Disease Predictive Modelling using Machine Learning in HealthCare : A Systematic Mapping Study

Samyak Jain[1], Mishal Shah[2]
Department of Computer Science and Engineering,
National Institute of Technology Karnataka Surathkal,
Mangalore, Karnataka.
[1]16co254.samyak@nitk.edu.in
[2]shahmishal1998@gmail.com

***Abstract*— The paper illustrates the approach, methodology and results from a Systematic Mapping Study (SMS) for Disease Predictive Modelling using Machine Learning for Healthcare Industry. The research comprised of studying and analyzing 50 Studies in the Healthcare Industry. The various Predictive Modelling techniques that are currently prevalent in industry and academia/ research are analyzed and presented. Also, methods to improve accuracy of Prediction have been discussed, also the factors that impact the prediction modelling and its accuracy are highlighted. Our research indicates that there are varying opinions and that there isnt́ a single classifier (model) that comes across as a best-in-class process for disease prediction. Various Studies have offered varying techniques and performance results, but all have concluded that Doctor opinion and analysis cannot be replaced by an algorithm or a set of algorithms and that learning systems need continuous trained or reinforced for better accuracy.**

***Index Terms*— Prediction Modelling, Machine Learning, Prediction Accuracy, Systematic Mapping Study (SMS)**

I. INTRODUCTION

Predicting Disease/ epidemic outbreaks is an intensely complex task that involves experts from various disciplines including healthcare agencies, Governments, Environmental agencies and Information Technologists. With increased digitization and governments focus on creating electronic records from cradle to grave for the population, it is now possible to apply predictive algorithms (Bayes Classifier/ Rule based) on real time data to predict disease at an individual or a mass population level. However, predicting the accuracy and therefore, the correctness and accuracy of the prediction is still circumspect.

The study strives to examine and review the literature around accuracy of Disease Prediction Modelling using Machine Learning. This could provide a baseline foundation for further analysis and research in this exciting but complex area.

- Section II defines the Approach/ methodology of research used for this study.
- Section III details the analysis of results.
- Section IV enumerates the research conclusions

II. SMS RESEARCH APPROACH/ METHODOLOGY

The SMS research process is based on Guidelines for performing Systematic Literature Reviews in Software Engineering Version 2.3, Technical Report (EBSE), EBSE-2007-01 as defined by Kitchenham and Charters [1]. The team of researchers, students of NITK, 4th Semester (Mishal Shah and Samyak Jain) have undertaken the features as mentioned in Fig.1 of SMS for conducting the study.

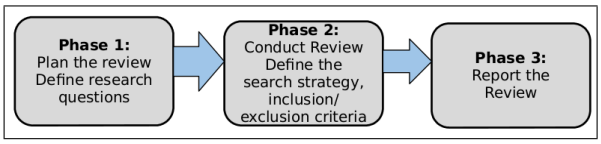

Fig. 1. Phases and steps in Systematic Mapping Study

A. Phase 1: Plan the SMS Study:

The framework used comprises of identifying the Population, defining the Intervention, identifying conventional Comparisons, expected Outcomes and the Context. The framework known as PICOC has been used to help frame the research questions.

- **Population:** Healthcare Software.
- **Intervention:** - The intervention used for Research was 'Machine Learning/ Artificial Intelligence'. The intervention basically is the software methodology/tool/technology/procedure that addresses a specific issue.
- **Comparison:** - The conventional comparison technology is Statistical Mathematical tools.
- **Outcomes:** - Outcomes relate to parameters that are important to practitioners e.g. improved accuracy, benefits, advantages, disadvantages, issues, problems etc.
- **Context:** - Small scale Academic Research by a group of NITK Students and performed at a small scale level.

After adequate deliberation, the following two research queries/ questions (RQx) were identified to be answered by the study:

- **RQ1:-** What are the techniques used for Prediction Modelling using Machine Learning for Healthcare Industry?
- **RQ2:-** What factors influence the accuracy of modelling for predictions in Healthcare Industry?

B. Phase 2 - Conduct the SMS Study, Search Strategy

Firstly, the two research questions (RQ1 and RQ2) were drafted and an exhaustive search was conducted as per the Search strategy outlined below. The Search Strategy keywords are enumerated in the Table I as defined below. The time period covered was 2000 to 2018 (inclusive).

