# Comparing Existing Methods for Predicting the Detection of Possibilities of Blood Cancer by Analyzing Health Data

**Gandhi Priyank Sanjay**
*School of Doctoral Research & Innovations, GLS University*

**Dr. Viral Nagori**
*School of Doctoral Research & Innovations, GLS University*

## Abstract

E-health is the word of this era and had been much popularized due to its advantages, but still doesn't have the clear definition of the word. Till 90's it was merely a word in a dictionary but now is often used in general practice and its more than just "internet medicine", it's all about virtually relating computer, medicines, technology and healthcare. This term was previous used by industrialists and people related to marketing and business, but now it's often used in academics. Machine Learning methods have change the way to deal with the diseases. For cancer, it may improve an understanding, the way how it progressed, the methods to validate and various ways to improve the health by minimizing its disastrous effect of the disease and improve the patients' health. The aim of the current research is to analyze large data obtained from health records using data mining and machine learning algorithms. The study of existing work is to understand better the current scenario and the work in medical science for detecting the blood cancer.

**Keywords: Blood Cancer, Mining, Cancer Detection, Health, ANN, Ehealth**

## I. INTRODUCTION

E-health is the word of this era and and had been much popularized due to its advantages, but still doesn't have the clear definition of the word. Till 90's it was merely a word in a dictionary but now is often used in general practice and its more than just "internet medicine", it's all about virtually relating computer, medicines, technology and healthcare. This term was previous used by industrialists and people related to marketing and business, but now it's often used in academics.[18] Internet and its related technologies gaps the bridge in providing various medical information, health related services and detailed informatics of the outcome. This word, E-health is not just only a word but a technical revolution along with a change in mindset to think, connect and an attitude to commit for networked technologies, thinking globally and improving healthcare worldwide by availing and refining the availed information. [18]

The second leading cause of death in the United States is Cancer or malignant neoplasm. Cancer consists of more than 100 different disease that may affect all major organs and every functioning of the body. [21] All cancers arise from abnormal and uncontrolled cell growth brought about by changes or damage in a cell's DNA. When mutations occur or a cell gains or loses a chromosome during mitosis, the biological pathways that normally inhibit cell division are disabled, resulting in a proliferation of damaged cells that may ultimately metastasize to other parts of the body. [3]

Machine Learning methods have change the way to deal with the diseases. For cancer, it may improve an understanding, the way how it progressed, the methods to validate and various ways to improve the health by minimizing its disastrous effect of the disease and improve the patients' health.

## II. VARIOUS METHODS OF ML

During the data preprocessing different types machine learning techniques are define such as

1) ANNs
2) DTs
3) SVMs
4) BNs is available

### *A. ANN*

Variety of pattern recognition and problems related to classification are handled by ANN's. Combination of various input variables generate an output via a trained mechanism.[18] For this process multiple hidden layers that are interconnected typically known as neural connections are used. However, some drawbacks are also associated with ANN as the layered generic structure of ANN is time consuming resulting poor performance. This technique is often known as "black-box" technology. Trying to find out how classification is done as well as the reason for failure of ANN is almost impossible to detect. [18]

All rights reserved by *www.ijirst.org*

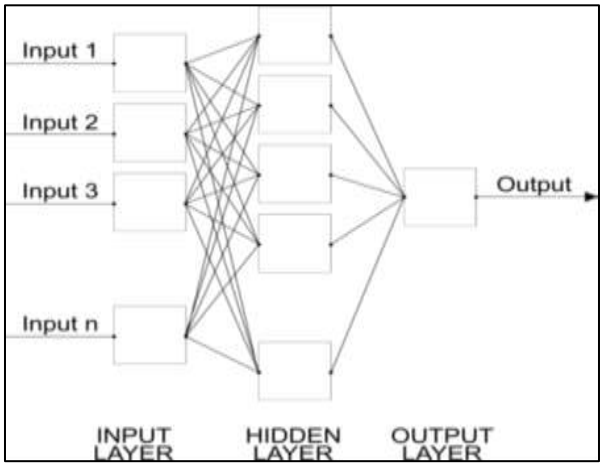

Fig. 1: ANN Architecture [18]

### *B. Decision Tree*

Decision Tree's almost follow the same way as a tree structure in which input is represented by nodes and output is represented as leaves. DT's are the oldest ML methods though robust classification purpose. DT's are simple to interpret, quick and easy to learn. A decision tree is a graph that uses a branching method to illustrate every possible outcome of a decision.[4]

