Online Healthcare Medium for Disease-Treatment using Modified ANN based Classification and Ranking

Mamatha Balipa, Balasubramani R.

Abstract: The fundamental purpose of the healthcare information medium in social networks is centered on ascertaining the opinions of several people regarding specific user queries. In the backdrop of ever-increasing accessibility and attractiveness of the opinion-rich resources as evidenced by the online review sites and personal blogs, the emerging opportunities and challenges dynamically make use information technologies to go in for and to comprehend the outlook of the vast majority of users. However, it is unfortunate that the time-honored finds its waterloo in locating the impending issue of deploying internet with a view to identify and generate appropriate conclusions regarding the specified ailments. The current investigation effectively carries out the function of processing the user query with the able assistance of the MedHelp website and subsequently forwards the pertinent traits to the sentiwordnet for performing the sentimental examination. It is followed by the creation of the score in accordance with the positivity and negativity of the content in the website. In this regard, the Artificial Neural Network (ANN) is ably guided with the aim of creating rank for the websites. And the weight optimization for ANN is elegantly executed by the efficient Grasshopper Optimization Algorithm (GOA). The technique is performed on the powerful platform of JAVA and the consequent outcomes assessed exhibits incredible decrease in the error rate.

Keywords: User query, Sentiwordnet, Artificial Neural Network(ANN), Grasshopper Optimization Algorithm(GOA), error rate.

I. INTRODUCTION

In everyday communication, the network of networks fondly named as the Internet has prominently emerged as an essential element of human existence, projecting itself as the most ideal and unique tool. No day passes without a multitude of people eagerly making intensive searches to quench their thirst for vital and fruitful data. It may be political information, information about government services, and innovative [1]. Of late, it has become a fashion for people to access the treasure house of medical resources offered by the internet through diverse platforms such as the health forums, discussion boards, blogs, and so on. In the course of their surfing, the users are thrown open to an amazing set of links. The user, in turn, will move through the specified links to access the data offered by the users containing a host of details such as the treatment undergone for the specific ailment and the clinic which offered the treatment in plain and simple language and the information is available mostly in the form of text. No wonder, it has become a Herculean task for the people to effectively obtain pertinent data from the corresponding text [2].A close scrutiny of the relative sentences and the detection of the disease-treatment association pave the way for effective ascertainment of the exact form of disease - treatment relation. The recognized disease – treatment association invariably comprises three basic semantic associations such as the cure, prevent and side effects [3]. In this regard, the Natural Language Processing (NLP) technology assumes huge significance with the responsibility of combining functional devices to usher in instantaneous insights to the decision makers regarding the supervision and pharmaco - vigilance functions. A host of endeavors have been carried out for the deft deployment of the smart NLP methods for gathering data on health problems, like the diverse ailments, symptoms, drugs, undesirable events and the like from the texts. [4]. The outstanding among the natural language processing methods consist of the text categorization, medical entity detection and entity relation mining. Naive Bayes and Support Vector Machines (SVM) have come to stay as the most appealing tools for classifying text. With a view to scrutinize natural language and speech, they employ the computational methods and linguistic ideas [5-7]. As regards the NLP data mining, Machine learning algorithms like the conditional random fields (CRFs) have proved their mettle, showcasing supreme efficiency [8]. The mining of informative sentences has itself emerged as an effective task in the domain of the NLP and ML [9]. In fact, there is a flood of Information Extraction techniques [10] which are in vogue for the purpose of mining relevant data from the appropriate online forums, though there are hassles in ascertaining a structured way for the purpose of saving the fruitful data. With ever-increasing quantity of data accessed and collected from several diverse sources, the paramount necessity for having a structure cannot but be overemphasized and there should also be certain type of semantics integrated with the data to enable it to be processed by independent agents obviating the requirement for human intervention [11].

Retrieval Number: C6535029320/2020©BEIESP
DOI: 10.35940/ijeat.C6535.029320
Journal Website: www.ijeat.org

4029

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

II. LITERATURE REVIEW

Hariprasad Sampathkumar et al by [13] worked towards bringing to light to the use of Hidden Markov Model based text mining mechanism which had the wherewithal of mining undesirable responses of drugs and was set up on the basis of analysis of content accessed from the online healthcare forums. Their mechanism was able to segregate the message as having drug side-effect data and thereafter mine the undesirable side-effect mentioned from it. However their technique failed to discern certain novel undesirable side-effect data which was capable of functioning as advance indicators for the health authorities so as to assist them in their endeavors towards the Pharmacovigilance.

Credit goes to Xiangfeng Dai et al [14] for proposing an innovative word embedding based clustering technique. In those days, the Word embedding held the sway as one of the most fashionable styles in the domain of the Natural Language Processing (NLP). It was equipped with the significant skills of comprehending the optimal vectors from the neighboring, and the vectors, in turn, were capable of referring the semantic information of words. Further, a tweet had the option of being referred as little vectors, so that they can be segregated into groups of identical words. Subjecting the tweet to the identical comparison measures of the entire groups, it was possible to categorize them as associated with a subject such as influenza or not. It was worth-mentioning that their novel technique scaled higher levels of accomplishment, by ushering in utmost precision, obviating the necessity for labor for labeling the training data, with the convenience of being extensively applied to the extant categorization issues or similar ailments.

