of the ODNN. This research work has more than 5000 records and which took the records of 400 patients and analyzed. The records are taken from Ashwin Clinic, Anna Nagar, Chennai which is famous for the heart disease.

A. Name Entity Recognition

Named Entity Recognition (NER), or single data removal is an NLP method which locate and classify the named data there in the text. NER classify the named data into predefined category such as the names of people, organization, locations, quantity, financial values, particular terms, product terms and expressions of times.

B. Optimal Deep Learning Model

The selected feature or text from the NER is given to the input for prediction phase. In these predictions, it is used the Optimal Deep Neural Network (ODNN). In our proposed method, the traditional deep neural network is adapted by means of optimization method. The whale optimization is utilized to optimize the parameter of deep neural network to learn high-level feature representations, capture long-term dependencies, and global features to help identify clinical entities. An ANN model with the multi layers of the hidden units and outputs is termed Deep Neural Networks. Moreover, it contains of pre-training (using DBN) and fine tuning phases in its restriction learning. The important goal of this research paper is to training the structures in the precise dataset, i.e. to finding the correct weight that can be used to rightly predict the text. The detailed process is described in further section.

Pre-training phase: In this training phase, it uses a Deep Belief Network (DBN) that is a profound architecture and typical it is forward network in which the dataset flows from the input layer to the output layer over lot of hidden layers which are more than two to three layers. This model certificates the network to yield evident beginnings on the basis of its hidden layers states, which typifies the network belief. Here, it is implemented with Restricted Boltzmann Machine (RBM) for pre-training process.

Restricted Boltzmann Machine: An RBM is a high-class category of Markov random field that has one layer of stochastic hidden pieces and one layer of stochastic visible or observable pieces.

Phase 1: Primarily, it initializes the visible pieces V to the training vector.

$$E(v, h) = -\sum_{x=1}^X \sum_{y=1}^Y W_{xy} v_x h_y - \sum_{x=1}^X \chi_x v_x - \sum_{y=1}^Y \xi_y h_y \quad (1)$$

Where, W_{xy} signifies the symmetric communication term between the visible piece v_x and the hidden piece h_y , χ, ξ is the bias term, X, Y is the digit of visible and hidden pieces. The derived of the log probability of a training vector regarding a weight is unpredictably humble. Between hidden pieces in an RBM, there are no straight effects; it is extremely easy to get an unbiased model of $(V_x, h_y)_{data}$

$$\rho(h_y = 1|v) = \delta\left(\sum_{x=1}^X W_{xy} v_x + \chi_x\right) \quad (2)$$

Where $\delta(a)$ is the logistic sigmoid purpose $\frac{1}{(1 + \exp(a))}$, v_x, h_y is the unbiased model.

Phase 2: It is to inform the hidden and visible pieces are similar in the providing visible and hidden pieces. For performing the stochastic steepest ascent in the log probability of the training data items, this will show the way to ample straightforward erudition rule as,

$$\Delta W_{xy} \theta(v_x h_y)_{data} (v_x h_y)_{reconstruction} \quad (3)$$

When the RBM is skilled, a different RBM can be “stacked” on upper of it to method a multi-layer model. Each period a different RBM is stacked, the input visible unit is a well-informed vector, and values for the pieces in the already skilled RBM units are apportioned by means of the current weights and biases. The concluding unit of the already-skilled layers is affianced as input to the new RBM. The talented deep network weights are affianced in preparing a fine tuning phase.

Fine tuning phase: The fine tuning phase is just the normal back propagation algorithm. To categorize the blow deepness and place, an output unit is planned in the top of the Deep Neural Network. Also, here it has N number of input value (depending upon the limits), and more than three hidden layers are utilized in our DNN. Also, the training dataset skilled until the enhanced weight is grabbed, or extreme accuracy is reached with the assistance of equation (1). Finally, on the base of the optimum weight (w), the weights are enhanced and reduced the Mean Square Error (MSE) by using WOA algorithm. The method of Whale optimization algorithm is clarified below.

