# Comparative Study of Recent Trends on Cancer Disease Prediction using Data Mining Techniques

Satyam Shukla[+], Dharmendra Lal Gupta[#] and Bakshi Rohit Prasad[*]

{[+]KNIT Sultanpur U.P., [#]KNIT Sultanpur U.P., *IIIT Allahabd,U.P.}
{[+]shuklasatyam55@gmail.com, [#]dlgupta2002@gmail.com and
*rohit.cs12@gmail.com}

## Abstract

*Technological advancements have evolved into several application domains to solve various problems. One such technological area is Data Mining. It has shown its significance and potential in health care industries to serve as a guiding and decision making component. Its potential in unveiling new trends in health care organizations has proved its importance for all people associated with this area. It is the most important and encouraging area of research which have the motive to find out the information from large data set. Advance researches in data mining had made it a key player in health care field. Good analytical techniques are of utmost requirement for detecting precious information lying hidden in health industry data. This survey paper presents the importance and usefulness of different Data mining techniques such as classification, clustering, Decision Tree, Naive Bayes etc. in health domain. Here the study and comparison is done of different data mining techniques used for prediction of cancer disease from clinical dataset with different accuracy.*

*Keywords*: Cancer. Disease Prediction, Data mining, Classification

## 1. Introduction

One of the most encouraging and motivating areas of computer science that uses different statistical techniques is Data mining. The Data mining techniques are useful in finding patterns and relationships for the prediction of various diseases from the medical field [1]. There are plenty of cases where data mining techniques are being applied for the diagnosis of different diseases like heart disease [2], diabetes [2-3] Parkinson [2,4] cancer [2,5], *etc.*

According to WHO (World Health Organization) 8.2 million people die each year from cancer and it is estimated that 13% of total death worldwide in caused due to cancer 70% increase in new cases of cancer is expected over the next two decades. Over all 100 types of cancer exist each requiring unique treatment and diagnosis. The most commonly diagnosed cancer worldwide is of the Lungs (1.8 million, 13% of total) Breast cancer (1.7 million, 12% of total) Colorectal (1.4 million, 9.7 of total). The most common causes of cancer death are cancer of lungs (1.6 million, 19.1% of total) Liver cancer (0.8 million, 9.1% of total) Stomach cancer (0.7 million, 8.8 of total). It is estimated that by year 2025 increase to 19.3 million new cancer cases per year can be noticed due to growth and aging of the population growth. Cancer is one of the most important reason of mortality in different countries of the world [6].

The Figure 1 shows the estimated figure of total cancer cases in India which shows the seriousness of the issue.
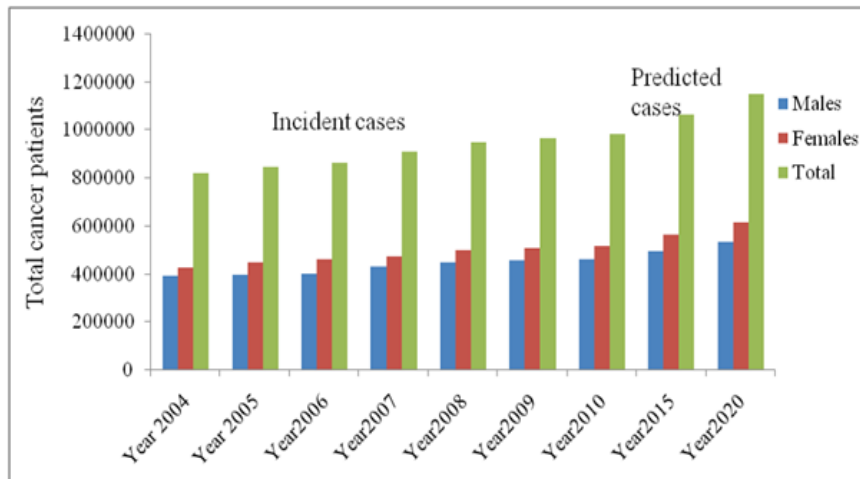
**Figure 1. World-wide Cancer Statistics**

This content of the paper proceeds with a brief introduction in Section I. Section II defines the statement of the problem that is targeted in this paper. Section III describes various types of cancer and major factors that cause it. In Section IV a detailed survey of various researches about several data mining techniques and methodologies that have been applied in reference to diagnosis and prediction of cancer disease is illustrated. Section V provides a comparative analysis of accuracies of different data mining techniques as reported by various researches in this domain. Section VI concludes the work and specifies the future work.

## 2. Problem Statement

New clinical techniques and database technologies are evolving rapidly. Clinical experts and personnel face various problems during collecting data and in applying those techniques in diagnosis and minimizing the disease. Several hurdles that often come across are as follows:

- Due to huge data availability information is not fully reliable and is incomplete.
- Cancer disease complication goes undiagnosed due to lack of good diagnostic level.
- Cancer remains undiagnosed due to no clear information, explanation and insufficient diagnostic criteria.
- Present cancer diagnostic methods are very expensive and time consuming.
- Difficulties are faced in identifying the types of cancer due to wrong entries in record.
- Lack of understanding of symptoms and risk factor from patient.