Yin Kang et al [15] was instrumental in developing a rule-based technique named RubE which was capable of mining the subjective features as well as the objective features from reviews posted by consumers. With indirect dependence as well as relative structure, the objective traits were queried by integrating the relation that is part–whole as well as patterns that are review specific and subjective traits were mined by way of extension of the double propagation. Their approach was found to excel in the parallel hi-tech methods as regards the product attribute mining. The Feature Selection(FS) technique intended for scaling up the Classification Performance of Support Vector Machines in Text Mining was the brainchild of S. Manochandar et al [16]. Their innovative technique was able to project a performance of superior accuracy, precision, and recall values with respect to whole datasets. The general strength of the expressions (SE (i)) was effectively evaluated by way of integrating the appraisal parameters and efficacy, which were exponentially weighted established on the representative rank positions of the diverse term weighting expressions. Their novel approach succeeded in achieving amazing accomplishment vis-à-vis those of the peer techniques in all cases admitting the mining of the less significant trait set so as to enhance the efficiency of the categorization method, thereby establishing without an iota of doubt that the recommended term weighting expressions were incredibly superior to those of the time-tested techniques for FS.

An efficient technique for the purpose of automatically locating the design and requirement self-admitted technical debt applying Natural Language Processing (NLP) was estimated by Everton da S.et al [17] Maldonado. The self-admitted technical debt could be identified using their method which uses fixed keywords and phrases. In this regard, the words associated with source code of mediocre quality emerged as the ideal pointers of design debt. Their innovative method provided excellent outcomes, realizing a whopping 90% of the categorization performance, employing an insignificant 23% of total comments for the design as well as requirement self-admitted technical debt, and 80% of the performance, utilizing a trivial 9% and 5% of the comments for design and requirement self-admitted technical debt, correspondingly. The key merit of their novel approach was its acumen to achieve an amazing accuracy even with a comparatively minute training dataset.

The Patient Reported Information Multidimensional Exploration (PRIME) framework, intended for the extraction, evaluation, and comparison self-reported data from Online Cancer Support Groups (OCSG) by insignificant intermediate-risk prostate cancer (PCa) enduring radical prostatectomy (RP), external beam radiotherapy (EBRT), and active surveillance (AS), and for the appraisal of its efficiency in quality-of-life (QoL) and emotion evaluates was effectively brought out by TharinduBandaragoda et al [18].

The scientific community acknowledged the work of Jing Liu et al [19] who developed the innovative ensemble learning technique named the stratified sampling-based random subspace method, (SSRS) which was dedicated to successfully tackling the hassles connected with the social-media Adverse Drug Event(ADE) identification. Their admirable algorithm was competent to outshine the parallel ensemble approaches and the modern social media-based ADE identification strategies. The superlatively efficient social media-based ADE identification could effectively boost the modern inactive systems and delivers regulatory authorities with decision aids, extend a helping hand to support patient empowerment, and achieve the data-driven medication security administration. Their technique was incredibly simple to execute thanks to the fact it was based on two distinct trait sets which were naturally obtained and were ever-accessible in the context of social media-based ADE recognition. As a result, the novel method was competent to successfully extract ADEs from social media.

A semi-supervised ensemble learning framework for mining adverse drug events (SSEL-ADE), which integrated diverse categories of traits and methods, admitting ensemble learning and semi-supervised learning, for mining adverse drug events from social media was intensely assessed by Jing Liu et al [20]. For improved ADE relation mining competencies, the novel technique exhibited excellent execution outdating the parallel approaches, thereby showcasing the feasibility of applying semi-supervised ensemble learning. The peak AUC value of 81.48% was realized by employing the Co-RS on the whole feature set SSEL-ADE. Further their technique was able to efficiently mine adverse drug events from the social media, and modernize the modern inert mechanisms, thereby offering the regulatory authorities decision tools, like the market withdrawals, drug recalls, and safety alerts, in addition to playing a significant role in considerably cutting back the monetary and legal loss risks for the pharmaceutical companies.

Moreover, the semi-supervised ensemble techniques were endowed with the requisite algorithms so that they could be fruitfully deployed to successfully cope with the related social media-based issues..

III. PROBLEM DEFINITION

The yesteryears have been silent spectators to the phenomenal growth of the online health social websites, which have appeared like a god-given gift and to al lot of people. When a user is on the web, making his searches for cure, or healing solutions which are offered online for a specific ailment, with the able assistance of a search engine, he is usually flooded with a list of links which have to be gone through to comprehend the details of information recorded by other users indicating the place of their treatment such as the hospital or clinic for the specific ailment and which were beneficial for them. Thus he is presented with massive amount of data such as huge number of questions regarding the ailment including the unsuccessful treatments, and the methods to attain the related data. Therefore, by way of extracting the treatments solutions offered by various users for the ailment through the web would make it very simple and convenient. Generally, the related postings are carried out by average users, there is the possibility for the occurrence of noise, and the data may be imprecise and inflated with errors in creation of words including spelling mistakes, and hence extracting such data paves the way for the false positives. However, the frequency of data may be helpful in finding a solution to the problem, as repeatedly occurring treatments may be construed as the true positives. Certain messages are circulated in the web by the vested interests for advocating the drugs or products produced by them. Nevertheless, the paucity of proper solutions regarding the health issues has acted as motivating factor for carrying the current investigation in this arena.