Whale Optimization algorithm: The Whale Optimization Algorithm (WOA) is a newly introduced optimization method that is useful in solving optimization problems. This is excerpted because of its resemblance to humpback whales. Three operators are consisted in the detailed algorithm to induce the prey searching, prey encircling and bubble-net obtaining prey habit of humpback whales. Here Whale Optimization Algorithm (WOA) is used for weight optimization process. The overall procedure of WOA algorithm is mentioned below by steps,

Phase 1: Initialization

The algorithm recognized by randomly making weight value that interconnects to the outcome in the examine space. Here the weight represents the random $w_i (i=1, 2, 3, \dots, n)$ where n represents the number of weight value. And initialize the factor vectors of value such as a, A , and C .

Phase 2: Fitness Calculation

Evaluate the fitness on basis of the equation (4) to become the finest weight value the fitness point of the solution is calculated. It's exposed in under,

$$Fit_i = MinMSE \quad (4)$$

Phase 3: Update the situation of current weight point

Encircling prey: The situation of prey is documented by humpback whale and it is encircled the prey. Towards the finest finding operator, the other finding operators will accordingly effort to inform their locations when the finest finding agent is considered. The update methods, is projected by the underneath equations:

$$\vec{U} = \left| \vec{C} \cdot w^{best}(t) - w(t) \right| \quad (5)$$

$$w(t+1) = w^{best}(t) - \vec{A} \cdot \vec{U} \quad (6)$$

Where t designates a current repetition, \vec{A} and \vec{C} designates a Coefficient vector, w^{best} designates a position vector for best solution, w signifies a current location Vector, $| \cdot |$ represents an absolute point.

The vectors \vec{A} and \vec{C} are calculated as follows:

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \quad (7)$$

$$\vec{C} = 2 \cdot \vec{r} \quad (8)$$

Where, a is the sequence of repetitions linearly abridged from 2 to 0 (in both exploration and exploitation phases), $r \in (0,1)$.

Bubble-net attacking method: To model the bubble net performance of hump back whales scientifically two methods are enhanced as trails:

Shrinking encircling method: The point of a in the equation (7) is reduced to attain this performance. Note that \vec{a} is used to reduction the difference range of \vec{A} . In other words, in the interval $[-a, a]$ \vec{A} is an accidental point where a is reduced from 2 to 0. The new location of a finding agent can be distinct wherever for $\vec{A} \in [-1, 1]$.

Spiral updating model: A spiral calculation between the position of whale and prey is shaped to imitator the helix shaped drive of hump back whales is as given:

$$w(t+1) = D \cdot e^{bk} \cdot \cos(2\pi k) + w^{best}(t) \quad (9)$$

Where $D = \left| w^{best}(t) - w(t) \right|$ and designates the distance of the d^{th} whale and prey which is the finest answer gotten so far, $k \in [-1, 1]$, b means the shape of the logarithmic spiral and it is a continuous point. During optimization the location of whales is efficient by pretentious a probability of 50 percent by selecting any of the spiral model or shrinking encircling method to model this concurrent performance.



The scientific model is given by equation (10).

$$w(t+1) = \begin{cases} w^{best}(t) - \vec{A} \cdot \vec{U} & \text{if } R < 0.5 \\ D \cdot e^{bk} \cdot \cos(2\pi k) + w^{best}(t) & \text{if } R \geq 0.5 \end{cases} \quad (10)$$

Where, $R \in [0, 1]$. The hump back whales finding for prey arbitrarily to form bubble net.

Search for prey: To find for prey the difference of the \vec{A} vector can be castoff by same method. In fact, rendering to the location of each extra hump back whales find randomly. So, to force find agent to transfer so far from a situation whale we use \vec{A} with the accidental points better than 1 or less than -1. The location of the find agent is efficient in manipulation phase. To perform a global search, this method and $|\vec{A}| > 1$

highlight examination allows the WOA algorithm. The precise model is given below:

$$\vec{U} = \left| \vec{C} \cdot w_{rand} - w \right| \quad (11)$$

$$w(t+1) = w_{rand} - \vec{A} \cdot \vec{U} \quad (12)$$

Where, w_{rand} is a current population chance location vector.

