ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

Symptoms Wise Disease Inference System Using Data Mining Technique

Tejal P. Burange[1], Prashant N. Chatur[2]

M.Tech Student, Department of Computer Science and Engineering, Government College of Engineering, Amravati, MH, India [1]

Head, Department of Computer Science and Engineering, Government College of Engineering, Amravati, MH, India[2]

ABSTRACT: Rare diseases are chronic diseases. Some of them are curable if detected in early stages. Rare disease symptoms are very common symptoms that can be ignored by the patients. In previous literature, the importance of rare disease discovery in early stage is discussed but practical solution is not given. In some literatures disease discovery system is given but solutions after disease discovery is not given. To improve disease discovery system proposes new disease discovery system. In proposed system, symptoms based disease discovery system which discovers diseases on the basis of deep learning algorithm. This system proposed a user friendly system in which patients will get suggestion about symptoms in their own language. This system proposes a medicine discovery system to implement personalized disease prevention techniques recommendation system. After disease discovery, the most relevant diseases will be shown to the patients along with detailed information on each disease. The patient will be able to verify his detailed symptoms with details of every disease. In case of any query, the patient can post queries to experts, the experts then reply to them. Along with disease discovery, we propose a intelligent question answer system which immediately deliver the questions and answers most relevant with the query/questions posted by the patient. System propose a medicine discovery system, to provide medicine order placing/processing facility for patients using which the suffered patient will be able to find out the medicals (with Google maps) where the corresponding medicines are available. The patients could place medicine order online to achieve anonymity.

KEYWORDS: Symptoms; Experts; Query

I. INTRODUCTION

Rare diseases are chronic diseases. Some of them are curable if detected in early stages. Rare disease symptoms are very common symptoms that can be ignored by the patients. In previous literature, the importance of rare disease discovery in early stage is discussed but practical solution is not given. In some literatures disease discovery system is given but solutions after disease discovery is not given. To improve disease discovery system we propose new disease discovery system. In this system, we propose symptoms based disease discovery system which discovers diseases on the basis of deep learning algorithm. Deep Learning is a subfield of machine learning which is concerned with algorithms and also inspired by the structure and brain of a function called artificial neural networks. Deep learning is a class of machine learning algorithms:

1) Which is use for feature extraction and transformation of a multiple layers of nonlinear processing units. So that each successive layer uses the output from the previous layer as input.
2) Supervised or unsupervised algorithm also applied. Pattern analysis (unsupervised) and classification (supervised) are included in applications.
3) Are based on the (unsupervised) learning of multiple levels of features or representations of the data. To form a hierarchical representation higher level features are derived from lower level features.
4) The broader machine learning field of learning representations of data is a part of Deep learning algorithm.
5) That corresponds to different levels of abstraction to learn multiple levels of representations and also the levels which is form a hierarchy of concepts.
6) Use some form of gradient descent for training via back-propagation.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

These definitions have in common: (1) multiple layers of nonlinear processing units (2) the supervised or unsupervised learning of feature representations in each layer, with the layers forming a hierarchy from low-level to high-level features.

Thedeep learning algorithm which is depends on the solution of the problems used in the composition of a layer of non-linear processing units. We propose a multi-layer deep learning algorithm to inference the diseases from symptoms.

Rare diseases are chronic diseases. Some of them are curable if detected in early stages. Rare disease symptoms are very common symptoms that can be ignored by the patients. In this system, we propose symptoms based disease discovery system which discovers diseases on the basis of deep learning algorithm. Deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks.

The organization of this paper is as follows: Section II provides proposed methodology. Finally, Section III concludes the paper.

II. RELATED WORK

In [2], Multi-Task Risk Prediction via Joint Co-morbidity Discovery (Multi-Risk) is discussed. In that system, the author proposed a multi-task risk prediction algorithm in which the symptoms specified by the user will be checked with multiple diseases and along with matching diseases, the diseases which are relevant(inter related) with the matching diseases will be shown to user. This paper only discussed about well-known diseases.

In [3], Regional Disease Distribution Based On Social Computing is discussed, In this system author focuses on the diseases may occur due to geographical conditions like temperature, quality of water, available vegetables etc. In this system, the medical dataset is considered and a decision support system is developed to spread awareness about the place wise diseases possibilities. Only geographical diseases are discussed but the information given is globalized. Some people's body can resist diseases infections germs depending on their immunity power. It can be dependent on some factors such as age, BMI, existing diseases etc.