IV. PROPOSED METHODOLOGY

In this investigation, an intelligent endeavor is made to give shape an innovative online healthcare solution extracting system for the purpose of enabling prospective users get in-time and improved treatment and healing care for their ailments. The innovative technique automatically offers quality medications from the websites and is devoted to offer the most fruitful recommendations from the healthcare professionals. At the outset, the novel technique performs the task of ascertaining the subject for the request is the user query from MedHelp website. Thereafter, in tune with the pertinent traits such as the comments, followers, the data accessed from the web are duly furnished to the sentiwordnet with the intention of recognizing the positivity and negativity in the data obtained, for which appropriate scores are allotted by the sentiwordnet. In accordance with the relative scores and related traits, the modified ANN is employed to determine the solution or non-solution pages with its rank. Further, the weight of the ANN gets optimized by means of the Grasshopper Optimization Algorithm (GOA). In the long run, both the medications and the recommendations from the healthcare professionals are effectively gathered from the relevant page and offered to the end users.
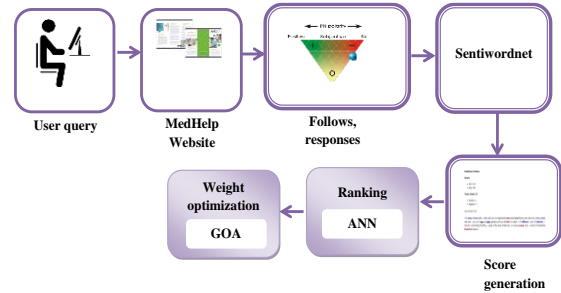

Fig. 1.Proposed block diagram

A. User Query:

A user query, in quintessence, represents the transactional query which smartly offers an introduction for the purpose of effectively executing any operations via the web like the web search, product purchase, downloading files or taking intelligent decisions in accordance with the reviews. In this backdrop, in the current investigation, the user prefers a request for attaining any data connected with diseases and medication thereof. The system effectively responds back by sending a detailed account of websites for results offering the related services.

Text analysis:

By the term 'Text analytics', what is meant, is the gathering of precise data from a gigantic quantity of data like a specific post in a website or from an evaluation of a product. When the specific bit of data is searched, Lexalytics a tool, proceeds to evaluate the text with the aim of offering a transparent version of the data such as the authors/orators, the subject of their discussion, their feelings and their recommendations/advices/conclusions, the questions assuming the forms of "who is talking and what they are talking about, what they feel about it, and what they say about it?", which are assessed by the core text analytics such the themes, categories, named entity mining, objectives, sentiment scrutiny and summarizations. The relative results are returned in terms of websites for the query preferred by the user. In tune with the traits like the followers and comments (positive, negative), the content is effectively subjected to the sentiwordnet investigation.

B. Sentiwordnet:

The Sentiwordnet elegantly executes the task of the sentiment scrutiny on the data to come to a tentative conclusion regarding the attitude of the clients regarding the product. Thereafter, it proceeds with the process of categorizing the positive and negative content leading to the creation of the appropriate scores in accordance with the investigation.

C. Score Generation:

The scores are invariably produced depending on the positivity and negativity remarks in the content. As a rule, the scores are efficiently produced as illustrated in Equations (1) and (2) given below [21]:

$$Positivity = \frac{Number\ of\ classifier\ stating\ positive}{Number\ of\ classifier} \quad (1)$$

$$\text{Negativity} = \frac{Number\ of\ classifier\ stating\ negative}{Number\ of\ classifier} \qquad (2)$$

The resulting scores are thereafter elegantly trained in the ANN with the intention of generating the ranks.

D. Ranking utilizing Artificial Neural Network:

The Artificial Neural Network (ANN) is efficiently trained so that it determines ranks in accordance with the scores and features. The ANN, in essence, represents an artificial intelligence technique which is competent to replicate the functioning of the human mind and computation efficiency, and reflect the fundamental linear or non-linear connections among input and target data. By and large, it is multi-layered, and represents a combination of three layers network housing colossal number of neurons which performs the task of efficiently adapting the input data into expected output. The proposed work has three distinct layers such as the input layer, output layer and hidden layer, which are linked with specific number of neurons. The neurons in each layer, in turn, are associated with certain synaptic weights. The data from the input layer is initially forwarded to the hidden layer and subsequently to the output layers. The number of hidden layers neurons is arbitrarily varied until the error value realized is the least one [22]. The output neurons are invariably predetermined. In the training algorithm, synaptic weights of the neurons are refreshed in every cycle till the error among output and input is within tolerance. Fig 2 beautifully pictures the broad ANN configuration.
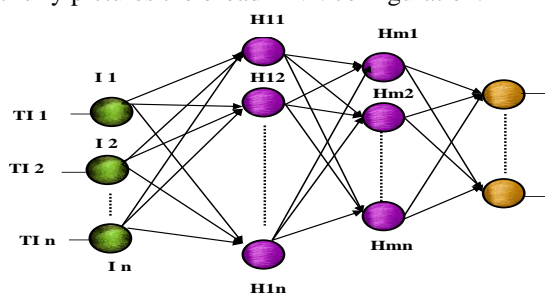

Fig. 2. General structure of Artificial Neural Network[22]

Structure initialization

Initialization processes the inputs based on the input layer weight α_j and the hidden layer weights β_{ij}. Input I_i represents the comments or messages.

