Journal of Computer and Communications, 2021, 9, 66-79
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Classifying Heart Disease in Medical Data Using Deep Learning Methods

T. Velmurugan, U. Latha

PG and Research Department of Computer Science, D. G. Vaishnav College, Chennai, India
Email: velmurugan_dgvc@yahoo.co.in, dgvclatha@gmail.com

How to cite this paper: Velmurugan, T. and Latha, U. (2021) Classifying Heart Disease in Medical Data Using Deep Learning Methods. *Journal of Computer and Communications*, **9**, 66-79.
https://doi.org/10.4236/jcc.2021.91007

Received: December 30, 2020
Accepted: January 26, 2021
Published: January 29, 2021

Abstract

Recent days, heart ailments assume a fundamental role in the world. The physician gives different name for heart disease, for example, cardiovascular failure, heart failure and so on. Among the automated techniques to discover the coronary illness, this research work uses Named Entity Recognition (NER) algorithm to discover the equivalent words for the coronary illness content to mine the significance in clinical reports and different applications. The Heart sickness text information given by the physician is taken for the preprocessing and changes the text information to the ideal meaning, at that point the resultant text data taken as input for the prediction of heart disease. This experimental work utilizes the NER to discover the equivalent words of the coronary illness text data and currently uses the two strategies namely Optimal Deep Learning and Whale Optimization which are consolidated and proposed another strategy Optimal Deep Neural Network (ODNN) for predicting the illness. For the prediction, weights and ranges of the patient affected information by means of chosen attributes are picked for the experiment. The outcome is then characterized with the Deep Neural Network and Artificial Neural Network to discover the accuracy of the algorithms. The performance of the ODNN is assessed by means for classification methods, for example, precision, recall and f-measure values.

Keywords

Named Entity Recognition Algorithm, Neural Network Methods, Whale Optimization Algorithm, F-Measure, Recall, Precision

1. Introduction

The biomedical content mining is changed consistently. The unique study diary demonstrates the universally useful content calculation and information mining apparatuses are not all around characterized for the biomedical area since it is profoundly particular. The data enlightens from [1] [2]. It's concerning ceaselessly to investigate, question, break down and deal with the underutilized data. Part of bio clinical research advises the rich set regarding information for biomedical research. Content mining gives the information from a pile of content and applied in biomedical research. It has numerous computational systems, for example, AI, regular language handling to locate the unstructured biomedical content. To characterize the helpful content digging errands for the particular objectives of scientists, biomedical content mining and clinicians are better situated [3].

The difficulty in gaining new knowledge and information due to extraordinary growth in experiment and literature of biomedical science has led to the loss of hypotheses in text data mining. To overcome this, we should recognize the biomedical named entities that expose the inter-related relationships of these biomedical entities [4]. To change the unstructured information into organized information, a wide range of the content data paying little heed to the configuration can be changed to numbers listed for every one of reports utilized in content mining [5]. These days content mining assumes a major job in the controls like AI, information mining, data recovery and measurements and so forth, which is utilized to arrange, bunch, condense, relationship, conveyance for an enormous number of dataset [6].

Content mining additionally called as content information mining or information found structure information and it is a part of information mining. It is characterized as a computational procedure of taking out the significant data from the large measure of unstructured content information [7]. Content mining is likewise utilized in assortment of utilizations like digital security, for example, extortion recognition and interruption location and so forth and to fathom the client related information, for example, client obtaining, advertise container investigation and so on [8]. Content mining has the exceptionally solid strength to identify the commotion and sporadic structure in content information [9]. NER is additionally used to discover the jargon like CoPub in liver pathology terms [10].

To perceive sickness and human qualities content and concentrate the qualities, there is an open-source content mining programming and it is adaptable innovation to apply in different assignments in clinical and science [11]. Computational treatment of choices, conclusions and subjectivity of content utilized in Sentiment Analysis (SA) and it is progressing field in content mining, additionally required for some other normal dialects [12]. These days, it is one of the significant same sorts of significant issue that is design finding or bunching in content mining [13]. It is additionally utilized for getting prescient results, keen ready frameworks and continuing the clinical in the dynamic procedure [14]. NLP is the unmistakable informational indexes to extricate the highlights which speak to the regular data. It is joined with AI methods to improve the characterization of online networking content containing clinical information [15].

An unaided AI model for finding idle irresistible maladies utilizing online life information is examined by Sunghoon Lim *et al.* [16] and it has introduced a technique called solo AI model to discover dormant irresistible ailments without given data, for example, the name of that illnesses and their side effects. In that diary, a national general wellbeing establishment and correspondence with the overall population gives an inactive irresistible infection was characterized as a transferable sickness that has not yet been formalized.

An exploration study was finished by Haghanikhameneh *et al.* [17] in their work. The general idea behind classification in Data mining is to predict the target class from analyzing the training dataset. Certainly, it is the most significant task that can be applied in different field of human life. Zamani *et al.* [18] has developed; a meta-heuristic algorithm is proposed named FSWOA for feature selection. This algorithm is based on the hunting methods of Humpback Whales consisting of three main steps: encircling prey, spiral bubble-net attacking and search for prey. The performance of this algorithm is evaluated conducted by four standard medical datasets: Pima Indians Diabetes, Original Wisconsin Breast Cancer, Statlog and Hepatitis.