TABLE I
SEARCH STRATEGY

RQ1	RQ2
(prediction OR prediction modelling AND (Healthcare OR Life Sciences OR Health OR Epidemic OR Medicine OR forecast OR machine OR learning OR supervised OR unsupervised learning)	(prediction modelling AND (advantage OR accuracy OR benefit OR success OR issues OR problems OR challenge OR disadvantage OR failure)

2.1 ***Primary Study Selection Criterion***

In order to improve the relevancy and comprehension of the chosen studies the exclusion criterion were as follows:

- Non-English studies,
- Very old studies (pre- 2000),
- studies relating to Statistical Modelling, Mathematical models,
- studies related to user interaction and modelling tools.

Study Selection Procedure: In accordance with Institute guidelines, only Institute approved Digital libraries were searched, this included IEEE, ACM, Elsevier and Springer. Table II depicts the Institution approved Digital library/Reference Material that was searched:

TABLE II
STUDY SELECTION CRITERION

Digital Library	Number of Final Studies
ACM	12
IEEE	17
Elsevier	19
Springer	2
Totals	50

2.2 ***Quality Verification/ Assessment*** A basic assessment of the identified Studies in terms of quality was done based on subjective judgement. As it is not possible to assess the author's credentials and background knowledge, the quality assessment was done purely based on the fact whether the research questions and allied subject matter have been deliberated upon by the chosen studies.

The Quality Assessment Checklist in Table III was used to further prune the list:

TABLE III
STUDY QUALITY ASSESSMENT

Number	Question?
1	How well does the Study align to research objectives?
2	How well defined is the approach and the analysis?
3	Are findings/ observations/ results stated in an unambiguous manner?
4	How clear are assumptions and/ or constraints?
5	How well have the depth and complexity of subject been articulated?

2.3 ***Study Data Retrieval and Synthesis***

The data from the studies were extracted using the properties as depicted in Table IV. The properties were identified and drawn from the studies and tables were populated to reflect on the research questions. At a based level, studies were classified as Industry or Non-Industry (i.e. academic setting or research environment).

TABLE IV
EXTRACTED DATA PROPERTY

Property	Research Question
What are Predictive Modelling techniques used in Healthcare Industry?	RQ1
What are the sources of Learning Data-sets for training the Computer?	RQ1
What are the Accuracy levels (if any) for defined Predictive Models in Healthcare Industry?	RQ2
What factors impact the Predictive Modelling Accuracy	RQ2

C. Report the SMS Study, Study report validation

The researchers identified two risks/ threats with respect to the review validity: Firstly, the Studies are mostly academic studies and report pilot results and hence the inferences need to be taken cautiously. Secondly, there is a lot of divergence in the way the authors have tried to solve the complex issue and therefore, there is a lot of potential inconsistency that may creep in as the observations / results are compared to draw the inferences.

3 Analysis of SMS Results

The SMS study comprised of studying 50 primary studies. The same were collected, studied and thoroughly analyzed as part of this research study. 72% of the studies were predominantly P&I (Population and Intervention) studies, while 22% were purely Intervention based and 6% studies were purely Population based.

3.1 RQ1: What are the techniques used for Prediction Modelling using Machine Learning for Healthcare Industry?

A total of 50 Studies were assessed - a vast majority of the studies (86%) were Academic/ Non-Industrial studies while 14% were Industrial studies. 60% of the studies had qualitative analysis while 40% had both qualitative and quantitative analysis. The observations from the data synthesized are as below:-