Find agents update their locations in every repetition with admiration to either the finest solution got so far or an arbitrarily chosen find agent. In order to deliver examination and manipulation the limit a is reduced from 2 to 0, correspondingly. For updating, an arbitrary find agent is selected when $|\vec{A}| > 1$, while the finest result is selected

when $|\vec{A}| < 1$ for the location of the search agents. Dependent on the point of R, WOA is able to change between either a circular or spiral drive.

Phase 4: Termination criteria: The WOA algorithm is terminated when finest weight point is obtained for the satisfaction of a closure principle.

III. EXPERIMENTAL RESULTS

The goal of this research work is to predict the heart disease using text data (Categorical data). Several methods were applied to healthcare data sets and for the forecast of upcoming health care use such as forecasting separate expenses and illness dangers for patients. The detailed steps involved clinical decision support systems are in two main process 1) Data preprocessing and 2) Prediction. Initially, the duplicate record, missing data, noisy in the consistent data will be removed from the database in preprocessing. For prediction, NER Using Optimal Deep Learning Model is proposed here. Named entity recognition (NER) promises to improve information extraction and retrieval. Here the prediction of heart disease is done by the optimal deep neural network. In our proposed method, the traditional deep neural network is modified by means of optimization technique. The whale optimization is utilized to optimize the parameter of deep neural network. The overall flow diagram is plotted in figure 1.

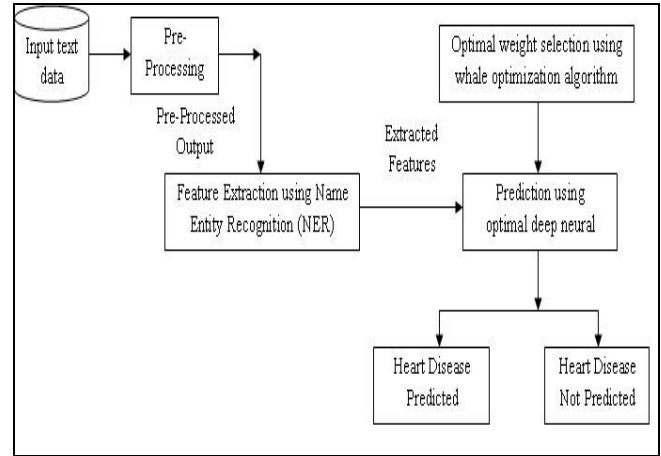


Fig. 1. Flow diagram of ODNN method

1.1 Data Set

This research work has more than 5000 records and which took the records of 400 patients and analyzed. The records are taken from Ashwin Clinic, Anna Nagar, Chennai which is famous for the heart disease.

Table-I. Sample Dataset

Date	Patient Number	Patient Name	Type	Description
'Dec 5, 2016'	'P1'	'Prabhakar.D'	'investigations'	'BP 120/70 mm of Hg PR 80/min. Weight 64 Kg'
'Dec 5, 2018'	'P5'	'Rajesh.R'	'complaints'	'Khan - Ghouse, Valliammal - CAG - PCI under scheme'
'Dec 6, 2014'	'P6'	'Pushpa.N'	'invesigations'	'BP 130/80 mm of Hg PR 93 /min. Weight 57.7 Kg'
'Apr 12, 2018'	'P178'	'VenuGopal.T'	'complaints'	'Dental extraction fitness'
'Nov 19, 2015'	'P180'	'Sakunthala C'	'observations'	'Presurgical evaluation - cataract surgery'
'Dec 9, 2016'	'P256'	'Jaganathan'	'investigations'	'Lipids 182/56/94/156 mg%.'

The dataset which is in .CSV format and it is transformed into the desired format which is used for this research work. In this, the patients which are affected heavily and mildly are taken. Table I shows the sample dataset. The description differs from patient to patient.