Input layer

The input layer embraces a multitude of neurons, which are associated with the hidden layer. The inputs are the text taken for training the network and the input neurons are $i_1, i_2, \ldots \ldots i_n$. The inputs are $W_1, W_2, \ldots \ldots W_n$ each neuron possesses the weight which is referred as the ith input layer neuron combined with the jth neuron of the hidden layer like $\beta_{12}, \beta_{12}, \ldots \ldots \beta_{ij}$.

Hidden layer

The hidden layer is home to a number of neurons labeled as $h_1, h_2, \ldots \ldots h_n$. The hidden layers are connected to the output layer by using the neurons [23]. Weight obtaining equation is α_i where L_l a number of input layers and the weight α_j and β_{ij} based on the activation function.

The relationship between the input trait d, and output of the first hidden layer, h_1 is effectively evaluated by means of the following Equation (3).

$$h_1 = A(w_1 d + bias_1) \qquad (3)$$

Where, w_1 and $bias_1$ represent the weight and bias at the first hidden layer. Further, $A(\cdot)$ illustrates the activation function.

Activation function

This function is capable of assuming two forms such as the linear, threshold or sigmoid function, the latter i.e., the sigmoid activation function being extensively employed for the hidden layer as it equates more or less linear conduct, curvilinear conduct and more or less steady conduct largely based on the input value. The activation function in ANN modeling process is given by means of the following Equation (4).

$$A_f = \sum_{j=1}^{h} \alpha_j * \left(\frac{1}{1 + \exp(-\sum_{i=1}^{N} M_i \beta_{ij})} \right) \qquad (4)$$

Where, F$_i$ signifies a fitness function α and β represent the weighted values, W indicates the input parameters, i depicts the number of inputs, j reveals the number of weights and characterizes the number of hidden neurons and N indicates the number of data. Based on this equation the fitness value is assessed.

Output layer

The output layer consists of a large number of neurons. The hidden layer neurons are connected to the output layer by means of these neurons. Each connection is credited with a weighted value like $\alpha_1, \alpha_2, \ldots \ldots \alpha_n$.

The fundamental function of the output units is vividly pictured in the following Equation (5).

$$O_l = \sum_{i=1}^{n} \alpha_i \sigma(F_{i(optimal)}) \qquad (5)$$

$$l = 1 \quad and \quad i \in [1, \ldots n]$$

Further, the error value is effectively estimated by means of the Equation (6) and is represented as E_i where, α and β characterize the weighted range from -50 to 50, 'i' signifies the number of inputs, 'j' reveals the number of weights and h indicates the number of hidden neurons.

$$E_i = \sqrt{\frac{\sum_{i=1}^{ND}(D_i - P_i)^2}{ND}} \qquad (6)$$

Where ND corresponds to the number of the data, D represents the desired value and P signifies the predicted value, i= 1, 2,.n. With the deployment of this Equation, the error value is derived as the difference between desired value and the predicted value.

A. Weight optimization utilizing Grass hopper Optimization Algorithm (GOA)

The Grasshopper optimization, in essence, represents a population based swarm intelligence technique which is the brainchild of Mirjalili et. al [25] and it imitates the social conduct of the grasshoppers. Incidentally, the grasshoppers are the insects which cause untold damages to the crops and adversely affect the agriculture. Their lifespan proceeds through two distinct phases such as the nymph and adulthood. In the nymph stage, the grasshoppers suffer from shortage of wings and consequently make sluggish movements to consume all the crops found during their journey. On attaining adulthood, they are competent to move swiftly to far distant places. These grasshoppers form large swarm of continental scale and it threatens farmers. A large number of nymph grasshoppers hop and travel just as the rolling cylinders. The vital attribute of the swarm technique is the sluggish movement of grasshopper during the larval stage and rapid movement in the adulthood. In this regard, hunt for food source emerges as one of the key attribute in the swarming of grasshoppers. The search procedure is segregated into distinct categories such as the exploration and exploitation phases. Conceived as a population based technique, each grasshopper effectively showcases a solution in the population. With the result, it is possible to scientifically model the nature-motivated conduct of the grasshopper as illustrated below.

The position update of each and every grasshopper largely relies on three distinct forces like the social interaction represented by 'si', gravitational force characterized by 'Gi' and wind advection force denoted by 'Ai'. The modeling of grasshopper employed for position X_i update is depicted below Equation (7)[26]:

$$X_i = S_i + G_i + A_i \qquad (7)$$

Social Interaction:
The social interaction 'S$_i$' is duly characterized by means of Equation (8) given hereunder.

$$S_i = \sum_{j=1}^{N} S(d_{ij})\,\overline{d}_{ij} \qquad j \neq i \qquad (8)$$

Where d_{ij} depicts the distance between i^{th} and j^{th} grasshopper and is represented as $d_{ij} = |\,x_j - x_i\,|$. 's' signifies the strength of social forces. $\overline{d}_{ij} = \dfrac{x_j - x_i}{d_{ij}}$

represents a unit vector from i^{th} and j^{th} grasshopper and 'N' indicates the number of grasshoppers. The social force is effectively illustrated by means of the following Equation (9) in terms of attraction and repulsion:

$$S(r) = f\, e^{\frac{-r}{l}} - e^{-r} \qquad (9)$$

Equation 9
Where 'f' indicates the intensity of attraction and 'l' stands for the length of attractive scale.

