International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

# A Review on Disease Inference Techniques

**Rasla Azeez[1], Anu Prabhakar[2]**

[1]KMCT College of Engineering, Calicut, Kerala, India
[2]KMCT College of Engineering, Calicut, Kerala, India

**Abstract:** ***An important problem of current Web search is that search queries are usually short and not enough for knowledge inferring, and thus are not good enough for specifying the precise user need. Automatic disease inference is of importance to reduce the distance between knowledge to be inferred and needs of the user. This knowledge mining is useful especially for community-based health services due to the vocabulary gap, incomplete information, correlated medical concepts, and limited high quality training samples. There are many online and offline methods are there to getting the information requested by the health seeker. In this paper, here first perform a health seeker study on the information needs of health seekers in terms of queries. Methods like questionnaire , deep learning are used as inferring methods .Here the sparse deep learning algorithm is used as the data mining technique .Some attributes used are raw features ,signatures ,medical attributes etc. Deep learning refers to the depth wise analysis of the raw features and their signatures as input nodes in one layer and hidden nodes in the next layers of the learning architecture. That is , it learns the internal relations between the data collected and several layers. Abstract signature mining will produces the deepest knowledge about enquiry process towards health seeking. Because of the deepest and alternative repeating of these features, our architecture a sparsely connected deep learning technique maintained with three hidden layers. It will give out specific tasks with fine-tuning, pre-tuning etc. Several experiments on day today dataset given by online doctors show the significant performance and importance of this disease inference.***

**Keywords:** SVM (Support Vector Machine), Sparse deep learning, Classifiers, Querying, Signature mining.

## 1. Introduction

The increasing in the development of each and every countries, expenditure of healthcare and emergence in computer technologies all are the major reasons for the innovation to the automatic health seeking system. Analyzing a survey it is easy to understand that most of the people are utilizes the availability of emerging technologies such as computers, journals, magazines, internet technologies etc. The Knowledge mining in health records is a key aspect for improved clinical decision making, and patient management, population management etc. Increased internet technologies in health seeking promises towards better data aggregation, automated requesting and updating, and clinical development etc. The large number of information they capture over time pose challenges not only for medical practitioners, but also for the information analysis by machines and local users they are intended to get satisfied with their need.

The main aim of this paper is to line up and emphasize the importance of exploratory analytics that are commensurate with human capabilities and constraints along with updates of existing exploratory data. In this realm we present a sparse deep learning architecture that discovers complex medical attributes, medical signatures and its corresponding matching patterns, which are easily understandable by users. Here enhance this framework based on users need and available prior data. The analysis study pointing towards several aspects of users need.

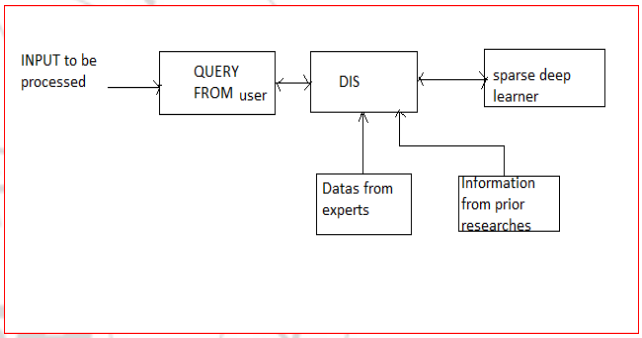


**Figure 1:** Disease Inferring Architecture

The above disease inferring system(DIS),illustrating about how an user or health seeker can be meet his health requirements.WebMD1 , health tap , MedlinePlus2 are the typical examples of health data providers. Community-based health services, such as HealthTap3 and HaoDF4 are examples of community based seeking. There are some interactive platforms for the seekers. They can anonymously ask health-oriented questions and at that time doctors provide the knowledgeable and trustworthy answers. Typical example can be explained as user can ask question regarding about their health need towards the system. Community based health service is very time consuming for health seekers to get their posted questions resolved in a particular period of time. The resolving time could vary from hours to days some may be more than these. Next limitation is that doctors having to cope with an ever-expanding workload, which leads to reduction in efficiency. While considering any other online methods we can see that doctors are compelled to finish replies for number of problems that are requested by users because all it is an important factor to identify solution for the above so here develop automatic and comprehensive wellness systems that can instantly answer all-round questions of health seekers and alleviate the doctors‟ workload. Along with the application of a bidirectional querying will improve the performance of the system in a greater amount. The

Support Vector Machine (SVM) is the classifier used to classify the data in a sparse deep learning manner.