A. Preprocessing

In preprocessing, the duplicate record, missing data, noisy in the consistent data will be removed from the database. The pre-processing usually includes changing .csv documents into text document, eliminating stop word, execution word stemming. Stop words are very often used mutual words like 'and', 'are', 'this' etc. They are not valuable in forecast of papers. So, they must be detached. Word stemming eliminates suffixes and generates the stemmed words example becomes recovery.



Then the resultant output is fed to prediction process. This pre-processing method is very useful for the proposed method to get the information with less error rate and gives the higher accuracy prediction.

B. The Proposed Method

For prediction, Named Entity Recognition (NER) Using Optimal Deep Learning Model is proposed here. In bio-medical area, the similar idea may have numerous dissimilar names (synonyms). For example, “heart attack” and “myocardial infarction” points to the similar idea. Using acronyms and abbreviations is very mutual in bio-medical literature which brands it complex to classify the ideas these terms fast. In order to overcome those drawbacks, the proposed method uses NER Using Optimal Deep Learning Models. NER potentials to recover info removal and recovery and the prediction of heart disease are done by the optimal deep neural network. The full procedure of the future technique is described in beneath. Then back propagation algorithm starts with the optimal weights. Primarily, the selected features or text are providing to the Deep Neural Network, though the weight is subjectively attuned.

Finally, on the basis of the optimal weight value, the selected features or text are predicted in testing stage by testing dataset. The performance of the ODNN method is evaluated and the effectiveness of the method is equaled with the other two existing algorithms in results and discussions. Analysis of proposed performance: The presentation examination of the proposed method is exposed in the beneath segment. At this point, the table I illustrates the performance analysis of suggested method by varying the iteration. In the suggested method, for training 90% of input data is used and for testing the remaining 10% of input data is used. The performance analysis of suggested method by varying the iteration is exposed in beneath.

Table-II. Performance analysis by varying no of iteration

Iteration	Precision	Recall	F-measure
10	79.63	70.37	74.71417
20	80.12	71.19	75.39149
30	81.27	72.07	76.39401
40	82.39	73.62	77.7585

From the table II, the proposed method analyzes the performance by changing the number of repetition. Here the ODNN method attains the precision value is 79.63%, recall value is 70.37% and the F-measure value is 74.71% for iteration 10. For iteration 20, the proposed method achieves the precision, recall and f-measure value is 80.12%, 71.19% and 75.39%. The proposed precision value is 81.27%, recall value is 72.07% and the F-measure value is 76.39% for iteration 30. For iteration 40, the proposed method achieves the precision, recall and f-measure value is 82.39%, 73.62% and 77.75%. The table III shows the execution time and memory utilization of proposed method. It is tabulated in below.

The proposed method analyzes the execution time and memory utilization by changing the number of repetition. Here the proposed method reaches the execution time is 12365ms and the memory utilization of ODNN method is 115441 bits for iteration 10. For iteration 20, the ODNN

method reaches the execution time is 21544ms and the memory utilization of ODNN method is 127844 bits. The proposed execution time is 26987ms and the memory utilization of ODNN method is 137457 bits for iteration 30. For iteration 40, the ODNN method reaches the execution time is 31587ms and the memory utilization of ODNN method is 150018 bits. The effectiveness of the proposed method is analyzed and the results are compared with the existing method in the following section.

Table-III. Time and memory utilization by varying no of iteration

Iteration	Execution time (ms)	Memory utilization (bits)
10	12365	115441
20	21544	127844
30	26987	137457
40	31587	150018

C. Results and discussion

This section provides the thorough view of the outcome that is obtained by optimal named entity recognition of heart disease which is performed in the occupied platform of JAVA program. The proposed heart disease prediction is done by optimal deep neural network. The traditional deep neural network is changed by whale optimization algorithm. The experimental result and the performance of the ODNN method are given below in detail.

Effectiveness proposed method: In order to confirm the efficiency of the recommended method, here the ODNN method is compared with the stated method. For comparison, the ODNN method considers the existing method is deep neural network without optimization and the traditional neural network. At this time, the ODNN method calculates the prediction accuracy, precision, recall and f-measure value for both proposed and existing method. The figure 2 specifies the contrast of the precision, recall and f-measure value of projected method and stated method.