Gravitational Force:
The gravitational component 'G$_i$' is convincingly computed by means of the Equation (10) given below.

$$G_i = -\,g\hat{e}_g \qquad (10)$$

Where 'g' indicates the gravitational constant and '$\hat{e}_g$' represents the unit vector towards the center of attraction.

Advection Force:
The advection component 'A$_i$' is aptly assessed as evident from the following Equation (11).

$$A_i = u\hat{e}_w \qquad (11)$$

Where 'u' stands for a constant drift '$\hat{e}_w$' exhibits a unit vector in the wind direction. The swarming conduct of the grasshoppers with the three forces gets illustrated as per the Equation appearing below.

$$X_i = \sum_{j=1}^{N} s\left(\left|x_j - x_i\right|\right)\frac{x_j - x_i}{d_{ij}} - g\hat{e}_g + u\hat{e}_w \qquad (12)$$

It is likely that the grasshoppers tend to arrive at the comfort zone in a swift manner and also the swarm fails miserably to converge to a specific point. These hassles are successfully managed by the optimization technique as exhibited in Equation (13) below.

$$X^d_{\ i} = c\left(\sum_{j=1}^{N} c\,\frac{ub_d - lb_d}{2} s\left(\left|x^d_{\ j} - x^d_{\ i}\right|\right)\frac{x_j - x_i}{d_{ij}}\right) + \hat{T}_d \qquad (13)$$

Where 'ub$_d$' and 'lb$_d$' respectively represent the upper bound and the lower bound in the Dth dimension. $\hat{T}_d$ Specifies target value. 'c' signifies the decreasing coefficient for dwindling comfort zone, attraction zone and repulsion zone. At this junction, gravitational force is absent and the wind direction is presumed to be towards the target ($\hat{T}_d$).

For each iteration, the subsequent location of the grasshopper is determined by taking into account the location of target, current location and location of all other grasshoppers.

Comfort Zone solution update: The coefficient 'c' incredibly cutbacks the comfort zone in relation to the number of iterations 'L' as illustrated in Equation (14) below

$$c = c_{max} - l\,\frac{c_{max} - c_{min}}{L} \qquad (14)$$

Where c$_{max}$ and c$_{min}$ represent the highest and lowest values of comfort zone, 'l' characterizes the existing iteration.

Optimal fitness generation:
The new solutions are produced in accordance with the procedure shown above. With respect to each and every solution produced, the best fitness value is effectively estimated for the new solution created. The minimized error rate is employed as optimal solution in the current investigation.

V. RESULT AND DISCUSSION

Here the outcomes for online healthcare medium for ailment with the deft deployment of the modified ANN based classification and ranking are vividly described.

A. Graphical User Interface (GUI)

Shown below are the GUI's attained subsequent to processing of the outcomes for online healthcare medium for ailment cure by effectively employing ANN and GOA.
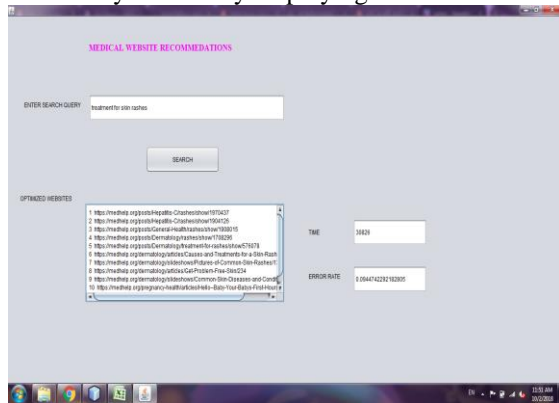

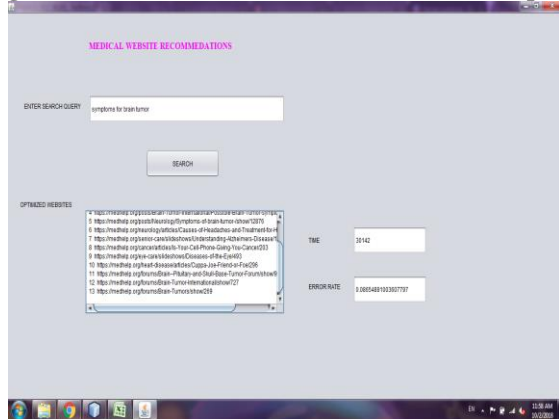**Fig. 3.GUI achieved for the cure of skin rashes query**

Fig. 4.GUI attained for the symptoms of brain tumor

B. Evaluation Metrics

The tools like the response time and error rate are evaluated with the intention of illustrating the effecient classification and ranking for online healthcare medium for ailment cure by employing ANN- GOA.

Response Time

The response time represents the duration needed for the purpose of reacting to any action or event in phase with the user demand or service, illustrating the dimension of system efficiency.

Error rate

By error rate what is meant is the divergence between the estimated value and an identified value. Thus the error rate showcases the deviation between experimentation and expected values.

C. Performance Metrics

The efficacies of the novel and current methods for superior categorization and ranking in online healthcare medium for ailment cure employing the ANN- GOA are brought out below.