In [4], Genetic disease discovery system is discussed, the experiment was held on the dataset of patients suffering from "Retinitis Pigmentosa". There is no recommendation system developed for live data.

In [5], why rare diseases are an important medical and social issue is discussed. In this paper, the author discussed about the awareness about the rare diseases should spread so that the human beings will be alert about the diseases and if the diseases are diagnosed in early stage, there is a possibility of cure. No solution is provided.

In [6], the author shows that the mining relationships among genetic diseases and among genes using vertex similarity measures can be an inexpensive and promising indicator for potential gene-disease relationship discovery. The research is very hard to implement and can be used for deep study in medical system.

In [7], this tool is designed to identify rare diseases on the basis of a patient's symptoms. To create it, several software entities were designed and integrated. A database of symptoms associated with every human rare disease known was designed and implemented. This database was derived from information available from Orphanet.

In [8], Orphanet provides gold-standard data regarding rare diseases in the world. A user-friendly website was also designed, implemented and connected to the database.

III. PROPOSED METHODOLOGY

This section provides the development of proposed system. This chapter describes the detailed information about the development of the proposed approach. These are further discussed in brief.

In existing system, BN algorithm is used to classify symptoms into rare diseases categories. The aim was to discover possible diseases on the basis of specified symptoms. But the relevant tests/preventive measures and remedies are not given. Some timesnon technical user will not be able to specify symptoms correctly in that case the existing system will not work properly.

The functionality of RDD is summarized as three modules are as follows:

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

*1)Disease Prediction:*The first RDD functionality is the prediction of a disease by entering the symptoms in the main panel. The symptom input field also has an auto-complete function. This function helps the user enter the symptom in the format expected by the application. Symptom input is always acknowledged by a message at the top of the application. Once the user has selected the symptoms, clicking the Submit Symptoms button begins communication with the server.

*2) List of diseases in the database:*The second functionality of RDD is the list of diseases in the database, together with their associated symptoms. This page appears on clicking the Disease List button. Then, one must select the first character of the name of the disease. Once a character is selected, the list of diseases with names starting with that character is shown. By default, the symptoms associated with a disease are hidden, and each disease can be clicked on to reveal them, making navigation more streamlined.

*3) List of symptoms the database:*The third functionality of RDD provides the list of symptoms in the database, to get her with their associated diseases. This page appears on clicking the Symptom List button (Figure 3). For each symptom, there is also a button that allows adding the symptom to the list of other symptoms that will be used for rare disease prediction by RDD. RDD was built to be modular. This makes the prototype highly flexible, allowing for easy updates to the disease and symptom databases. In addition, the modularity also facilitates future addition, replacement, and/or removal of alternative prediction engines without restarting the application.

In this paper, we propose a symptoms /questions/query based disease discovery system which classifies the symptoms specified by the user into most relevant diseases. We will include common as well as rare diseases in our dataset. We provide the easiest way to specify symptoms using language translator and symptoms suggestions service. We are going to use Deep learning algorithm (which is based on artificial neural network) to classify the symptoms into appropriate diseases. We also provide more detailed symptoms to user about the detected diseases so that the user will be able to verify that he is suffering from that disease or not. We also provide the preventive measures/ remedies and medicine stores(including Google map details) where the medicines are available. In case of urgency, patient will place an order to purchase the medicine. As compared to the literature, we are providing the diseases discovery as well as solutions over the diseases.

To improve efficiency and reduce time is the main objective of the system. This scheme builds a novel deep learning model. In this system, first mines the latent medical signatures from the health related reviews which are compact patterns of interdependent medical terminology. In one layer raw features and signatures serves as input nodes. And in the subsequent layers hidden nodes are serves. Later using pre-training the system learns about inter-relations between the two layers. For more abstract signature mining the hidden nodes are viewed as raw features. This scheme builds a sparsely connected deep learning architecture with three hidden layers. The sparse connection between two adjacent layers makes it faster. Our model is automatically determined the number of hidden nodes in each layer. Deep learning consist of two component given below:

1)Keyword Extraction

- Specify symptoms
- Tokenize symptoms
- Remove stop words
- Stemming
- Apply NLP
- Take nouns as keywords
- Submit keywords to deep learning system

2)Disease Detection via Sparsely Connected Deep Learning

The deep learning model has L layers which is sparsely connected with d1 ($1 \leq 1 \leq L$) nodes in each layer.