Support Vector Machine is a best classifier and we used it here, Support vector machines (SVMs) are a well- known supervised learning technique for performing binary classification in data classification. They are very accurate and generalize well to a wide range of operations. To understand this just considers an SVM used by a bank to determine to whom they will loan money. If a user‟s loan application is rejected and they would like to getting the information about why it is rejected , here it is not very useful to only be able to say that the algorithm came back with a number lower than some required threshold. There will be notification, and that would be much more satisfactory to the customer to be able to tell them that they were rejection of credit because their income is too low and they have six outstanding loans. The Decision trees can provide a structure that is much more easily interpreted, but unable to interpret always to achieve results as accurate as those produced by an SVM. Since support vector machines are "black-box" classifiers, the decisions they make are not always easily explainable. Because of this we mean that the model produced does not automatically provide any helpful intuitive reasons about why a particular point is classified into a class rather than another classifier.

## 2. Literature Survey

Because of the growing aged population coupled with lack of medical services and healthcare services in most of the developing countries, the traditional health- care system meets challenging problems caused by its high operating cost and un scalability. Compared to the conventional healthcare system, there is a need of more accurate and easy to access system to improve quality of health care services. The following are the some of the parameters we needed to be improved for the health care services. They are improving the quality of medical service, Improving the utilization of medical helps and care by enabling remote medical services, and supporting the development of the health industry. The existing system mainly focuses on healthcare service in a physiological and psychological aspect with the following two undesirable features etc. The existing literatures [9],[5] and [8],[6] are explaining some discriminant features. Where the basic methodology used in this advanced health care is explained by support Vector Machine classifier[1].The literatures[3],[4],[2],[7] following some earlier methods of health services.

David Barbella [1] proposed a system the where SVM or the Support vector machines are a valuable and useful tool for making classifications. But their black-box nature means that they lack the natural explanatory value that many other classifiers possess. In the first, we report the support vectors most touching in the final classification for a particular test location. Next we determine which features of that test location would need to be changed in order to be placed on the separating surface between the two classifications. Another technique is called "border classification." In order to explaining these explanatory techniques, we also present a free-for-download software tool that enables users to visualize these insights graphically. In other way, many useful and great famous websites have shown recent success in explaining recommendations based on nature of other users. Accepting by these ideas, we recommend two novel methods for providing insight into local classifications produced by a SVM.

Here introduces two new techniques for explaining support vector machines on continuous data. Both techniques explain the model on the local level, i.e. for an individual test point, as a recommender system might. One involves finding the support vectors that make the strongest contribution to the classification of a particular test point. The next technique is similar to inverse classification technique: the aim is to find a relatively minimal change in order to switch the classification of a test point. Anyway instead of minimally switching the classification, we propose finding the locally minimal change required to move the point to the different surface of the classes. Here these techniques add wide details to the results of an SVM classifier in a format with which users of online recommendation systems are quite familiar. We have showed a software tool named SVM-zen that allows users to see these explanations graphically. A SVM requester can look at a particular test location and determine that the test location was classified in that class due to a specific group of highly weighted support vectors that is, this classification is based on the classification of a point of particular similarity threshold with another class.

F. Wang [2] proposed a temporal knowledge representation and learning framework to perform large scale temporal signature mining of longitudinal heterogeneous event data occurrences .We present a doubly constrained convolutional sparse coding architecture that learns interpretable and shift-invariant latent temporal event signatures. Novel stochastic optimization architecture performs large-scale incremental learning of group-specific temporal event signatures. It evaluates the framework on synthetic data and on an electronic health record dataset and its manipulation.