Table 1: Follows and responses attained from users for each and every query

Query	Follows			Responses		
	ANN-GOA	ANN	ANN-GA	ANN-GOA	ANN	ANN-GA
Treatment for skin rashes	1	1	1	6	5	6
Reasons for retinal disorders	1	1	1	43	43	43
Symptoms for brain tumor	2	1	2	3	5	3
Problems in preterm birth	1	1	1	4	4	3
Treatment for muscular dystrophy	1	1	1	2	2	2

Discussion:

Table 1 shows with the follows and responses for the searched queries. Here, the Follows exhibit the number of users following the comment from another user. In contrast, the responses represent the frequency of visitors/users who have replied to the comment made by the specific users. The corresponding follows and response measure values achieved by the innovative and time-honored techniques in the course of their functioning are vividly described.

Table 2: Response time (ms) and Error rate attained for the innovative ANN- GOA and the conventional ANN, ANN- GA methods

Query	Response time (ms)			Error rate		
	ANN-GOA	ANN	ANN-GA	ANN-GOA	ANN	ANN-GA
Treatment for skin rashes	30826	41254	36454	0.09448	0.099484	0.095454
Reasons for retinal disorders	153957	185454	162548	0.09845	0.099944	0.099244
Symptoms for brain tumor	30142	42315	35264	0.08654	0.096854	0.091478
Problems in preterm birth	41594	48657	43268	0.074888	0.09154	0.089444
Treatment for muscular dystrophy	17402	20145	18957	0.052342	0.086454	0.062854

Discussion:

In Table 2, the response time (ms) and error rate realized for the novel and traditional methods are listed. The average duration of time consumed in replying to the preferred queries in the novel ANN- GOA is shown to be *54784 ms as against the values of* 67565 *ms and* 59298 *ms* for the current ANN and ANN- GA approaches. Further the error rate for the projected technique is a mere *0.08134* vis-à-vis the values of *0.094855* and *0.087695* for the traditional ANN and ANN- GA techniques correspondingly. Thus it is clear that the novel technique projects an incredibly lower value for the response time duration, thereby it steals a march over its peers by accomplishing the response in the shortest possible time. In the case of the error rate also, the projected approach comes out with a successful performance by scaling down the error rate to the least possible, due to the training of the network for ranking, pushing the modern methods to the behind weighed in with their high error rates. Depicted below in Figures 5 and 6 are the comparison graphs for the response time and error rate.
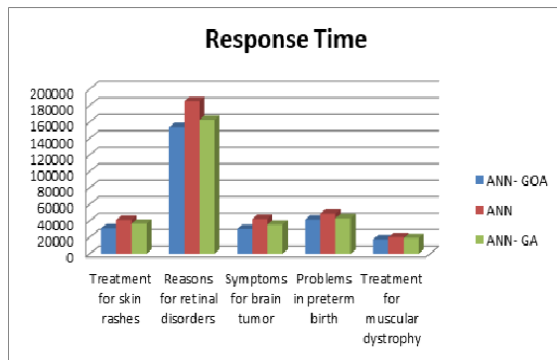
Fig. 5.Comparison graph for the response time achieved by the novel ANN- GOA and conventional ANN, ANN- GA methods.
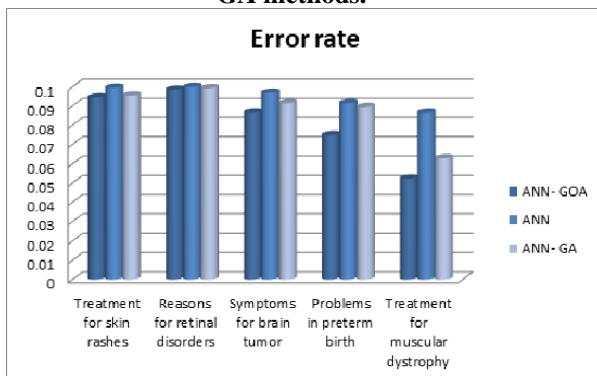

Fig. 6.Figure 6: Comparison graph for the error rate ushered in by the novel ANN- GOA and the modern ANN, ANN- GA methods.

VI. CONCLUSION

The document offers a view to building a system for an online healthcare medium for ailment cure with the deft deployment of an innovative technique technique viz. the ANN- GOA. Here, the user query is duly processed with the help of the MedHelp website. The queries are furnished to the sentiwordnet for the purpose of effectively assessing the scores for the websites. In accordance with the scores produced, the ANN is duly trained with the aim of ranking the websites. In the long run, the user receives the best ranked website as the response for his query. The task of optimization of the weight in the neural network is duly discharged by the Grasshopper Optimization Algorithm (GOA). The encouraging outcomes proclaim the decreased error rate of around 0.08134. There is huge potential for escalating the efficacy of mechanism by way of encrypting and saving the resultant outcomes for future use.