In the first layers the input n-dimension raw features are contained. The output disease types denoted in the L-th layers. The last layer is intermediate layers which are hidden layers It is unseen from the data.

In this work, nodes at higher layercontains the signatures and are connected to the adjacent lower layer nodes.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

The hidden layers are alternating between sub graph mining and pre-training which are initially regarded the learning model with only one hidden layer.

In hidden layer each node is corresponding to a signature which is obtained via dense sub graph mining from a large graph.

Fig. 1 describes the Deep learning model using Multi-layers. Initially in Fig. 1there are three layers containing Input layer, Hidden layer and output layer. In input layer raw features are contained. Hidden layers shows the possible signature regarding to the input layer. The hidden layer is unseen from the data so that the user only see the result from the hidden layers. The last layer is output layer which is also called as intermediate layer. The output layer shows the predicated result from the input layer and from hidden layer.

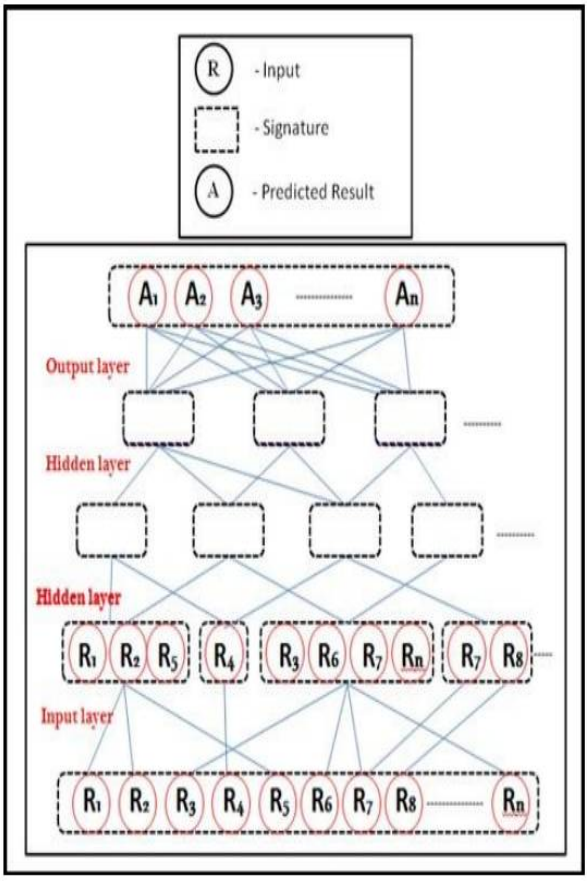

Fig. 1 Deep Learning Model using Multi-layers

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

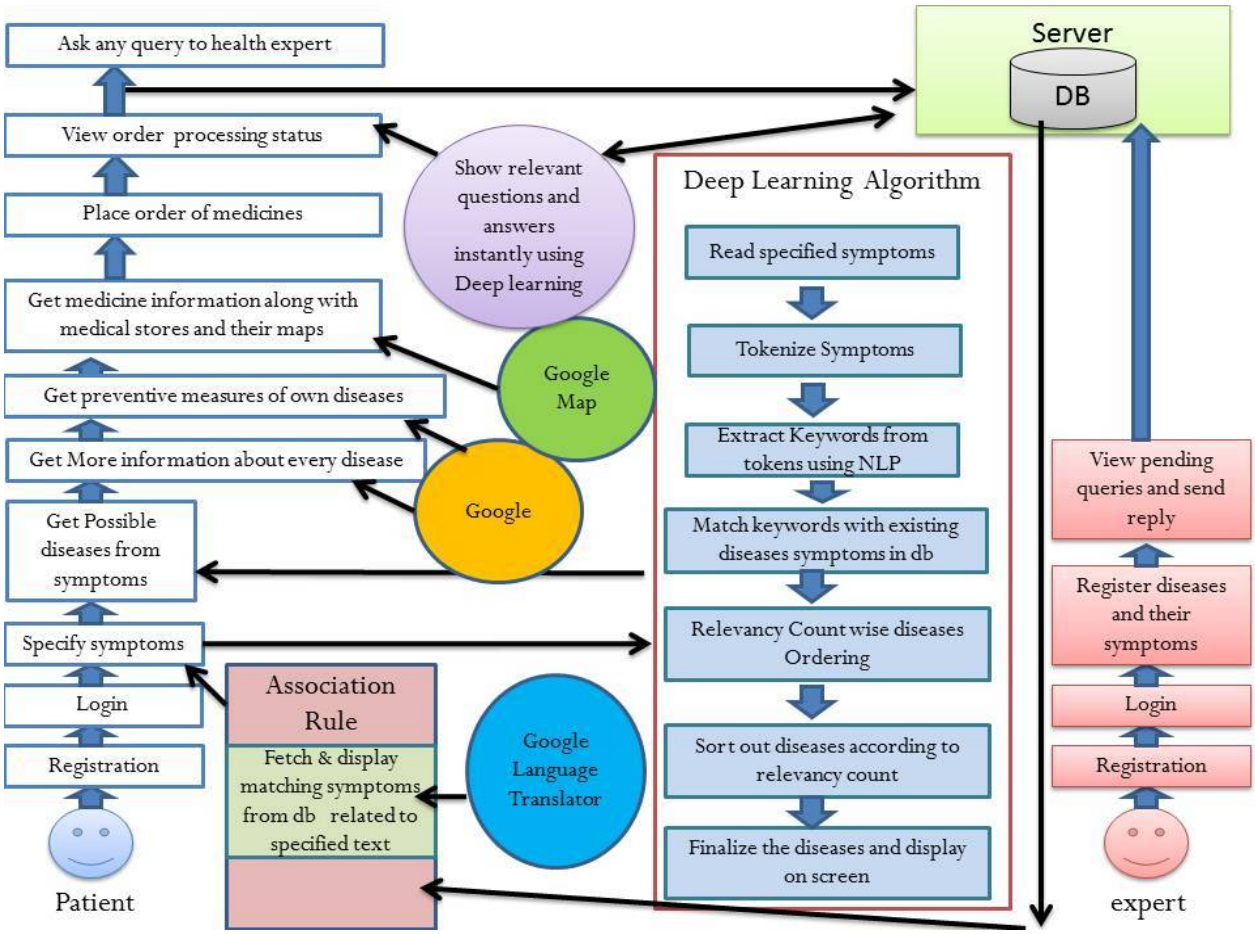

Fig. 2 Flowchart of Proposed System

3) Modules

a) Admin panel

- Login/logout
- Register disease and details
- View users list
- View user search details

b) Health Expert

- Registration
- Login
- Post any symptoms/query
- Get symptoms/query wise probable diseases information
- View search history of symptoms/query

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

c) End user

- Login/logout
- Patient communication
- View patient pending queries
- Send answer to queries
- Inbox
- Preventive measures registration
- Outbox

d) Medicine management

- Chemist will maintain track of medicines stock
- System will suggest patients about predicted disease relevant medicines and stores location with Google map where the medicine is available
- Patient will place order of any medicine if needed