A fully automated deep learning framework for EAT and thoracic adipose tissue (TAT) quantification from noncontrast coronary artery calcium CT scans [19]. A first multi-task convolutional neural network (ConvNet) is used to determine heart limits and perform segmentation of heart and adipose tissues. A second ConvNet, combined with a statistical shape model (SSM), allows for pericardium detection. EAT and TAT segmentations are then obtained from outputs of both ConvNets.

Crisis office (ED) clog is a major issue for emergency clinics. An exploration work by Filipe R. Lucini *et al.* [20] advises to anticipate future hospitalizations and releases, a book mining techniques to process information from early crisis understanding records utilizing the SOAP system. They are attempted different methodologies for pre-handling of content information records and to anticipate hospitalization. Twofold portrayal, term recurrence, and term recurrence opposite record recurrence were accustomed to getting a lot of words. A finding of Ebola on US soil activated far reaching alarm. Accordingly, the Centers for Disease Control and Prevention held a live Twitter talk to address open concerns. Another work did by Allison J. Lazard *et al.* [21] in their research work, a literary experiment procedure to uncover bits of knowledge from those tweets that can illuminate correspondence technique. Client produced tweets were gathered, arranged, and investigated is significant subject of their article.

An experimental work did by Jitendra Jonnagaddala *et al.* [22] tells from unstructured electronic wellbeing information, a strategy called clinical content mining to extract Framingham hazard factors are utilized. That is likewise utilized for the ascertaining the diabetic patients for 10-year coronary supply route illness hazard scores. With the assistance old enough, sexual orientation, absolute cholesterol, HDL-C, pulse, diabetes history and smoking history, they will discover the hazard factor. Melissa Ailem*et al.* [23] has introduced a nonexclusive structure which we used to encapsulate the relations between 10 qualities report related with asthma by a past GWAS. The goal is to use unaided content information mining methods utilizing content based cosine closeness examinations and bunching applied to competitor and arbitrary quality vectors, so as to increase the GWAS results. Xia *et al.* [24] have presented, advanced natural language processing and deep learning for high-performance ADE extraction. The framework consists of training the word embeddings using a large medical domain corpus to capture precise semantic and syntactic word relationships, and a deep learning based named entity recognition method for drug and ADE entity identification and prediction.

Roygaga *et al.* [25] have deals with the analysis of parameters of Error-Back Propagation algorithm that would provide the best accuracy for diagnosing heart disease in a patient. The network with 13 hidden neurons provides the best Accuracy, Specificity and Positive Prediction Value. Another exploration work completed by M.A. Jabbar *et al.* [26] in their experimental work. A strategy is there to find affiliation runs in medical text information to discover coronary illness for Andhra Pradesh. That approach was required to help specialists to settle on exact choice. A hybrid methodology based on the fuzzy analytic hierarchy process and fuzzy inference system to design a clinical decision support system (CDSS) [27] with the aim of evaluating the likelihood of developing heart diseases. CDSS assesses patients' conditions at a cheaper cost because additional expensive diagnosis and clinical tests will only be prescribed once the CDSS reports a high likelihood of developing heart diseases.

Latha *et al.* [28] in their research work, the proposed Optimal Neural network algorithm is more efficient than traditional neural network by means of accuracy, sensitivity and specificity. In another research work, Velmurugan *et al.* [29] tells that the newly invented ODNN method is more proficient than Deep Neural Network algorithm by means of precision, recall and f-measure.

2. Materials and Methods

The problem definition of this work is discussed in this section. The significant issue in clinical content information mining undertakings is association the different idea of unstructured story message in the clinical record. The precise location of ailment status from clinical content requires a comprehension of example and key expressions in a subject's clinical history, which can differ widely. The availability of mammoth measure of clinical information prompts the requirement for powerful information examination apparatuses to take out valuable information. The dataset comprises of excess information, missing information and insignificant characteristics is preprocessed by methods for name element acknowledgment and resultant information is put away in a book record named sentiwordnet and the cleaned coronary illness dataset given in an Optimal Deep Neural Network (ODNN) to anticipate which patient is influenced intensely and gently with the assistance of loads that are taken from the use of the ODNN. This examination work has in excess of 5000 records and which took the records of 400 patients and broke down. The records are taken from Ashwin Clinic, Anna Nagar, Chennai which is well known for the coronary illness.

2.1. Data Set

This work has in excess of 5000 records and which took the records of 400 patients and broke down. The records are taken from Ashwin Clinic, Anna Nagar, Chennai which is acclaimed for the coronary illness. The dataset which is in.CSV configuration and it is changed into the ideal arrangement which is utilized for this examination work. In this, those are influenced intensely and gently are taken for the prediction. The portrayal contrasts from patient to quiet. **Table 1** shows the sample dataset given by the physician.

2.2. Named Entity Recognition

Named Entity Recognition (NER), or substance extraction is a NLP strategy which finds and arranges the named elements present in the content. Named Entity Recognition groups the named substances into pre-characterized classifications, for example, the names of people, associations, areas, amounts, financial qualities, specific terms, item wording and articulations of times.