REFERENCES

1. Christian Halim, Alfan F. Wicaksono, MirnaAdriani,"Extracting Disease-Symptom Relationships from Health Question and Answer Forum",International Conference on Asian Language Processing (IALP),2017
2. Ms.MamathaBalipa, Dr.BalasubramaniR,"Disease-Treatment Relationship Extraction for Psoriasis from Online Healthcare Forums using NLP and Classification Techniques", International Journal of Applied Engineering Research, pp. 3568-3573,2018
3. Keerrthega.M.C,Ms.D.Thenmozhi," Identifying Disease -Treatment Relations using Machine Learning Approach", Procedia Computer Science, Volume 87, Pages 306-315,2016
4. Paloma Martınez , Jose L. Martınez , Isabel Segura-Bedmar , Julian Moreno-Schneider , Adrian Luna , Ricardo Revert," Turning user generated health-related content into actionable knowledge through text analytics services"Computers in Industry, Volume 78, Pages 43-56,2016
5. Xiao Liu , Hsinchun Chen," A research framework for pharmacovigilance in health social media: Identification and evaluation of patient adverse drug event reports",Journal of Biomedical Informatics, Volume 58, Pages 268-279,2015
6. Xiao Liu, Jing Liu , Hsinchun Chen," Identifying Adverse Drug Events from Health Social Media: A Case Study on Heart Disease Discussion Forums"International Conference on Smart Health, pp 25-36, 2014
7. Xiao Liu , Hsinchun Chen,"AZDrugMiner: An Information Extraction System for Mining Patient-Reported Adverse Drug Events in Online Patient Forums"International Conference on Smart Health, pp 134-150, 2013
8. AbulHasan , Mark Levene, David J. Weston," Natural Language Analysis of Online Health Forums"International Symposium on Intelligent Data Analysis, Advances in Intelligent Data Analysis XVI pp 125-137, 2017
9. OanaFrunza, Diana Inkpen, Thomas Tran," A Machine Learning Approach for Identifying Disease-Treatment Relations in Short Texts",IEEE Transactions on Knowledge and Data Engineering , Volume: 23, Issue: 6, 2011
10. Ferrara, P. De Meo, G. Fiumara, and R. Baumgartner, "Web data extraction, applications and techniques: A survey," Knowledge-Based Systems, vol. 70, pp. 301–323, 2014
11. HariprasadSampathkumar ,Xue-wen Chen ,Bo Luo," Ontology-based visualization of healthcare data mined from Online Healthcare Forums", International Conference on Healthcare Informatics,2015
12. A.Akilan," Text Mining: Challenges and Future Directions",International Conference on Electronics and Communication Systems (ICECS),2015
13. HariprasadSampathkumar, Xue-wen Chen2 and Bo Luo," Mining Adverse Drug Reactions from onlinehealthcare forums using Hidden Markov Model"BMC Medical Informatics and Decision Making, 2014
14. Xiangfeng Dai, Marwan Bikdash, Bradley Meyer," From Social Media to Public Health Surveillance: Word Embedding based Clustering Method for Twitter Classification", SoutheastCon 2017
15. Yin Kang, Lina Zhou,"RubE: Rule-based Methods for Extracting Product Features from Online Consumer Reviews",Information& Management, Volume 54, Issue 2, Pages 166-176,2017
16. S. Manochandar," Scaling Feature Selection Method for Enhancing the Classification Performance of Support Vector Machines in Text Mining", Computers & Industrial Engineering, Volume 124, Pages 139-156,2018
17. Everton da S. Maldonado, EmadShihab, NikolaosTsantalis," Using Natural Language Processing to Automatically Detect Self-Admitted Technical Debt", IEEE Transactions on Software Engineering , Volume: 43, Issue: 11, 2017
18. TharinduBandaragoda,WeranjaRanasinghe, AchiniAdikari, Daswin de Silva, Nathan Lawrentschuk, DammindaAlahakoon, Raj Persad, Damien Bolton,"The Patient-Reported Information Multidimensional Exploration (PRIME) Framework for Investigating Emotions and Other Factors of Prostate Cancer Patients with Low Intermediate Risk Based on Online Cancer Support Group Discussions",Annals of Surgical Oncology, Volume 25, Issue 6, pp 1737–1745,2018
19. Jing Liu, Gang Wang,"Pharmacovigilance from social media: An improved random subspace method for identifying adverse drug events", International Journal of Medical Informatics, Volume 117, Pages 33-43,2018
20. Jing Liu, Songzheng Zhao , Gang Wang,"SSEL-ADE: A semi-supervised ensemble learning framework for extracting adverse drug events from social media", Artificial Intelligence in Medicine, Volume 84, Pages 34-49,2018
21. Ali, Farman, DaehanKwak, Pervez Khan, SM Riazul Islam, Kye Hyun Kim, and Kyung Sup Kwak. "Fuzzy ontology-based sentiment analysis of transportation and city feature reviews for safe traveling." *Transportation Research Part C: Emerging Technologies* 77 (2017): 33-48.
22. Fister, Iztok, PonnuthuraiNagaratnamSuganthan, Salahuddin M. Kamal, Fahad M. Al-Marzouki, MatjažPerc, and DamjanStrnad. "Artificial neural network regression as a local search heuristic for ensemble strategies in differential evolution." *Nonlinear Dynamics* 84, no. 2 (2016): 895-914.