This architecture enables the representation, extraction, and mining of high- order latent event structure and relationships within single and multiple event sequences. This data representation point s the heterogeneous event sequences to a geometric image by encoding events as a structured spatial-temporal shape process. It empirically showing that stochastic optimization scheme converges to a fixed point and we have demonstrated that our framework can learn the latent event patterns within a group. Future work will be developed to a thorough clinical assessment for visual interactive knowledge discovery in large electronic health record databases for user‟s need.

N. Lee [3] established the knowledge discovery in electronic health records (EHRs) as a central aspect for improved clinical decision making, prognosis, health data management and patient management. Where EHRs show great promise towards better data integration, automated access, and clinical workflow improvement, the detailed information they collect over time face challenges not only for medical practitioners, but also for the information analysis by machines.

The aim of this is to motivate the importance of exploratory analytics that are commensurate with human capabilities and constraints to be meet. Here this architecture on synthetic data and on EHRs together with an extensive validation involving many computed latent factor models. The current study is the first to link temporal patterns of healthcare resource utilization (HRU) against a diabetic disease complications severity index to better understand the relationships between disease severity and care delivery that will useful for further motivations. While using this realm we present a novel temporal event matrix representation and learning architecture that discovers complex latent event patterns, which are easily interpretable by human beings.

In Amitpande[4] SSIEEE, used built-in smartphone sensors (accelerometer and barometer sensor), sampled at low frequency, to accurately estimate Energy Expenditure. Here also using a barometer sensor, in addition to an accelerometer sensor, greatly increases the accuracy of Energy Expenditure estimation. The Energy expenditure (EE) estimation is an important parameter in chasing personal activity and stopping chronic diseases, such as obesity and diabetes. Maximum correct and timely EE estimation utilizing small wearable sensors is a difficult task, firstly because of the most existing schemes work offline or use experience.

In this, pointing on accurate EE estimation for chasing ambulatory routines (walking, standing, climbing upstairs, or downstairs) of a typical smartphone user. Considering bagged regression trees, a machine learning technique, here enhanced a generic regression model for EE estimation that yields upto 96% correlation with actual Energy Expenditure. Here compare our results against the state-of-the-art calorie measuring meter equations and consumer electronics devices (Fitbit and Nike+ Fuel Band are considered). The latest developed EE estimation algorithm demonstrated superior accuracy compared with currently available methods.

Lejun Gong [5] proposed a system where latest disease holding genes could be detected. Understanding the hand of genetics in diseases is one of the most important and greedy tasks in the post genome era. Genetic association analysis and diversions has proven to be a successful tool to enhance the knowledge about genetic risk components to a variety of complex diseases. Measuring the functional similarity between known disease susceptibility genes and unknown genes is to predict new disease susceptibility genes.

There are wide applications of computational methods in dis-covering gene responsible for human disease. Here propose an approach to prioritize disease susceptibility genes using LSM/SVD. Measuring the functional similarity between known disease susceptibility genes and unknown genes is to predict new disease susceptibility genes. It could discover again latest disease holding genes. This new approach of disease gene prioritization could discover new disease affecting genes.

M.Shouman[6] established the availability of large amounts of medical data that leads to the need for powerful data analysis tools to extract useful knowledge for finding a particular need towards health. While using single data mining technique in the diagnosis of a disease especially in heart, has been comprehensively investigated showing acceptable levels of correctness. Here as a result, researchers have been investigating the result of hybridizing more than one technique showing enhanced results in the diagnosis of a disease (heart disease). Health researchers have long been considering with applying statistical and data mining tools to enhance data analysis on large data sets of health data.

It is done by the motivation by the world-wide increasing mortality of heart disease patients each year and the availability of huge amounts of data, researchers are using data mining techniques in the diagnosis of disease especially heart disease. Also applying data mining techniques to help health care professionals in the diagnosis of heart disease is having some success, the use of data mining techniques to identify a suitable treatment for heart disease patients has received less attention, these all lead to improve the health care techniques.