2.3. Optimal Deep Learning Models

The selected highlight or content from the NER is given to the contribution for expectation arrange. In this, the utilized method is Optimal Deep Neural Network (ODNN). In our proposed strategy, the customary profound neural system

Table 1. 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%.

is changed by methods for advancement procedure. The whale improvement is used to advance the parameter of profound neural system to learn elevated level component portrayals, catch long haul conditions, and worldwide highlights to help recognize clinical substances. A counterfeit neural system model with the different layers of the concealed units and yields is named DNNs. In addition, it comprises of both pre-preparing (utilizing generative profound conviction system or DBN) and calibrating stages in its parameter learning. The principle point of this paper is to prepare the highlights in the specific informational index, for example to locate the correct weight that can be utilized to effectively anticipate the content. Using this weight and score, the prediction is made with the help of accuracy, sensitivity and specificity.

3. Experimental Results

The objective of this examination work is to anticipate the coronary illness utilizing content information (Categorical information). A few procedures were applied to human services informational collections and for the expectation of future social insurance usage, for example, anticipating singular uses and ailment dangers for patients. The definite advances included clinical choice emotionally supportive networks are in two fundamental procedure 1) Data preprocessing and 2) Prediction. At first, the copy record, missing information, loud in the reliable information will be expelled from the database in preprocessing. For forecast, Named Entity Recognition (NER) Using Optimal Deep Learning Model is proposed here. Named substance acknowledgment (NER) vows to improve data extraction and recovery. Here the expectation of coronary illness is finished by the ideal profound neural system. In our proposed strategy, the customary profound neural system is changed by methods for streamlining procedure. The whale improvement is used to enhance the parameter of profound neural system. The architecture of ODNN method is shown in **Figure 1**.

3.1. Preprocessing

In preprocessing, the duplicate record, missing information, loud in the predictable information will be expelled from the database. The preprocessing

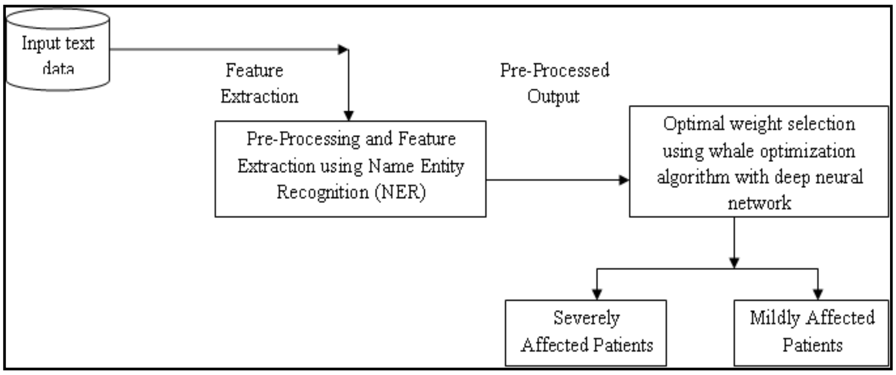

Figure 1. Architecture of ODNN method.

ordinarily incorporates changing over xml archives into content report, evacuating stop word, performing word stemming. Stop words are every now and again utilized normal words like "and" "are" "this" and so forth. They are not helpful in expectation of archives. So they should be expelled. Word stemming evacuates additions and creates the stemmed words model recovery becomes retries. At that point the resultant yield is taken care of to forecast process.

3.2. The Optimal Deep Neural Network

For expectation, Named Entity Recognition (NER) Using Optimal Deep Learning Model is proposed here. In biomedical area, a similar idea may have various names (equivalent words). For instance, "cardiovascular failure" and "myocardial localized necrosis" point to a similar idea. Utilizing abbreviations and shortened forms is normal in biomedical writing which makes it confounded to recognize the ideas these terms express. So as to conquer those disadvantages, the proposed strategy utilizes Named Entity Recognition (NER) Using Optimal Deep Learning Models. Named element acknowledgment (NER) vows to improve data extraction and recovery and the expectation of coronary illness are finished by the ideal profound neural system. The point by point procedure of the proposed strategy is portrayed in underneath.

At that point back engendering calculation begins with the ideal loads. Essentially, the chose highlights or content are given to the DNN, however the weight is subjectively balanced. At long last, based on the ideal weight esteem, the chose highlights or content are anticipated in testing stage by testing dataset. The exhibition of the proposed technique is assessed and the adequacy of the proposed strategy is contrasted and the current calculation in results and conversations.

The effectiveness of the proposed method is analyzed and the results are compared with the existing method in the following section.

3.3. Results and Discussions

This area gives the detailed perspective on the outcome that is gotten by proposed optimal named entity recognition of coronary illness which is acted in the working foundation of JAVA. The proposed coronary illness expectation is finished by ideal profound neural system. Here the conventional profound neural system is modified by methods for whale enhancement calculation. The test result and the presentation of the proposed strategy are given beneath in detail.

The experiments were carried out on Intel core i3 processor with 2.0 GHz, 2 GB RAM memory which works on windows 7 operating system. The computational time and the memory space may vary depending upon the system requirements; for this hardware specification the coronary illness dataset produces the results that are given below. The resultant dataset are compared with the existing algorithms to validate the efficiency and accuracy and in finding the best algorithm. Validation and comparison is based on time, space, precision, recall and f-measure for the medical dataset.