23. Sadegh, Hosseini, AhmadiNajafabadi Mehdi, and Akhlaghi Mehdi. "Classification of acoustic emission signals generated from journal bearing at different lubrication conditions based on wavelet analysis in combination with artificial neural network and genetic algorithm." *Tribology International* 95 (2016): 426-434.
24. Lawan, S. M., W. A. W. Z. Abidin, T. Masri, W. Y. Chai, and A. Baharun. "Wind power generation via ground wind station and topographical feedforward neural network (T-FFNN) model for small-scale applications." *Journal of cleaner production* 143 (2017): 1246-1259.
25. Mirjalili, Seyedeh Zahra, SeyedaliMirjalili, ShahrzadSaremi, HossamFaris, and Ibrahim Aljarah. "Grasshopper optimization algorithm for multi-objective optimization problems." *Applied Intelligence* 48, no. 4 (2018): 805-820.
26. Saremi, Shahrzad, SeyedaliMirjalili, and Andrew Lewis. "Grasshopper optimisation algorithm: theory and application." *Advances in Engineering Software* 105 (2017): 30-47.

AUTHORS PROFILE

Ms. Mamatha Balipa, Associate Professor, Department of MCA, NMAM Institute of Technology, Nitte, MCA, (Ph.D), MCA, Email:mamathabalipa@nitte.edu.in, Publications: MamathaBalipa and Balasubramani R., "Search Engine using Apache Lucene", IJCA (International Journal of Computer Applications), ISBN : 973-93-80889-58-8, DOI: 10.5120/ijca2015906476, pp 27-30, Volume 127, Number 9, 2015, Foundation of Computer Science (FCS), NY, USA., Balipa, Mamatha, and R. Balasubramani. "Search engine using apache lucene." International Journal of Computer Applications 127.9 (2015): 27-30.
Balipa, Mamatha. "OPINION ANALYSIS OF TREATMENTS FROM ONLINE HEALTHCARE FORUMS FOR THE DISEASE PSORIASIS USING NLP AND ARTIFICIAL NEURAL NETWORK." Development 5.07 (2018).
Balipa, Ms Mamatha, and R. Balasubramani. "Disease-Treatment Relationship Extraction for Psoriasis from Online Healthcare Forums using NLP and Classification Techniques." International Journal of Applied Engineering Research 13.6 (2018): 3568-3573.
Mamatha Balipa, Dr Balasubramani R, Harolin Vaz, Christina Shilpa Jathanna, Text Summarization for Psoriasis of Text Extracted from Online Health Forums using TextRank Algorithm, International Journal of Engineering & Technology, 7 (3.34) (2018) 872-873
Mamatha Balipa, Dr. Balasubramani R, Retrieval of Disease-Treatment Information from Online Forums Using NLP and Convoluted Neural Network Pipeline, International Journal of Research in Advent Technology, Vol.7 , No.1, January 2019, E-ISSN: 2321-963

Dr. Balasubramani R, Professor, EDC Head, Department of Information Science and Engineering, NMAM Institute of Technology, Nitte, Ph.D. in Content Based Imaged Retrieval for Stock Photography using Semantic and Visual Descriptors from Vinayaka Missions Research Foundation (Deemed-to-be-University) Salem in 2011, M.Tech. in Information Technology from Allahabad Agricultural Institute (Deemed-to-be-University), Allahabad in 2005. Qualified in the AICTE Deemed to be University Examination 2018. B.E. in Electronics and Communication Engineering from Madurai Kamaraj University, Madurai in 1990, Email: balasubramani.r@nitte.edu.in, Publications: "Attainment of Programme Outcomes through Course Outcomes in Outcome Based Education: A Case Study", Journal of Engineering Education Transformations (JEET) | Online ISSN: 2394-1707, Print ISSN: 2349-2473 | DOI: 10.16920/jeet/2017/v31i2/119555 |pp 26-30 | Volume: 31 | No: 02 | October 2017.personal profile which contains their education details, their publications, research work, membership, achievements, with photo that will be maximum 200-400 words.
Balasubramani R., Devidas, "Enhancing Employability Skills through Infosys Campus Connect Programme: A Case Study", JEET (Journal of Engineering Education Transformations), Online ISSN: 2394-1707, Print ISSN: 2349-2473, DOI: 10.16920/jeet/2015/v29i1/77113, pp 60-66, Volume 29, Issue 1, July 2015.
Balasubramani R, Manasa I P, "Multi-Level Network Scheduling Scheme for Wireless Sensor Network", IJSR-International Journal of Scientific Research, ISSN No. 2277 – 8179, pp 282-285, Volume: 3 , Issue: 5, May 2014.
Balasubramani R, Kannan V, "Semantic Annotation of Stock Photography for CBIR using MPEG-7 standards",Global Journal of Computer Science & Technology 2009(GJCST 2009), Publisher: Global Journals Inc.,(USA), ISSN: 0975-4172(Online), 0975-4350(Print), pp 7-11 , Vol. 9 , Issue 5 , September 2009.
Balasubramani R, Kannan V, "Semantic Search and Retrieval of Stock Photography based on MPEG-7 Descriptors", Global Journal of Computer Science & Technology 2009(GJCST 2009), Publisher: Global Journals Inc.,(USA), ISSN: 0975-4172(Online), 0975-4350(Print), pp 146-156 , Vol. 9 , Issue 5 , September 2009.
Balasubramani R, Kannan V, "Efficient use of MPEG-7 Color Layout and Edge Histogram Descriptors in CBIR Systems", Global Journal of Computer Science & Technology 2009(GJCST 2009), Publisher: Global Journals Inc.,(USA), ISSN: 0975-4172(Online), 0975-4350(Print), pp 157-163 , Vol. 9 , Issue 5 , September 2009.