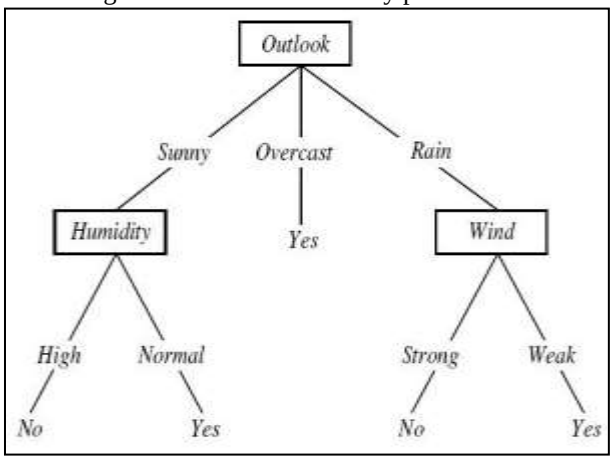

Fig. 2: Simple Decision Tree [4]

### *C. Support Vector Machine*

SVMs are a more recent approach of ML methods applied in the field of cancer prediction/prognosis. In a high or infinite-dimensional space hyper planes or a set of hyper planes are constructed by SVM's. Outlier's detection, regression or classification can be done with the help of this. [2] Support vector machines are simply the coordinates with different individual observations. It basically separates two classes with similar properties.

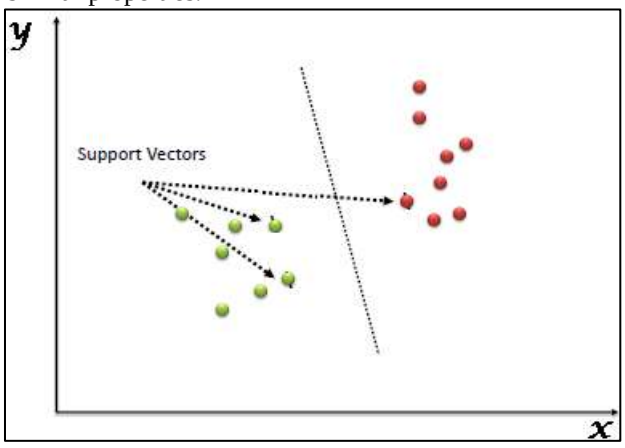

Fig. 3: Support Vector Machine [2]

All rights reserved by *www.ijirst.org*

### *D. Naïve Bayes Classifier*

Probability estimation is made by BN classifier rather than predictions. They represent directed acyclic graph of probabilistic knowledge to that of its related dependencies. It's widely used for reasoning purpose and knowledge representation. One of the powerful and straight forward method of classification is this method. [4] It works best with millions of records with attributes and also gives great results when used for textual data analysis.