Figure 2, **Figure 3** and **Figure 4** show the sample result of all the three algorithms because the data are big in numbers shows different patients report. Implementation of various algorithms is performed on the dataset and the results are evaluated and compared to find the efficient algorithm.

After the pre-processing with NER the resultant data is then implemented with ANN algorithm and the sample results are shown in **Figure 2**. **Figure 3** shows the results of DNN classification algorithm.

After the pre-processing with NER the resultant data is then implemented with DNN algorithm and the sample results are shown in **Figure 3** for those who

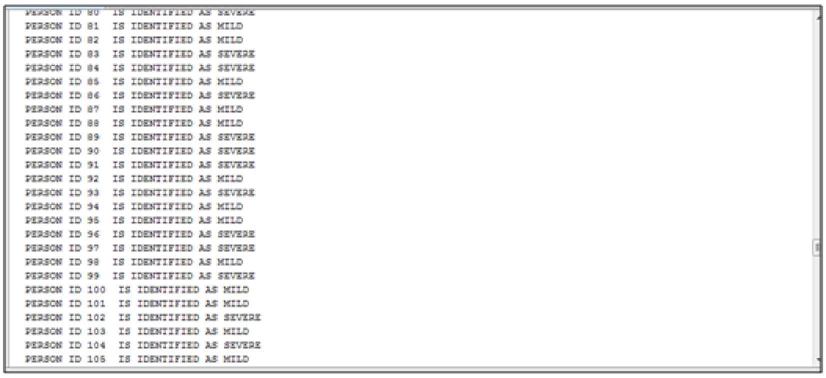
PERSON ID 80 IS IDENTIFIED AS SEVERE
PERSON ID 81 IS IDENTIFIED AS MILD
PERSON ID 82 IS IDENTIFIED AS MILD
PERSON ID 83 IS IDENTIFIED AS SEVERE
PERSON ID 84 IS IDENTIFIED AS SEVERE
PERSON ID 85 IS IDENTIFIED AS MILD
PERSON ID 86 IS IDENTIFIED AS SEVERE
PERSON ID 87 IS IDENTIFIED AS MILD
PERSON ID 88 IS IDENTIFIED AS MILD
PERSON ID 89 IS IDENTIFIED AS SEVERE
PERSON ID 90 IS IDENTIFIED AS SEVERE
PERSON ID 91 IS IDENTIFIED AS SEVERE
PERSON ID 92 IS IDENTIFIED AS MILD
PERSON ID 93 IS IDENTIFIED AS SEVERE
PERSON ID 94 IS IDENTIFIED AS MILD
PERSON ID 95 IS IDENTIFIED AS MILD
PERSON ID 96 IS IDENTIFIED AS SEVERE
PERSON ID 97 IS IDENTIFIED AS SEVERE
PERSON ID 98 IS IDENTIFIED AS MILD
PERSON ID 99 IS IDENTIFIED AS SEVERE
PERSON ID 100 IS IDENTIFIED AS MILD
PERSON ID 101 IS IDENTIFIED AS MILD
PERSON ID 102 IS IDENTIFIED AS SEVERE
PERSON ID 103 IS IDENTIFIED AS MILD
PERSON ID 104 IS IDENTIFIED AS SEVERE
PERSON ID 105 IS IDENTIFIED AS MILD

Figure 2. Results of ANN Algorithm.

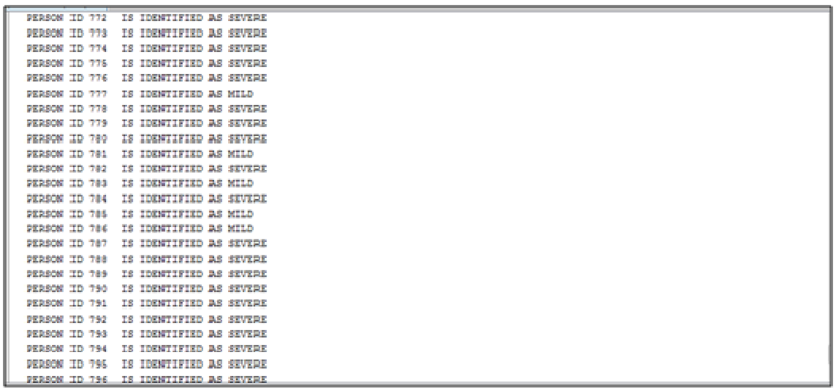
PERSON ID 772 IS IDENTIFIED AS SEVERE
PERSON ID 773 IS IDENTIFIED AS SEVERE
PERSON ID 774 IS IDENTIFIED AS SEVERE
PERSON ID 775 IS IDENTIFIED AS SEVERE
PERSON ID 776 IS IDENTIFIED AS SEVERE
PERSON ID 777 IS IDENTIFIED AS MILD
PERSON ID 778 IS IDENTIFIED AS SEVERE
PERSON ID 779 IS IDENTIFIED AS SEVERE
PERSON ID 780 IS IDENTIFIED AS SEVERE
PERSON ID 781 IS IDENTIFIED AS MILD
PERSON ID 782 IS IDENTIFIED AS SEVERE
PERSON ID 783 IS IDENTIFIED AS MILD
PERSON ID 784 IS IDENTIFIED AS SEVERE
PERSON ID 785 IS IDENTIFIED AS MILD
PERSON ID 786 IS IDENTIFIED AS MILD
PERSON ID 787 IS IDENTIFIED AS SEVERE
PERSON ID 788 IS IDENTIFIED AS SEVERE
PERSON ID 789 IS IDENTIFIED AS SEVERE
PERSON ID 790 IS IDENTIFIED AS SEVERE
PERSON ID 791 IS IDENTIFIED AS SEVERE
PERSON ID 792 IS IDENTIFIED AS SEVERE
PERSON ID 793 IS IDENTIFIED AS SEVERE
PERSON ID 794 IS IDENTIFIED AS SEVERE
PERSON ID 795 IS IDENTIFIED AS SEVERE
PERSON ID 796 IS IDENTIFIED AS SEVERE