- 100% of the studies observed that Disease Prediction/prognosis is vital for medical system to provide best Healthcare services. However, there is no convergence on the most appropriate predictive modelling technique - Bayesian classification (30%) [5,10,23], Decision Tree(10%)[17], Regression model(10%)[12,17], Support Vector Machine[20,25,46], Neural Network(55%)[3,5,12,13,14,16,17,18,20,51], Ensemble model[12,13,37] etc.
- 100% of the studies observed that while the prediction techniques provide good performance (60%), accuracy is lower than while using a medical dataset. Also, the studies indicated that the learning data needs to be continuously upgraded for new symptoms and disease pattern. Also, 50% of the studies indicated that the same set of symptoms and laboratory data could actually be diagnosed as different diseases depending upon the environmental factors (season/ climate, external events etc.).
- 100% of the studies observed that Doctor Analysis cannot be replaced by Algorithmic prediction and that the procedure to be adopted should be that the algorithms predict the disease based on the patients symptoms and laboratory data before doctor analyzes the disease [2].
- 72% of the studies also indicated that highly accurate algorithms can be developed to predict the outbreak of a singular disease e.g. Swine flu or dengue or for critical illnesses like Cancer, heart disease etc. But the same algorithms cannot be extended to generic scenarios.
- 55% of the studies indicated that Neural Network is improving the observation capacity of information systems by training a limited number of neural networks nodes and collecting the results. Further improvement in accuracy of predictions is proposed through Fuzzy Hierarchical Approach [4]. Alternately, the output of multiple neural networks running in parallel feeding into a final rule based engine is also suggested[19].
- Finally, 35% studies indicated that composite algorithms suite or a hybrid Intelligent model using Hebbian Learning model, Naive Bayes Classification and back propagation further enhanced using kNN [smart pattern matching], Hopfield neural network and Self Organizing Maps(SOM) have provided better results [3].

The Table V summarized the techniques used in Prediction Modelling for Diseases.

TABLE V
PREDICTION TECHNIQUES

Technique	Description
Naive Bayes(30%)	Probability based Checking Bias towards diseases of High Probability
Decision Trees (20%)	Complexity Increase + Time Consuming
Clustering Method (10%)	Cluster Changes based on Symptoms, Unreliable results
Logistic Regression (15%)	Recursive, Time Consuming Process
Back-propagation (20%)	Time Complexity
Support Vector Machine (15%)	SVM Method works well in bits and Pieces for particular disease set
Random Forest (10%)	Multiple Decision Trees, Random Vectors, Voting for every tree with majority voting class returned (as Disease prediction)
Fuzzy Hierarchical Approach (30%)	Fuzzy/Mathematical Model-Fuzzy Systems
kNN, Hopfield Neural + SOM (25%)	K- Nearest Neighbors + Hopfield Network + Self Organise Maps to detect Diagonal Diseases (Weighted prediction of Symptoms)

3.2 *RQ2: What factors influence the accuracy of modelling for predictions for Healthcare Industry?*

The observations in terms of factors influencing the accuracy of prediction are:

- 100% of the studies observed that Medical Diagnosis is very complex and the same set of symptoms can lead to completely different diagnosis based on other parameters (patient's history, environmental factors etc.)
- 80% of the studies observed that quality and sufficiency of training data-set is the most important factor influencing the accuracy of Disease prediction. 60% of the studies used only one data set for validation hence the confidence levels are lower. Also, 100% of the studies explored common generic symptoms and diseases.
- 50% of the studies indicated that the Accuracy levels are currently between 50-60%. Accuracy levels are higher for predicting the outbreak of a disease (e.g. Swine flu, Dengue, Viral..) than generic diseases. Also, accuracy levels tend to be higher for specific diseases like lung cancer, heart disease etc. The Table VI below depicts the type of Charts demonstrated in 25% of the studies:
- 50% of the studies observed that Medical diagnosis is based on a large quantum of knowledge Basic Knowledge and relational knowledge [4] and has a large amount of uncertainty Diagnosis based on Fuzzy systems is purported in 25% of the studies[7]
- 40% of the studies indicated that weightages needed to be accorded to symptoms to improve accuracy and the same set of symptoms with varying weightages may change the disease prediction ranking.
- 30% of the studies observed that running multiple

classifiers improves the accuracy of prediction and that the classification improved with time. Multiple fold cross validation methods have been used to check accuracy and sensitivity[12,14,16,50].

- 25% of the Studies indicated that issues like a) Dealing with Missing data b) Noisy Data and c) reducing number of tests improved the quality of prediction. Also, 20% of the studies used noisy data-sets to test the robustness of the classifiers.
- 25% of the studies included multi-level architecture to improve accuracy of prediction - the key aim being to successively improve the accuracy by recording false negatives/ positives[13,14,18].
- 20% of the studies indicated that retrieving relevant and accurate medical data is a major problem plagued with issues such as data integrity, data confidentiality, data consistency, data security and data ambiguity.