## III. COMPARISON AMONG VARIOUS RELATED WORK

| *Sr No.* | *Title* | *Approach* | *Pros* | *Cons* |
|---|---|---|---|---|
| *1* | *Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network[9]* | *Convolutional neural network* | – *Best in class performance.*<br>– *Reduces the need for feature engineering.*<br>– *can be adapted to new problems relatively easily* | – *Requires a large amount of data.*<br>– *Extremely computationally expensive to train.* |
| *2* | *Detection of Leukemia in Human blood sample through Image Processing[10]* | *Naïve Bayesian (NB), K-Nearest Neighbor (KNN), Multilayer Perceptron (MLP), Radial Basis Functional Network (RBFN), and Support Vector Machine (SVM).* | – *Produce very accurate classifiers.*<br>– *Less overfitting, robust to noise.*<br>– *Simple to implement*<br>– *Flexible to feature / distance choices* | – *SVM is a binary classifier.*<br>– *Computationally expensive, thus runs slow.*<br>– *Large search problem to find nearest neighbours*<br>– *Storage of data.*<br>– *Must know we have a meaningful distance function.* |
| *3* | *The Detection and Classification of blast cell in Leukaemia Acute Promyelocytic Leukaemia ( AML M3) blood using Simulated Annealing and Neural Networks[11]* | *Hill Climbing, Simulated Annealing and Genetic Algorithms* | – *Depend on a suitable starting point.*<br>– *Finding this starting point has been found to be the major hurdle when trying to do unattended, automatic optimizations.* | – *Increasing the number of variables, the number of evaluations increases as well.*<br>– *Optimizations with many variables are usually impractical.* |
| *4* | *Comparative Analysis of Various Techniques Based on Leukemia Detection[12]* | *ALL and AML by using K-NN* | – *Very easy to understand and implement.*<br>– *Does not assume any probability distributions on the input data.*<br>– *Can quickly respond to changes in input* | – *Sensitive to localized data.*<br>– *Computation time.*<br>– *Normalization*<br>– *Dimensions* |
| *5* | *Machine Learning for Medical Diagnosis[14]* | *Naive Bayesian classifer, neural networks and decision trees.* | – *Feature Learning*<br>– *Parameter optimization* | – *Works with continuous loss functions*<br>– *Limited*<br>– *Large data requirement.* |
| *6* | *Optimistic Diagnosis of Acute Leukemia Based On Human Blood Sample Using Feed Forward Back Propagation Neural Network[15]* | *Feed Forward Back Propagation Neural Network, Statistical Approach and Fuzzy Inference System* | – *The ease to model your reasoning;*<br>– *The ability to deal with uncertainty and nonlinearity;*<br>– *The ease of implementation;*<br>– *The use of linguistic variables.*<br>– *can deal hybrid decision systems (datasets)* | – *Requires high computation time.*<br>– *Not robust at all* |
| *7* | *A Framework for White Blood Cell Segmentation in Microscopic Blood Images Using Digital Image Processing[16]* | *Gradient vector flow* | – *Autonomously and adaptively search for the minimum state.*<br>– *External image forces act upon the snake in an intuitive manner.* | – *Sensitive to local minima states.*<br>– *Minute features are often ignored during energy minimization over the entire contour.* |

All rights reserved by *www.ijirst.org*

| | | | | |
|---|---|---|---|---|
| | | | – *Incorporating Gaussian smoothing in the image energy function introduces scale sensitivity.*<br>– *Can be used to track dynamic objects.* | – *Accuracy depends on the convergence policy.* |
| *8* | *Comparative Analysis of Classification Algorithms for the Prediction of Leukemia Cancer[17]* | *Naïve Bayes Classifier, Decision Tree Classifier , Lazy Classifier (LC)* | – *Feature Learning*<br>– *Parameter optimization* | – *Works with continuous loss functions*<br>– *Limited*<br>– *Large data requirement.* |
| *9* | *Combined Numerical and Linguistic Knowledge Representation and Its Application to Medical Diagnosis[18]* | *Hybrid Intelligent System, Incremental Learning Fuzzy Network (ILFN)* | – *Relatively easy to use*<br>– *Can approximate any function, regardless of its linearity*<br>– *Great for complex/abstract problems like image recognition.*<br>– *Mimics the brain (but there is some controversy over the degree of this)* | – *Often abused in cases where simpler solutions like linear regression would be best*<br>– *Requires a shit load of training and cases*<br>– *Black box that not much can be gleaned from*<br>– *Increasing accuracy by a few percent can bump up the scale by several magnitudes* |
| *10* | *CANFIS—a computer aided diagnostic tool for cancer detection* | *Neural Network; Coactive Neuro-Fuzzy ,Inference Systems, Probab ilistic Neural Network[19]* | – *Relatively easy to use.*<br>– *Can approximate any function, regardless of its linearity.*<br>– *Great for complex/abstract problems like image recognition.* | – *Often abused in cases where simpler solutions like linear regression would be best*<br>– *Requires a load of training and cases*<br>– *Increasing accuracy by a few percent can bump up the scale by several magnitudes.* |

## IV. UNDERSTANDING OF THE VARIOUS WORK DONE

The aim of the current research is to analyze large data obtained from health records using data mining and machine learning algorithms. The study of existing work is to understand better the current scenario and the work in medical science for detecting the blood cancer. Our aim to detect the possibility of blood cancer can be find out by analyzing various work done in past to gain the consciousness of what is to be done in future to achieve our goal of benefiting the mankind.