Figure 3. Results of DNN Algorithm.

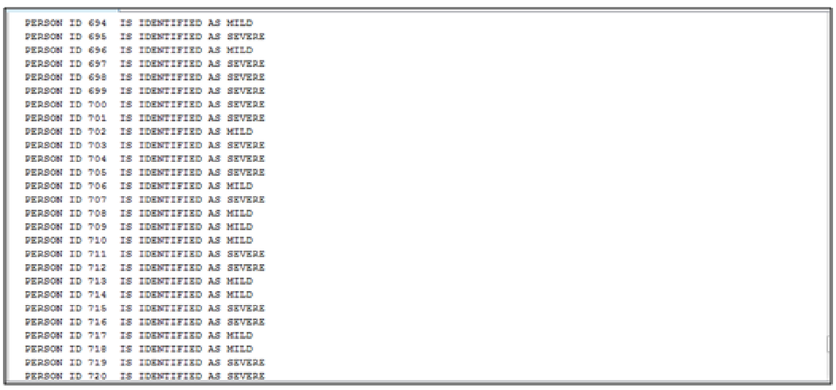
PERSON ID 694 IS IDENTIFIED AS MILD
PERSON ID 695 IS IDENTIFIED AS SEVERE
PERSON ID 696 IS IDENTIFIED AS MILD
PERSON ID 697 IS IDENTIFIED AS SEVERE
PERSON ID 698 IS IDENTIFIED AS SEVERE
PERSON ID 699 IS IDENTIFIED AS SEVERE
PERSON ID 700 IS IDENTIFIED AS SEVERE
PERSON ID 701 IS IDENTIFIED AS SEVERE
PERSON ID 702 IS IDENTIFIED AS MILD
PERSON ID 703 IS IDENTIFIED AS SEVERE
PERSON ID 704 IS IDENTIFIED AS SEVERE
PERSON ID 705 IS IDENTIFIED AS SEVERE
PERSON ID 706 IS IDENTIFIED AS MILD
PERSON ID 707 IS IDENTIFIED AS SEVERE
PERSON ID 708 IS IDENTIFIED AS MILD
PERSON ID 709 IS IDENTIFIED AS MILD
PERSON ID 710 IS IDENTIFIED AS MILD
PERSON ID 711 IS IDENTIFIED AS SEVERE
PERSON ID 712 IS IDENTIFIED AS SEVERE
PERSON ID 713 IS IDENTIFIED AS MILD
PERSON ID 714 IS IDENTIFIED AS MILD
PERSON ID 715 IS IDENTIFIED AS SEVERE
PERSON ID 716 IS IDENTIFIED AS SEVERE
PERSON ID 717 IS IDENTIFIED AS MILD
PERSON ID 718 IS IDENTIFIED AS MILD
PERSON ID 719 IS IDENTIFIED AS SEVERE
PERSON ID 720 IS IDENTIFIED AS SEVERE

Figure 4. Results of ODNN Algorithm.

are affected severely and mildly from heart disease.

After pre-processing the medical text data with Name Entity Recognition, the ODNN algorithm is implemented and results are shown in **Figure 4**. **Table 2** shows the identification of patients by all the three algorithms those who are affected severely and mildly from the real world medical text dataset.

Figure 5 shows the identification of patients those who are affected severely and mildly from heart disease by all the three algorithms to the medical text dataset. The efficiency and accuracy of the ODNN algorithm is been validated by comparing the results with the ANN algorithm and with the DNN classification algorithm. The efficiency of the algorithm is perform by two factors, one is the speed which is been calculated by the time taken to implement the algorithm and another factor is storage space calculated by the algorithm for the resultant data.

The effectiveness of the suggested technique, here the proposed strategies are contrasted and the expressed technique. The underneath **Figure 6** determine the differentiation of the precision, recall and f-measure estimation of the anticipated technique and expressed strategy. **Table 3** shows the performance analysis

Table 2. Identification of Patients by all the Algorithms.