TABLE VI
ACCURACY PREDICTION (ILLUSTRATIVE CHART)

Category	Total Cases	Correct Prediction	Incorrect Prediction	Prediction Accuracy
1st Run	20	12	8	60%
2nd Run	20	12	8	60%
3rd Run	20	14	6	70%

III. CONCLUSION

Disease prediction is one of the critical task while designing medical diagnosis software. Machine Learning techniques have been successfully utilized in number of studies to assist in medical diagnosis. In all the studies, disease prediction is based on clinical /symptoms and laboratory data. However, since the learning data set used in all studies (100%) limited and the fact that the same symptoms can mean different diseases depending upon other factors such as patient profile and external environment the efficacy of disease prediction is mid-50s (80% studies).

The techniques identified as part of RQ1 suggest that there is a difference in opinion on the usage across the board, however all studies indicate that expert doctor analysis cannot be replaced using current techniques, this is due to the fact that the Medical knowledge is very vast and there is an inherent uncertainty so, the disease prediction can at best supplement and enhance Doctor expert opinion. In RQ2, the key factor impacting the accuracy came out to be the quality and sufficiency of data set and the insight that the data sets have to be multiple and dynamic and continuously updated with newer data sets. Finally, the researchers conclude that more research needs to be done and that there is a need to collate and map clinical data with Medical data for enhanced decision making.

REFERENCES

[1] 1. Kitchenham, B.A., Charters, S, Guidelines for performing Systematic Literature Reviews in Software Engineering V 2.3. School of Computer Science and Mathematics Keele University, Department of Computer Science University of Durham, 2007

[2] Dhaval Raval, Dvijesh Bhatt, Malaram K Kumhar, Vishal Parikh, Daiwat Vyas, Medical Diagnosis System Using Machine Learning, International Journal of Computer Science, vol. 7, issue 1, pp 177-182, Mar 2016

[3] R. Isola, R. Carvalho, M. Iyer, and A. K. Tripathy, Automated differential diagnosis in medical systems using neural networks, knn and som, in Developments in E-systems Engineering (DeSE), 2011,pp. 6267, IEEE, 2011.

[4] S. Zahan, C. Michael, and S. Nikolakeas, A fuzzy hierarchical approach to medical diagnosis, Proceedings of the Sixth IEEE International Conference on Fuzzy Systems, 1997 pp. 319324

[5] B. Thakkar, M. Hasan, M. Desai, et al., Health care decision support system for swine flu prediction using naive bayes classifier, in Advances in Recent Technologies in Communication and Computing(ARTCom), 2010 International Conference, pp. 101105

[6] A. A. Bakar, Z. Kefli, S. Abdullah, and M. Sahani, Predictive models for dengue outbreak using multiple rule base classifiers, in Electrical Engineering and Informatics (ICEEI), 2011 International Conference, 2011, pp. 16

[7] Manjeevan Seera, Chee Peng Lim, A hybrid intelligent system for medical data classification, Expert Systems with applications vol. 41 issue 5, pp. 22392249, 2014

[8] Min Chen , Yixue Hao , Kai Hwang , Lu Wang , and Lin Wang, Disease Prediction by Machine Learning Over Big Data From Healthcare Communities, IEEE Access pp.8869 8879, 2017

[9] Sellappan Palaniappan, Rafiah Awang , Intelligent Heart Disease Prediction System Using Data Mining Techniques, IEEE/ACS International Conference on Computer Systems and Applications, 2008

[10] Shadab Adam Pattekari and Asma Parveen, Prediction System For Heart Disease Using Naive Bayes, International Journal of Advanced Computer and Mathematical Sciences, pp. 290-294, 2012

[11] Leroy Hood and Mauricio Flores, A personal view on systems medicine and the emergence of proactive P4 medicine predictive, preventive, personalized and participatory, New Biotechnology, vol. 29, issue 6, 2012.

[12] Manjeevan Seera, Chee Peng Lim, Wei Shiung Liew, Einly Lim, Chu Kiong Loo - Classification of electrocardiogram and auscultatory blood pressure signals using Machine learning models, Expert Systems with Applications, Elsevier, 2014

[13] Zhia-Hua Zhou, Yuan Jiang, Yu-Bin Yang, Shi-Fu Chen, Lung Cancer cell identification based on artificial neural network ensembles, Artificial Intelligence in Medicine, vol. 24, pp. 25-36, 2002

[14] Jia-jia Sim, Garry Wei-Han Tan, Jessica C.J. Wong, Keng-Boon Ooi, Teck-Soon-Hew, Understanding and predicting the motivators of mobile music acceptance - A multi-stage MRA artificial neural network, Telematics and Informatics, pp 569-584, 2014

[15] Wei-Yen Hsu, Brain-Computer interface - The next frontier of telemedicine in human-computer interaction, Telematics and Informatics, pp. 180-192, 2015

[16] Cheng-Hsuing Weng, Tony Cheng-Kui Huang, Ruo-Ping Han, Disease Prediction with different types of neural network classifiers, Telematics and Informatics, pp. 277-292, 2016.