Y. Bengio[7] proposed a representation learning that evaluate the learning process in a better way. The performance of machine learning techniques is heavily dependent on the choice of data representation (or features) on which they are applied. To expand the scope and ease of applicability of machine learning, it would be highly desirable to make learning algorithms less dependent on feature engineering so that novel applications could be constructed faster, and more importantly, to make progress toward artificial intelligence (AI.)Because of that reason, much of the actual effort in deploying machine learning algorithms goes into the design of pre-processing pipelines and data transformations that result in a representation of the data that can support effective machine learning techniques. This is important but labour intensive and highlights the weakness of current learning techniques .Its disability to extract and organize the discriminative information from the data. Feature engineering is a way to take advantage of human unawareness and prior knowledge to compensate for that disability. An Artificial Intelligence must fundamentally understand the world around us, and we go straight forward that this can only be achieved if it can learn to identify and disentangle the underlying explanatory factors hidden in the observed of low-level sensory data for any learning scheme.

In Olumurejiwa[8], the Medical equipment in developing world health facilities is largely insufficient and inappropriate, resulting in reduced capacity for hospitals and clinics, suboptimal outcomes for patients, and wasted money for donors and investors. The equipment donation is currently the largest source of equipment for developing world health needs and caring facilities. While the majority of these health instruments fail within 6 months of arrival, with others never being usable, this will be the status of the available instruments. So the result would be corrupted.

The current situation in developing countries is a one way flow of technology down the wealth and development gradient also their major concern is about health caring. Promoting the use of appropriate technology goes beyond adjusting the criteria for equipment procurement to building local design, manufacturing, and management capabilities on

the ground. . Appropriateness encapsulates „„effectiveness, safety, the ability of the community to pay for it, and the availability of expertise to utilize and maintain the technology"". An important question is whether „„high-tech"" equipment even has a place, or is appropriate, in low-resource set- tings. The MTS needs to aid national health administrators in this last task by giving a single measure of adherence to national standards and progress in capacity formation.

In Gerberding[9] public health is an important factor for the wellness of a country, so public health surveillance is the on-going, systematic collection, analysis, interpretation, and dissemination of data about a health-related event for use in public health action to reduce morbidity and mortality and to improve health status. The foundation of communicable disease inference in the United States is the state and local application of the reportable disease inference system known as the National Fortifiable Disease Surveillance System (NNDSS), which holds the listing of diseases and laboratory findings of public health interest, the publication of case definitions for their surveillance, and a system for passing case reports from local to state to CDC and that will perform an efficient health wellness system.

The valuation should be shortened to convey the strengths and weaknesses of the system under scrutiny. Shortening and reporting evaluation findings should facilitate the comparison of systems for those making decisions about new or existing surveillance methods. An Institute of Medicine study concluded that although innovative surveillance methods might be increasingly useful in the detection and monitoring of outbreaks, a balance is needed between strengthening proven approaches (e.g., diagnosis of infectious illness and strengthening the difference between clinical facilities providers and health controllers) and the exploration and evaluation of new methods.

Min Chen [10] proposed a system called AIWAC to reduce the heavy burden from rapidly growing demands of healthcare service. A wearable computing-assisted healthcare has been proposed for health monitoring and remote medical care services. Lots of existing healthcare systems are targeted at caring for elderly people"s physiological status rather than psychological status. So, without efficient mechanisms of affective interaction, the traditional wearable technology is not adequate to provide advanced healthcare services involving both physical and emotional care that is especially in psychological care. This psychological aspect becomes more and more important to improve seniors" quality of life and health status maintenance.

In AIWAC it supports Hybrid emotional data analysis, which supports computation-intensive analysis of various emotional data from CPS-Spaces, Dynamic resource perception and allocation by which user"s status can be analyzed, which provides users with real-time, available, and effective and affective interaction. At last an AIWAC test bed is used for emotion-aware application, based on a robot has been presented. Using all this architecture-level design, we will investigate how to provide mobile users with resource-intensive and emotion-aware services while achieving a user friendly trade-off between communication and computation as future work.