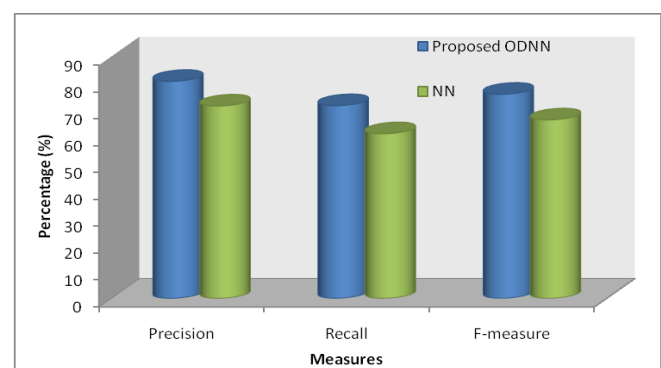


Fig. 2. Result of ODNN and NN Method

When analyzing the figure 2, the proposed method attains the precision value is 80.85% but the exiting traditional neural network (NN) achieves the precision value is 71.68% which is minimum value when compared to the existing method. The recall value for the proposed method is 71.81% but the NN attains the 61.36% value.

When compared to the proposed method the existing method attains the minimum recall value. The f-measure value for the proposed method is 76.06% but the NN achieves the f-measure value is 66.52%. When compared to the existing method, the proposed attains the maximum f-measure value. From the results, it is clear that the proposed method outperforms better when compared to the existing methods.

Table-IV. Performance of NN and ODNN

Algorithms	Precision	Recall	F-measure
ODNN	80.85	71.81	76.06
NN	71.68	61.36	66.52

The table IV shows the performance analysis and table V shows the time and memory utilization of NN and proposed ODNN methods.

Table-V. Performance of DNN and ODNN

Algorithm	Accuracy
ODNN	79.36
NN	69.48

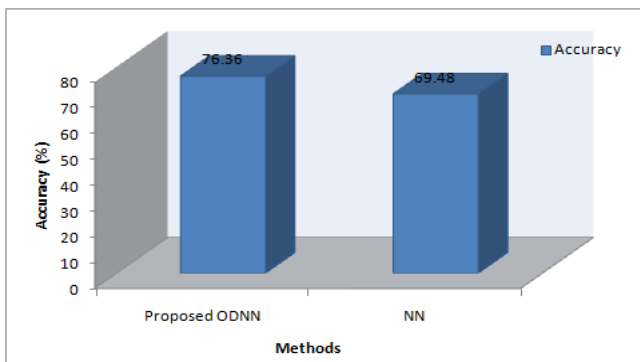


Fig. 3. Accuracy comparison of proposed method

When analyzing the figure 3, it is clear that the proposed method attains the maximum prediction accuracy when compared to the existing method. Here the proposed method attains the maximum prediction accuracy is 79.36% but the existing method NN achieves the accuracy value is 69.48%. When compared to the existing method the proposed method attains the maximum prediction accuracy value because the proposed method selects the optimal weight value using whale optimization algorithm. If the error value is reduced in neural network structure it will produce the more prediction accuracy. From the experimental results, the proposed method gives more prediction accuracy when compared to the existing methods. The proposed method recognizes the entities of other types on other genres of text without needing manual annotation.

IV. CONCLUSION

Optimal named entity recognition of heart disease is proposed in this research work. The heart disease prediction is done by optimal deep neural network with whale optimization algorithms. There are two types of output obtained which gives the patients affected by heart disease or not and it tells the effective result that twenty five percentage of the patients

are affected from heart disease from the given dataset and it is very much confirmed with the physician in practice. The assessment metrics of proposed method is prediction accuracy, precision, recall and f-measure. The proposed method attains the prediction accuracy value which is maximum value when compared to the existing method. The proposed method also gets the maximum precision, recall and f-measure value. From the experimental results, it is clear that the proposed method outperforms better when compared to the existing method. It is confirmed that from the calculated weights, the heart disease predicted by the ODNN method is very high and by NN method is less compared with the ODNN. The proposed method recognizes the entities of other types on other genres of text without needing manual annotation. In future, the researcher has sufficient opportunity to incorporate various prediction algorithms in order to improve the prediction accuracy.