[17] Dursun Delen, Glenn Walker, Amit Kadam, Predicting breast cancer survivability: a comparison of three data mining methods, Artificial Intelligence in Medicine, pp. 113-127, 2005

[18] Hasan Temurtas, Nejat Yumusak, Feyzullah Temurtas, A comparative study on diabetes disease diagnosis using neural networks, Expert Systems with Applications, pp. 8610-8615, 2009

[19] Freddie Astrom, Rasit Koker, A parallel network approach to prediction of Parkinsons Disease, Expert Systems with Applications, pp. 12470-12474, 2011

[20] Song-Pan, Serdar Iplikci, Kevin Warwick, Tipu Z. Aziz, Parkinsons Disease Tremor Classification - A comparison between Support Vector Machines and neural networks, Expert Systems with Applications, pp. 10764-10771, 2012

[21] E de Bruin, J G Loeber, Ameijer, G Martinze Castillo, ML Granados Cepeda, M Rosario Torres-Sepulveda, GJC Borrajo, M Caggana, Y Giguere, M Meyer, M Fukushi, AR Rama Devi, I Khneisser, L Vilarinho, U von Dobeln, T Torresani, J Mackenzie, I Zutt, M Schipper, L H Elvers, MPG Koopmans, Evolution of an influenza pandemic in 13 countries from 5 continents monitored by protein microarray from neonatal screening blood spots, Journal of Virology, pp. 74-80

[22] Igor Kononenko, Machine Learning for Medical Diagnosis: History,

State of the Art and perspective, Artificial Intelligence in Medicine, 2001

[23] Ankita R Borkar, Prashant R Deshmukh, Naive bayes Classifier for Prediction of Swine Flu Disease, International Journal of Advanced Research in Computer Science and Engineering, vol. 5, issue 4, 2015

[24] Saroj K Biswas, Manomita Chakraborty, Biswajit Purkayustha, An ANN based Classification Algorithm for Swine Flu Diagnosis, National Institute of Technology Silchar, May 2016

[25] Chiung Ching Ho, Choo Yee Ting, Dhesi Baha Raja, Using Public Open Data to Predict Dengue Epidemic: Assessment of Weather Variability, Population Density, and Land use as Predictor Variables for Dengue Outbreak Prediction using Support vector machine, Indian Journal of Science and Technology vol.11, 2018

[26] R Isola, Rebeck Carvalhi, Amiya Kumar Tripathy, Knowledge Discovery in Medical Systems Using Differential Diagnosis, LAMSTAR, and k-NN, IEEE Transactions on Information Technology in Biomedicine, vol. 16, 2012

[27] Noor Diana Ahmad Tarmizi, Fraha Jamaluddin, Azuraliza Abu Bakar, Zulaiha Ali Othman, Abdul Razak Hamdan, Classification of Dengue Outbreak using Data Mining Models, Research Notes in Information Science, vol. 12, 2013

[28] Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni, Predictive Data Mining for Medical Diagnosis: An overview of Heart Disease Prediction, International Journal of Computer Applications, vol. 17, 2011

[29] Jean Patric, Using different characteristics of machine learners to identify different defect families, Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering, 2016

[30] Chris Cummins, Pavlos Petoumenos, Zheng Wang, Hugh Leather, Synthesizing benchmarks for predictive modeling, Proceedings of the 2017 International Symposium on Code Generation and Optimization, 2017

[31] Daniele Di Mitri, Maren Scheffel, Hendrik Drachsler, Dirk Brner, Stefaan Ternier, Marcus Specht, Learning pulse: a machine learning approach for predicting performance in self-regulated learning using multimodal data, Proceedings of the Seventh International Learning Analytics & Knowledge Conference, 2017

[32] Tushar Chandra, Regressing Towards Simpler Prediction Systems, Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, 2015