Liqiang Nie[11] established a system in which disease is inferred from Health-Related Questions via Sparse Deep Learning technique is an efficient technique to identify diseases, or to monitor the health status. With incorporating the bidirectional querying technique it is possible to getting the real time health status as correct as possible. As mentioned earlier, vocabulary gap, incomplete information, inter-dependent medical attributes and limited ground truth have greatly hindered the performance of classic shallow machine learning approaches. There are several techniques for the primary analysis of a disease; they are online health data providers such as WebMD1, MedlinePlus2. The others are online talking schemes with doctors such as HealthTap3 and HaoDF4.While reading health magazines and journals it is also possible to getting more information. Online doctors offer interactive platforms, where health seekers can anonymously ask health-related questions while doctors provide the knowledgeable and trustworthy replies.

Here collected more than 900 popular disease concepts from EveryoneHealthy5, WebMD and Medline Plus. Also handled with a wide range of diseases, including endocrine, urinary, neurological and other aspects. Using these disease concepts as queries, we crawled more than 220 thousand community generated QA pairs from Health Tap. In order to increase the effectiveness of our proposed disease inference scheme, we compare it against three state-of-the-art techniques. Most of them can benefit from labeled data; unlabeled data supervised and unsupervised data, which ensures fair comparison. This technique mainly focused on sparse deep learning technique where each layer is incrementally added based on the user"s need. SVM is implemented here as a classifying tool. Overall it will give a better performance in inferring a disease.

## References

[1] David Barbella1, Sami Benzaid2, Janara Christensen3,Bret Jackson4, X. Victor Qin "Understanding Support Vector Machine Classifications via a Recommender System-Like Approach"
[2] F. Wang, N. Lee, J. Hu, J. Sun, S. Ebadollah , and A. Laine, "A framework for mining signatures from event sequences and its applications in healthcare data," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013.
[3] F. Wang, N. Lee, J. Hu, J. Sun, and S. Ebadollahi, "Towards heterogeneous temporal clinical event pattern discovery: A convolutional approach," in The ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2012.
[4] Amitpande, JindanZhu1, Aveek K. Das1, Yunze Zeng1, Prasanth Mohapatra1, (Fellow, IEEE), And Jay J. Han " Using Smartphone Sensors for Improving Energy Expenditure Estimation"
[5] Lejun Gong∗, Ronggen Yang, Qin Yan, and Xiao Sun, " Prioritization of Disease Susceptibility Genes Using LSM/SVD"

[6] M.Shouman, T. Turner, and R. Stocker, "Using decision tree for diagnosing heart disease patients," in Proceedings of the Australasian Data Mining Conference, 2011.
[7] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013.
[8] Refinement of the Facility-Level Medical Technology Score to Reflect Key Disease Response Capacity and Personnel Availability, Olumurejiwa A. Fatunde (Student Member, IEEE)1, And Timothy W. KOTIN (Student Member, IEEE).
[9] Gerberding, M.D., M.P.H."Centers for Disease Control and Prevention" Julie L.
[10] AIWAC: Affective Interaction Through Wearable Computing And Cloud Technology, Min Chen, Yin Zhang, Yong Li, Mohammad Mehedi Hassan, And Atif Alamri
[11] Liqiang Nie, Meng Wang, Luming Zhang, Shuicheng Yan, Member, IEEE, Bo Zhang, Senior Member, IEEE, Tat-Seng Chua, Senior Member, IEEE "Disease Inference from Health-Related Questions via Sparse Deep Learning".

## Author Profile

**Rasla Azeez** is pursuing her M.Tech degree in Computer Science and Engineering from KMCT College of Engineering, Calicut University. She obtained her B.Tech Degree in Computer Science and Engineering from KMCT College of Engineering, in 2013.