Affected Patients	ODNN	DNN	ANN
Severe	78	70	83
Mild	322	330	317

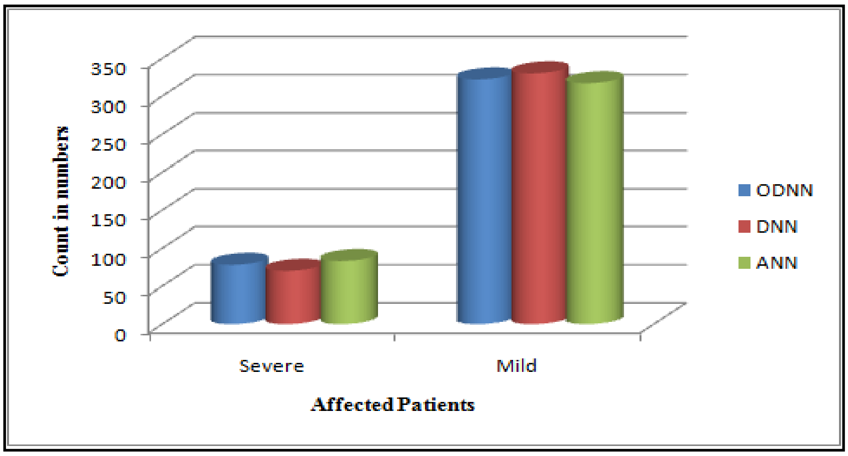

Figure 5. Identification of patients by all the algorithms.

Table 3. Performance Analysis of ODNN, DNN and ANN.

Algorithms	Precision	Recall	F-measure
ODNN	79.63	70.37	74.71
DNN	73.94	65.22	69.56
ANN	70.46	60.92	65.17

When analyzing **Figure 6**, the performance analysis of all the algorithms after iterations and the ODNN algorithm achieves the precision value is 79.63%, recall value is 70.37% and the F-measure value is 74.71%. The DNN algorithm attains the precision, recall and f-measure values are 73.94%, 65.22% and 69.56%. The precision value is 70.46%, recall value is 60.92% and the F-measure value is 65.17% achieves by the ANN Algorithm. From the above results, it is clearly known that the proposed method outperforms better when compared to the existing methods and **Table 4** shows the execution time and memory utilization of all the algorithms.

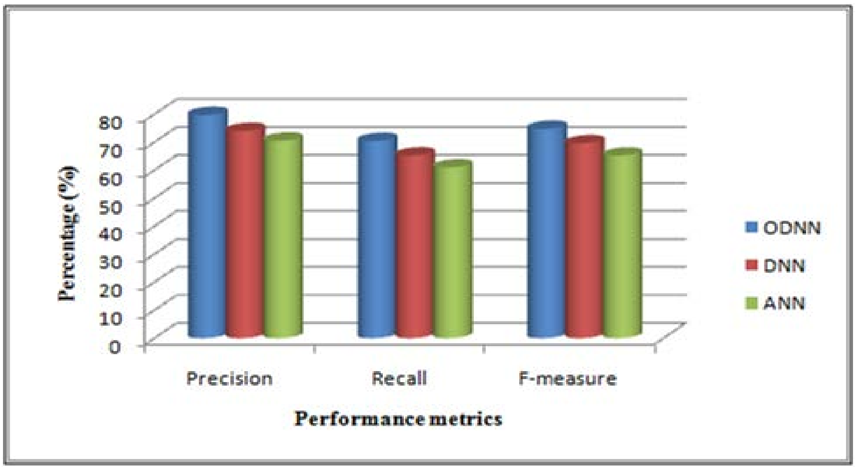

Figure 6. Performance analysis.

Table 4. Average computational time and memory utilization of algorithms.

Algorithms	Execution time (ms)	Memory utilization (bits)
ODNN	2854	134469
DNN	3332	188256
ANN	4101	246289

Figure 7. Results based on run time.

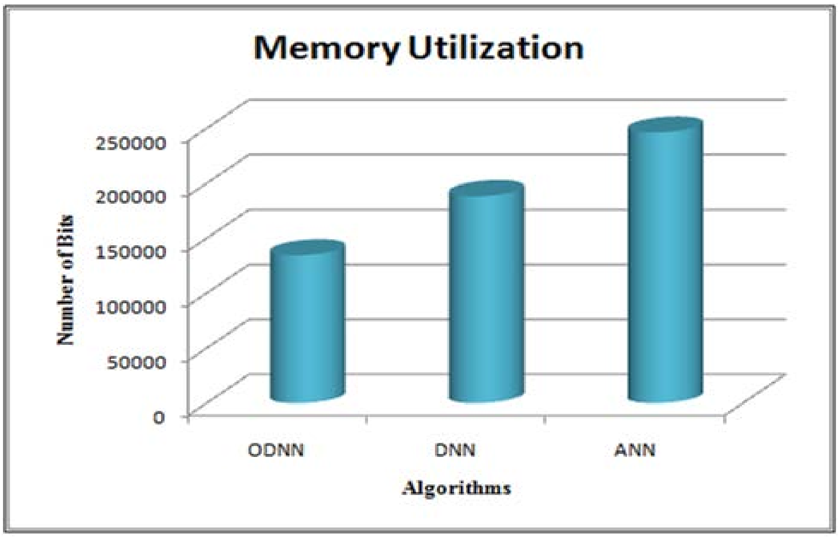

Figure 8. Results based on memory space.