[33] Jesse Davis, Vitor Santos Costa, Soumya Ray, David Page, An integrated approach to feature invention and model construction for drug activity prediction, Proceedings of the 24th international conference on Machine learning, 2007

[34] Josh Gardner, Christopher Brooks, A Statistical Framework for Predictive Model Evaluation in MOOCs, Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale, 2017

[35] Bo Long, Large scale machine learning for response prediction, Proceedings of the 2nd Workshop on Parallel Programming for Analytics Applications, 2015

[36] Dong Wang ; Kwok-Leung Tsui ; Qiang Miao, Prognostics and Health Management: A Review of Vibration Based Bearing and Gear Health Indicators, IEEE Access, 2017

[37] Dipali Singla ; Prashant Singh Rana, Eye state prediction using ensembled machine learning models, Inventive Computation Technologies (ICICT), International Conference, 2015

[38] Anis Davoudi ; Tezcan Ozrazgat-Baslanti ; Ashkan Ebadi ; Alberto C. Bursian ; Azra Bihorac ; Parisa Rashidi, Delirium Prediction using Machine Learning Models on Predictive Electronic Health Records Data, Bioinformatics and Bioengineering (BIBE), IEEE 17th International Conference, 2017

[39] V.U.B. Challagulla ; F.B. Bastani ; I-Ling Yen ; R.A. Paul, Empirical assessment of machine learning based software defect prediction techniques, Object-Oriented Real-Time Dependable Systems, 10th IEEE International Workshop,2005

[40] Durga L. Shrestha, Dimitri P. Solomatine, Machine learning approaches for estimation of prediction interval for the model output, Springer Special Issue, 2006

[41] Sapan Shah, Sreedhar Reddy, Application of Machine Learning Techniques for Inverse Prediction in Manufacturing Process Chains, Proceedings of the 3rd World Congress on Integrated Computational Material Engineering (ICME), 2015

[42] Ted Gueniche, Philippe Fournier-Viger, Vincent S. Tseng, Compact Prediction Tree: A Lossless Model for Accurate Sequence Prediction, Part II of the Proceedings of the 9th International Conference on Advanced Data Mining and Applications, vol. 8347, ADMA, 2013

[43] Yunlong Liu, Yifeng Zeng, Hexing Zhu, Yun Tang, Making and Improving Predictions of Interest Using an MDP Model, Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, 2017

[44] Ben Taskar, Vassil Chatalbashev, Daphne Koller, Carlos Guestrin, Learning structured prediction models: a large margin approach, Proceedings of the 22nd international conference on Machine learning, 2005

[45] Md. Osman Gani, Sarah Isnain Binte Ashraf, Nafia Malik, Bushra Hossain, M. Afzal Hossain, Hasan Sarwar, Recursive Implementation of Markov Model, A New Approach, Proceedings of the 2011 IEEE Workshops of International Conference on Advanced Information Networking and Applications, 2011

[46] Ioannis Tsochantaridis, Thomas Hofmann, Thorsten Joachims, Yasemin Altun, Support vector machine learning for interdependent and structured output spaces, Proceedings of the twenty-first international conference on Machine learning, 2004

[47] V.S.H Rao, M N Kumar, A new Intelligence based approach for computer aided diagnosis of dengue fever, Information Technology in Biomedicine, IEEE Transactions, vol. 16, pp. 112-118, 2012

[48] Joanne M Meyer, Geoffrey S Ginsburg, The path to personalized medicine, Current opinion in Chemical Biology, vol. 6, pp 434-438, 2002

[49] Michele Berlingerio, Francesco Bonchi Fosca Giannotti, Fraco Turini, Mining Clinical Data with a Temporal Dimension: a Case Study:, IEEE International Conference on Bioinformatics and Biomedicine, 2007

[50] Renqiang Min, Stanley DA, Zineng Yuan, Bonner A, Zhaolei Zhang, A Deep Non-linear Feature Mapping for Large Margin kNN Classification, Data Mining, Ninth IEEE International Conference, 2009

[51] Chien-Pen Chuang, Shiunn Shin Lee, Jia-Shiunn Tsai, Tai-Jung Kao, Detecting mammograhpy of breast microcalcification with SOL-based self-organization neural network, Natural Computation (ICNC), Sixth International Conference, 2010