- Chemist will process the pending orders
- Patients & chemist will get notifications accordingly

e) Deep Learning

- Read query as a input
- Query tokenization
- Filter Medical signatures from tokens
- Fetch matching signature wise probable diseases
- Arrange signature matching count in descending order
- Filter diseases having low matching counts
- Show diseases information as a output

f) Questions-Answers System

- User will send any question to health expert
- System will process that questions using deep learning algorithm and deliver most relevant answers from existing dataset

Fig. 2 describes the Flowchart of Proposed System. Initially in Fig. 2 there are three modules Admin panel, Health Expert and End User. For login the End User i.e. Patienthave to registered on the system. To fetching and displaying the matching symptoms from database Association Rule is used. After that to get possible diseases from the symptoms user have to specify the symptoms. Then the user will get to know the preventive measures and medicines of the predicted disease so that he can place the order of medicines and ask queries regarding the disease to the health expert from the database. The database shows the relevant questions and answers instantly using Deep learning.

Further in Fig. 2, the Health Expert also have to logged in on the system. After performing login, Expert will registered diseases and their symptoms. And also send reply to the pending queries of the patients from the database.

In Fig. 2, the Deep Learning Algorithm plays an important role for specification of the symptoms. First, entering the symptoms on the system the Deep Learning Algorithm read the specified symptoms and tokenize the symptoms and extract keywords from tokens using NLP (Natural Language processor). After that using Deep learning the system matches all the keywords with existing diseases symptoms in the database and also ordering the diseases by the relevancy count. After ordering the diseases sort out all the diseases by relevancy count. Sorting all the diseases the only best possible diseases will finalize and then displays on the screen to the patient.