Figure 7 shows the graphical illustration of the execution time taken by the resulting dataset by all the three classification algorithms. **Figure 8** shows the graphical illustration of the memory space occupied by the resulting dataset of the three classification algorithms. **Figure 7** shows that the time to compute the ODNN algorithm is very less when compared to other two algorithms that is DNN and ANN algorithm. **Figure 8** shows that the memory space occupied by ODNN algorithm is also relatively less when compared to ANN and DNN algorithms for the medical text dataset.

4. Conclusion

The computerized techniques utilized right now one of the applications calculations to break down the clinical content information. The exhibition of the cross breed calculation is talked about for irregular initialization, quick combination, hearty division, and to gain shorter CPU time. With the blend of existing procedure, this examination work gives an imaginative methodology, ODNN model to anticipate coronary illness. The created model is tried with the clinical content dataset and delivered results are confirmed by clinical specialists. On breaking down the exhibition and aftereffects of the calculations, mixture ODNN calculation is the best and progressively appropriate for distinguishing proof of coronary illness influenced patients from the clinical content information. The proposed ODNN method is executed with the content of the chosen medical dataset. The result of the proposed ODNN accurately recognizes the coronary illness and the outcomes are enormously acknowledged by physicians. Also the results are verified by the clinical specialists. The particular word which affects the disease in patients dataset is exactly fetch out by the Named Entity Recognition Algorithm. Results from the experiments show that it is identified that the ODNN method predicts the heart disease affected patients very efficiently. In future, the proposed technique can be applied to locate the enormous or increasingly number of dataset all the more precisely. The improvement of other half breed calculations for clinical content mining and testing this cross breed ODNN method for different constant dataset is likewise remembered for what's to come. The work has been reached out to different calculations like bunching calculations and furthermore a portion of the strategies applied to discover the coronary illness impeccably so as to progress the forecast exactness.

Acknowledgements

We sincerely thank our College Management for encouraging and supporting us by providing good infrastructure. Also, we thank Dr. Prabhakaran, Cardio specialist from Apollo Hospital, Chennai for providing constant support and his valuable time.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Zhu, F., Patumcharoenpol, P., Zhang, C., Yang, Y., Chan, J., Meechai, A., Vongsangnak, W. and Shen, B.R. (2013) Biomedical Text Mining and its Applications in Cancer Research. *Biomedical Informatics*, **46**, 200-211. https://doi.org/10.1016/j.jbi.2012.10.007

[2] Harpaz, R., Callahan, A., Tamang, S., Low, Y., Odgers, D., Finlayson, S., Jung, K., Le Pendu, P. and Shah, N.H. (2014) Text Mining for Adverse Drug Events: The Promise, Challenges, and State of the Art. *Drug Safety*, **37**, 777-790. https://doi.org/10.1007/s40264-014-0218-z

[3] Simpson, M.S. and Demner-Fushman, D. (2012) Biomedical Text Mining: A Survey of Recent Progress. Mining Text Data, 465-517. https://doi.org/10.1007/978-1-4614-3223-4_14

[4] Zhou, X.Z., Peng, Y.H. and Liu, B.Y. (2010) Text Mining for Traditional Chinese Medical Knowledge Discovery: A Survey. *Biomedical Informatics*, **43**, 650-660. https://doi.org/10.1016/j.jbi.2010.01.002

[5] Eskici, H.B. and AlpayKoçak, N. (2018) A Text Mining Application on Monthly Price Developments Reports. *Central Bank Review*, **18**, 51-60. https://doi.org/10.1016/j.cbrev.2018.05.001

[6] You, X.G. (2014) Text Mining Software and Their Applications. In *Process of Fourth International Conference on Instrumentation and Measurement*, *Communication and Control*, China, 902-905. https://doi.org/10.1109/IMCCC.2014.189

[7] Shi, G. and Kong, Y. (2009) Advances in Theories and Applications of Text Mining. In *Process of First International Conference on Information Science and Engineering*, Nanjing, China, 4167-4170. https://doi.org/10.1109/ICISE.2009.214

[8] Kumar, B.S. and Ravi, V. (2016) A Survey of the Applications of Text Mining in Financial Domain. *Knowledge-Based Systems*, **114**, 28-147. https://doi.org/10.1016/j.knosys.2016.10.003

[9] Sheng, X.W., Bao, X. and Luo, Y.M. (2016) A Novel Text Mining Algorithm Based on Deep Neural Network. In *Process of Inventive Computation Technologies* (*ICICT*)

International Conference, Vol. 2.
https://doi.org/10.1109/INVENTIVE.2016.7824810

[10] Fleuren, W.W.M. and Alkema, W. (2015) Application of Text Mining in the Biomedical Domain. *Methods*, **74**, 97-106. https://doi.org/10.1016/j.ymeth.2015.01.015

[11] Pletscher-Frankild, S., Palleja, A., Tsafou, K., Binder, J.X. and Jensen, L.J. (2015) DISEASES: Text Mining and Data Integration of Disease-Gene Associations. *Methods*, **74**, 83-89. https://doi.org/10.1016/j.ymeth.2014.11.020

[12] Medhat, W., Hassan, A. and Korashy, H. (2014) Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Engineering Journal*, **5**, 1093-1113.
https://doi.org/10.1016/j.asej.2014.04.011