REFERENCES

- Allison J. Lazard, Emily Scheinfeld, Jay M. Bernhard, Gary B. Wilcox, Melissa Suran, "Detecting themes of public concern: a text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat," *American journal of infection Control*, Vol. 43, No. 10, pp. 1109-1111, 2015.
- Bei Zhong, L. C. Kong, Jin Liu, "Entity Attribute Extraction from Unstructured Text with Deep Belief Network", *International journal of database theory and application*, Vol.9, No.5, pp. 187-196, 2016.
- D. Agnihotri, K. Verma and P. Tripathi, "Pattern and Cluster Mining on Text Data," in *process of Fourth International Conference on Communication Systems and Network Technologies*, Bhopal, pp. 428-432, 2014.
- Eskici, HaticeBurcu, and Necmettin Alpay Koçak, "A text mining application on monthly price developments reports," *Central Bank Review*, Vol. 18, No. 2, pp. 51-60, 2018.
- Fei Zhu, Preecha Patumcharoenpol, Cheng Zhang, Yang Yang, Jonathan Chan, Asawin Meechai, Wanwipa Vongsangnak, BairongShen, "Biomedical text mining and its applications in cancer Research", Vol. 46, No.2, pp.200-211, 2013.
- Filipe R. Lucini, Flavio S. Fogliatto, Giovanni J.C. da Silveira, Jeruza Neyeloff, Michel J. Anzanello, Ricardo de S. Kuchenbecker, Beatriz D. Schaanc, "Text mining approach to predict hospital admissions using early medical records from the emergency department," *Medical Informatics*, Vol. 100, pp.1-8, 2017.
- Fleuren, Wilco WM, and Wynand Alkema, "Application of text mining in the biomedical domain," *Methods*, Vol. 74, No. 97, pp. 97-106, 2015.
- G. Shi and Y. Kong, "Advances in Theories and Applications of Text Mining," in *process of first International Conference on Information Science and Engineering*, Nanjing china, pp. 4167-4170, 2009.
- Jain, Vinay Kumar, Shishir Kumar, and Steven Lawrence Fernandes, "Extraction of emotions from multilingual text using intelligent text processing and computational linguistics," *Computational Science*, Vol. 21, pp.:316-326, 2017.
- Jitendra Jonnagaddala, Siaw-Teng Liaw, Pradeep Ray, Manish Kumar, Nai-Wen Chang, Hong-JieDai, "Coronary artery disease risk assessment from unstructured electronic health records using text mining," *Biomedicalinformatics*, Vol.58 .pp. 203-210, 2015.
- Kumar, B. Shravan, and Vadlamani Ravi, "A survey of the applications of text mining in financial domain," *Knowledge-Based Systems*, Vol. 114, pp.28-147, 2016.
- Lim, Sunghoon, Conrad S. Tucker, and Soundar Kumara, "An unsupervised machine learning model for discovering latent infectious diseases using social media data," *Journal of biomedical informatics*, Vol. 66, pp. 82-94, 2017.
- M. A. Jabbar, Priti Chandra, B. L Deekshatulu, "Knowledge Discovery from Mining Association Rules for Heart Disease Prediction", *Journal of Theoretical and Applied Information Technology*, Vol.41, No.2, pp., July 2012