The Admin will give the detailed information about the registered diseases. The Admin shows the whole users list on the system so that he can search all the details about the users. Therefore, the whole proposed system will be applicable on the End User, Expert and Admin Panel.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

4) Algorithms

a) Deep Learning Algorithm: This uses a cascade of many layers of nonlinear processing units for feature extraction and transformation. This is also use in some form of gradient descent for training.

b) Associated Rule mining: For discovering relations between variables association rule mining is used in large database.

IV.CONCLUSIONS

In our proposed system we propose the aim is to improve disease discovery system. We propose symptoms based disease discovery system which discovers diseases on the basis of Deep learning algorithm. To find the most relevant diseases will be shown to the patients along with detailed information on each disease. To get the medicine patients could place medicine order online to achieve anonymity.

REFERENCES

1. Marc Pinol, RuiAlves, Ivan Teixido, Jordi Mateo, FrancescSolsona, Ester Vilaprinyo, "Rare Disease Discovery: an optimized disease ranking system", IEEE Transactions on Industrial Informatics, 2017.
2. Xiang Wang, Fei Wang, Jianying Hu IBM T. J. Watson Research Center, Yorktown Heights, NY. "A Multi-Task Learning Framework for Joint Disease Risk Prediction and Comorbidity Discovery". 2014 22nd International Conference on Pattern Recognition
3. SHEN Peng-Fei, ZUO Wan-Li. "Discovery Of The Regional Disease Distribution Based On Social Computing". 2015 7th International Conference on Information Technology in Medicine and Education.
4. Chunxiao Xing, Chunxiao Xing. "Disease Gene Discovery of Single-gene Disorders Based on Complex Network". 2016 IEEE/ACM 3rd International Conference on Big Data Computing, Applications and Technologies.
5. ArrigoSchieppati, Jan-IngeHenter, Erica Daina, Anita Aperia." Why rare diseases are an important medical and social issue". www.thelancet.com Vol. 371 June 14, 2008.
6. AbolfazlDoostparastTorshizi, Linda Petzold." Sparse Pathway-Induced Dynamic Network Biomarker Discovery for Early Warning Signal Detection in Complex Diseases" Journal Of Latex Class Files, Vol. 14, No. 8, August 2015.
7. Maiella S, Rath A, Angin C, Mousson F, Kremp O. "Orphanet and its consortium: where to find expert-validated information on rare diseases." Science Direct 2013.
8. Rath A, Olry A, Dhombres F, Brant MM, Urbero B, Ayme S. "Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users." HUMAN MUTATION, Vol. 33, No. 5, 803–808, 2012 2012.
9. Curilem M Et Al. Pattern recognition applied to seismic signals of the Llaima volcano (Chile): An analysis of the events' features. Journal of Volcanology and Geothermal Research 282: 134–147. 2014.
10. Dragusin R, Petcu P, Lioma C, Larsen B, Jørgensen Hl, Cox Ij, Hansen Lk, Ingwersen P, Winther O. FindZebra: a search engine for rare diseases. International Journal of Medical Informat- ics 82:528-538. 2013.
11. Prashanth R, Roy Sd, MandalPk, Ghosh S. Automatic classificationandpredictionmodelsforearlyParkinson'sdiseasediagnosis from SPECT imaging. Expert Syst. Appl. 41(7), 3333-3342. 2014.
12. Yadav G, Kumar Y, Sahoo G. Predication of Parkinson's disease using data mining methods: A comparative analysis of tree, statistical, and support vector machine classifiers. Indian Journal of Medical Sciences 65(6): 231-242. 2011.
13. Mumford Ja, Ramsey Jd. Bayesian networks for fMRI: a primer. Neuro image 86: 573–582. 2014.
14. Morales Da Et Al. Predicting dementia development in Parkin- son's disease using Bayesian network classifiers. Psychiatry Research: Neuroimaging 213(2):92-98. 2013.
15. EomJh, Kim Sc, Zhang Bt. Aptacdss-E: A classifier ensemble- based clinical decision support system for cardiovascular disease level prediction. Expert Systems with Applications 34 (4) 2465-2479. 2008.
16. EadieLh, Taylor P, Gibson Ap. Recommendations for research design and reporting in computer-assisted diagnosis to facilitate meta-analysis. J Biomed Inform 45: 390–397. 2012.