## REFERENCES

[1] Marios Anthimopoulos, Stergios Christodoulidis, Lukas Ebner, Andreas Christe, Stavroula Mougiakakou, Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network, DOI 10.1109/TMI.2016.2535865, IEEE.
[2] Rege M.V,Dr.B.W. Gawli, Detection of Leukemia in Human blood sample through Image Processing, International Journal of Modern trends in Engineering & Research, ISSN:-2349-9745.
[3] Waidah Ismail, Rosline Hassan, Annette Payne, Stephen Swift, The Detection and Classification of blast cell in Leukaemia Acute Promyelocytic Leukaemia ( AML M3) blood using Simulated Annealing and Neural Networks ,Department of information System & computing,Brunel University,West London,UK.
[4] Inderjeet kaur, Jyoti Rani, Comparative Analysis of Various Techniques Based on Leukemia Detection, International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 4, April 2016.
[5] Yu. Kiselev, Blood Analysis for Early Cancer Detection, Physics of Particles and Nuclei, 2014, Vol. 45, No. 1, pp. 344–346. Pleiades Publishing, Ltd., 2014, ISSN 1063_7796.
[6] Igor Kononenko, Machine Learning for Medical Diagnosis, University of Ljubljana, Faculty of Computer and Information Science, Tr·za·ska 25, 1001 Ljubljana, Slovenia.
[7] M.R. Thansekhar , N. Balaji, Optimistic Diagnosis of Acute Leukemia Based On Human Blood Sample Using Feed Forward Back Propagation Neural Network, International Journal of Innovative Research in Science, Engineering and Technology, Volume 3, Special Issue 3, March 2014.
[8] Farnoosh Sadeghian, Zainina Seman, Abdul Rahman Ramli, Badrul Hisham Abdul Kahar, and M-Iqbal Saripan, A Framework for White Blood Cell Segmentation in Microscopic Blood Images Using Digital Image Processing, Shulin Li (ed.), Biological Procedures Online, Volume 11, Number 1, DOI: 10.1007/s12575-009-9011-2 URL: springerprotocols.com; springerlink.com.
[9] Durairaj M, Deepika R, Comparative Analysis of Classification Algorithms for the Prediction of Leukemia Cancer, International Journal of Advanced research in Computer science and software engineering, Volume 5, Issue 8, August 2015.
[10] Phayung Meesad, Gary G. Yen, Combined Numerical And Linguistic Knowledge Representation And Its Application To Medical Diagnosis, IEEE Transactions On Systems, Man, And Cybernetics—Part A: Systems And Humans, Vol. 33, No. 2, March 2003.
[11] Latha Parthiban, R. Subramanian, CANFIS—a computer aided diagnostic tool for cancer detection, J. Biomedical Science and Engineering, 2009, 2, 323-335

All rights reserved by *www.ijirst.org*

[12] Htet Thazin Tike , An Approach For Breast Cancer Diagnosis Classification Using Neural Network, Myanmar, Advanced Computing: An International Journal (ACIJ), Vol.6, No.1, January 2015
[13] Xie, J., et al., Novel Hybrid Feature Selection Algorithms for Diagnosing Erythemato-Squamous Diseases, J. He, et al., (Editors). 2012, Springer Berlin Heidelberg. p. 173-185.
[14] Liao, B., et al., A Novel Hybrid Method for Gene Selection of Microarray Data. Journal of Computational and Theoretical Nanoscience, 2012. 9(1): p. 5-9.
[15] Grzymala-Busse, J.W. and W.J. Grzymala-Busse, Handling Missing Attribute Values Data Mining and Knowledge Discovery Handbook, O. Maimon and L. Rokach, (Editors). 2010, Springer US. p. 33-51.
[16] Hall, M.A. and G. Holmes, Benchmarking Attribute Selection Techniques for Discrete Class Data Mining. IEEE Transactions on Knowledge And Data Engineering, 2003. 15(3).
[17] Feng, D.D., Biomedical information technology. 2008, Amsterdam; Boston: Elsevier/Academic Press.
[18] Song, H., et al., New methodology of computer aided diagnostic system on breast cancer, in Proceedings of the Second international conference on Advances in Neural Networks - Volume Part III. 2005, Springer-Verlag: Chongqing, China. p. 780-789.
[19] Han, J. and K. M, Data Mining Concepts and Techniques. Vol. 3. 2011, San Franscisco: Morgan Kaufmann.
[20] http://www.mayoclinic.org/diseases-conditions/cancer/in-depth/cancer-diagnosis/art-20046459 Dt:-15/02/2017 11:39 AM
[21] http://www.cancercenter.com/treatments/diagnostics/ Dt:-12/02/2017 06:24 PM
[22] http://whatis.techtarget.com/definition/machine-learning dt:-28/01/2017 03:34 PM.
[23] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3830014/ 25/07/2017 04:55 PM.

All rights reserved by *www.ijirst.org*