14. Medhat, Walaa, Ahmed Hassan, and HodaKorashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, Vol. 5, No.4, pp.1093-1113, 2014.
15. Melissa Ailem, François Role, Mohamed Nadif, Florence Demenais, "Unsupervised text mining for assessing and augmenting GWAS results," *Biomedical Informatics*, Vol. 60, pp. 252-259, 2016.
16. P. K. Anooj. "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," *Elsevier Computer and Information Sciences*, Vol. 24, no. 1, pp. 27-40, January 2012.
17. Piedra, David, Antoni Ferrer, Joaquim Gea, "Text mining and medicine: usefulness in respiratory diseases", *Archivos de Bronconeumología*, Vol.50, No.3, pp.113-119, 2014.
18. Rave Harpaz, Alison Callahan, Suzanne Tamang, Yen Low, David Odgers, Sam Finlayson, Kenneth Jung, Paea Le Pendu, and Nigam H. Shah, "Text mining for adverse drug events: the promise, challenges, and state of the art," *Drug safety*, Vol. 37, No.10, pp.777-790, 2014.
19. Sahu, Sunil Kumar, Ashish Anand, Krishnadev Oruganty, Mahanandeeshwar Gattu. "Relation extraction from clinical texts using domain invariant convolutional neural network." *arXiv preprint arXiv:1606.09370* (2016).
20. Sarker, Abeer, Graciela Gonzalez, "Portable automatic text classification for adverse drug reaction detection via multi-corpus training", *Biomedical Informatics*, Vol. 53, pp.196-207, 2015.
21. Sheng, Xiaobao, Xin Wu, Yimin Luo, "A novel text mining algorithm based on deep neural network," in *process of Inventive Computation Technologies (ICICT) International Conference*, Vol. 2, 2016.
22. Simpson, Matthew S, Dina Demner- Fushman, "Biomedical text mining: a survey of recent progress," *Mining Text Data*, pp.465-517, 2012.
23. SunePletscher-Frankild, Albert Palleja, KalliopiTsafou, Janos X. Binder, Lars Juhl Jensen, "DISEASES: Text mining and data integration of disease-gene associations," *Methods*, Vol. 74, pp.83-89, 2015.
24. Y. Xianguang, "Text Mining Software and Their Applications," in *process of Fourth International Conference on Instrumentation and Measurement, Communication and Control*, China, pp. 902-905, 2014.
25. Zhou, Xuezhong, YonghongPeng, and Baoyan Liu, "Text mining for traditional Chinese medical knowledge discovery: a survey," *Biomedical Informatics*, Vol. 43, No.4, pp.650-660, 2010.-vidmar

arranged a Certificate program in "Android Applications". Participated in various International/National level workshops and seminars. She acted one of the organizers in International Conference on Computing and Intelligence Systems (ICCIS 2015).

AUTHORS PROFILE



Dr. T. Velmurugan is working as an Associate Professor in the PG and Research Department of Computer Science and Applications, D. G. Vaishnav College, Chennai-600106, India. Also, he is the Head of the Department of Computer Science and Applications (UG). He holds a Ph.D. degree in Computer Science from the University of Madras. He

is elected as a Senate Member, University of Madras. He has published more than 90 articles indexed in SCOPUS and SCI such as Applied Soft Computing, Journal of Computer Science and etc. He was an invited speaker of the 10th Int'l Conference on Computational Intelligence and Software Engineering (CiSE 2018) held from January 5-7, 2018 in Bangkok, Thailand. He has guided more than 300 M.Phil. Research Scholars in the field of Computer Science. He guided 7 Ph.D. scholars and currently guiding 10 Ph.D. scholars in the same field. He served as a nominated Senate Member in Middle East University, Dubai, UAE for a period of three years. He is a member in Board of studies for many autonomous institutions and Universities like Periyar University, Salem, India. He hosted a lot of programmes in Doordharsan television about recent topics in Information Technology field. He arranged and acted as an Organizing Secretary of International Conference on Computing and Intelligence Systems (ICCIS 2015). In addition, he was a resource person for various national workshops entitled "Scientific Research Article Writing and Journal Publications". He is an editorial board member of many International Journals. He is also a reviewer in many peer reviewed journals like Elsevier, Springer, IOS Press Journals etc. Further, he is a visiting faculty for M.Phil. Courses for various universities throughout India.



U. Latha is working as an Assistant Professor in the Department of Computer Applications, D. G. Vaishnav College, Chennai-600106, India. Also, she is research scholar under the guidance of Dr. T. Velmurugan in University of Madras, Chennai-600005. Participated and presented research papers in International/National Conferences and published two journals. Also, she