[13] Agnihotri, D., Verma, K. and Tripathi, P. (2014) Pattern and Cluster Mining on Text Data. In *Process of Fourth International Conference on Communication Systems and Network Technologies*, Bhopal, 428-432.
https://doi.org/10.1109/CSNT.2014.92

[14] Piedra, D., Ferrer, A. and Gea, J. (2014) Text Mining and Medicine: Usefulness in Respiratory Diseases. *Archivos de Bronconeumología*, **50**, 113-119.
https://doi.org/10.1016/j.arbr.2014.02.008

[15] Sarker, A. and Gonzalez, G. (2015) Portable Automatic Text Classification for Adverse Drug Reaction Detection via Multi-Corpus Training. *Biomedical Informatics*, **53**, 196-207. https://doi.org/10.1016/j.jbi.2014.11.002

[16] Lim, S., Tucker, C.S. and Kumara, S. (2017) An Unsupervised Machine Learning Model for Discovering Latent Infectious Diseases Using Social Media Data. *Journal of Biomedical Informatics*, **66**, 82-94. https://doi.org/10.1016/j.jbi.2016.12.007

[17] Haghanikhameneh, F., Panahy, P.H.S., Khanahmadliravi, N. and Mousavi, S.A. (2012) A Comparison Study between Data Mining Algorithms over Classification Techniques in Squid Dataset. *International Journal of Artificial Intelligence* (*IJAI*), Vol. 9.

[18] Zamani, H. and Nadimi-Shahraki, M.-H. (2016) Feature Selection Based on Whale Optimization Algorithm for Diseases Diagnosis. *International Journal of Computer Science and Information Security*, **9**, 1243.

[19] Commandeur, F., Goeller, M., Betancur, J., Cadet, S., Doris, M., Chen, X., Berman, D.S., Slomka, P.J., Tamarappoo, B.K. and Dey, D. (2018) Deep Learning for Quantification of Epicardial and Thoracic Adipose Tissue from Non-Contrast CT. *IEEE Transactions on Medical Imaging*, **8**, 1835-1846.
https://doi.org/10.1109/TMI.2018.2804799

[20] Lucini, F.R., Fogliatto, F.S., da Silveira, G.J.C., Neyeloff, J., Anzanello, M.J., Kuchenbecker, R.S. and Schaanc, B.D. (2017) Text Mining Approach to Predict Hospital Admissions Using Early Medical Records from the Emergency Department. *Medical Informatics*, **100**, 1-8. https://doi.org/10.1016/j.ijmedinf.2017.01.001

[21] Lazard, A.J., Scheinfeld, E., Bernhard, J.M., Wilcox, G.B. and Suran, M. (2015) 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*, **43**, 1109-1111. https://doi.org/10.1016/j.ijmedinf.2017.01.001

[22] Jonnagaddala, J., Liaw, S.-T., Ray, P., Kumar, M., Chang, N.-W. and Dai, H.-J. (2015) Coronary Artery Disease Risk Assessment from Unstructured Electronic Health Records Using Text Mining. *Biomedical Informatics*, **58**, 203-210.
https://doi.org/10.1016/j.jbi.2015.08.003

[23] Ailem, M., Role, F., Nadif, M. and Demenais, F. (2016) Unsupervised Text Mining

for Assessing and Augmenting GWAS Results. *Biomedical Informatics*, **60**, 252-259. https://doi.org/10.1016/j.jbi.2016.02.008

[24] Xia, L., Wang, G.A. and Fan, W.G. (2017) A Deep Learning Based Named Entity Recognition Approach for Adverse Drug Events Identification and Extraction in Health Social Media. In *International Conference on Smart Health*, Springer, Cham, 237-248. https://doi.org/10.1007/978-3-319-67964-8_23

[25] Roygaga, C., Punjabi, S., Sampat, S. and Sarode, T.K. (2018) Neural Networks Application for Detecting Heart Disease. i-Manager'*s Journal on Information Technology*, **3**, 24. https://doi.org/10.26634/jit.7.3.14409

[26] Jabbar, M.A., Chandra, P. and Deekshatuluknowledge, B.L. (2012) Knowledge Discovery From Mining Association Rules For Heart Disease Prediction. *Journal of Theoretical and Applied Information Technology*, **41**, No. 2.

[27] Nazari, S., Fallah, M., Kazemipoor, H. and Salehipour, A. (2018) A Fuzzy Inference-Fuzzy Analytic Hierarchy Process-Based Clinical Decision Support System for Diagnosis of Heart Diseases. *Expert Systems with Applications*, **95**, 261-271. https://doi.org/10.1016/j.eswa.2017.11.001

[28] Latha, U. and Velmurugan, T. (2019) Heart Disease Prediction using Optimal Name Recognition Based on Deep Learning Models and Whale Optimization Algorithms. *Journal of Advanced Research in Dynamic and Control System*, **11**, 808-816.

[29] Thambusamy, V. and Umasankar, L. (2019) Prediction of Heart Disease Using Name Entity Recognition Based on Back Propagation and Whale Optimization Algorithms. 4*th International Conference on Management*, *Engineering, Science*, *Social Science and Humanities & International Journal of Innovative Technology and Exploring Engineering*, **8**, 437–443. https://doi.org/10.35940/ijitee.J1081.08810S19