The proposed work presents a detailed survey of recent trends and technologies from data mining domain that are applicable to medical diagnosis. In following sections and subsections paper covers the severity scenario of cancer disease world-wide along with its types and symptoms. Moreover, it investigates various researches done in this field to diagnose different categories of cancers using data mining technologies. Several advantages and disadvantages are discussed about various works and technologies.

## 3. Cancer

Cancer is that deadly disease which is caused due to change in normal cells of the body and as a result there is uncontrolled growth of cells which give rise to tumor, except leukemia this is the main cause of cancer. If tumor is not treated in time it grows and spread into surrounding areas through bloodstream and affect the digestive, resource and

circulatory system and cause severe health consequences which is the important cause of death .Men are more prone to Lung, prostate, stomach, liver cancer. While women are more prone to breast, colorectal, lung, cervix uteri, and stomach cancer.

Collection of related disease is called cancer. If the proper treatment and diagnosis of the disease is not done in time as it is found in most of the cases this malignant disease can even cause death.

### 3.1. Classification of Cancer

In over all more than 200 types of cancer are found. The name of the cancer is given on the basis where the cell or tissue of the cancer form. For *e.g.* in lung cancer the cells of lung is affected so it is called lung cancer, likewise of brain and other cancer types. The type of cell that formed them are also responsible for their name and classification as epithelial or squamous cell. The following are some types of cancer that start in specific types of cell [6].

- **Carcinoma:** That arises from the epithelial cells. It may invade the surrounding tissues. The example of carcinoma forms of cancer are breast, prostate, lung and colon. Carcinomas have special name due to formation in different epithelial cells like Aden carcinoma is that type of cancer which is formed in that epithelial cells which produces fluids, such types of tissues are called glandules tissues.
- **Sarcoma:** A type of malignant tumour of the bone, or soft tissue *e.g.* are leiomyosarcoma, lip sarcoma, and osteo sarcoma. Sarcoma is the cancer of bone mainly. The most commonly found soft types of tissue sarcoma are leimyosarcoma, Kaposi sarcoma, malignard Fibrous histiocytoma, lip sarcoma, and dermatofibln sarcoma protuberant.
- **Lymphoma:** cancer that affect lymphatic system, and travels throughout the body and can affect any part of the body
- **Leukaemia:** Is a cancer of the tissue which forms blood cells, white blood cells and bone marrow. There are many subtypes like lymphatic leukaemia and chronic lymphocytic leukaemia. These four types of leukaemia are classified on the basis how quickly it get worse cell the cancer starts on this type of blood cells.

### 3.2. Causes of Cancer

- **Alcohol:** One of the main cause of cancer is alcohol. Almost all alcoholic drinks cause cancer. It causes different types of cancers including mouth, throat, liver, and breast.
- **Tobacco:** Tobacco had such 80 substances that are the reasons for cancer. When tobacco is inhaled in the form of cigarette the chemicals enter into lungs reaches the lungs and then it is transported to blood stream and then throughout the body.
- **Overweight or Obesity:** Due to increase in higher insulin level overweight people are more prone to of cancer. It also increases the chances of cancer of food pipe, Kidney, gallbladders, breast and womb cancer in women.
- **The Immune System:** People who have weak immune system are more at risk of developing cancer. People having HIV or AIDS have weaker immune system and they are more prone to the cancer.
- **Carcinogens:** Genes have coded messages which tell them which protein to make. Change in gene like damage can stop their normal behaviour or working and then cell is more likely to be cancerous is called carcinogen.
- **Infection:** Due to different types of infection there is a risk of cancer mainly infection caused by virus.

## 4. Related Work

Plenty of work and researches have been done to find out different methods of diagnosis of various cancers types. It is an attempt to predict and diagnose the cancer disease based on symptoms that occurs at an early stage.

### 4.1.1. Predicting Lung Cancer using KNN, Logistic Regression, Random Tree Classifier [7]

This paper discusses the lung cancer which is one of the deadly diseases of lungs. Based on that, feature selection process short lists 20 such parameters. Some of those influencing parameters are weight loss, bloody mucus, back pain *etc.* Here, researchers focus on pre-diagnosis which is considered as the most vital stage to know the susceptible patients for going through special diagnosis process.  They noticed that supervised learning ways are far more better than to the cross validation approach. From this research, it was found that random tree classifier, KNN, logistic, multilayer perception, sequential minimal, optimization has given much more better and reliable performance in this domain.

### 4.1.2. Diagnosis of Cancer through Rank Based <ethod [8]

A molecular alterations in carcinogenesis, one of the most common forms, is characterized by Aberrant DNA methylation cancer. In this research, authors first applied different ways to reveal several methylation patterns that include meaningful rules. These identified rules from methylation profile were capable to exactly differentiate between samples representing tumor and normal case. Thus, the main focus of their approach was DNA methylation profiles. These profiles were collected from TCGA database. Then after, a Rank-Based method is applied to discriminate effectively a normal tissue from tumor tissues. This technique used the relative methylation level reversal pairs as candidate markers pairs. Then identification of the most different R-pairs in all candidates markers pairs are done by analyzing the CpG sites having highest appearance frequencies.

### 4.1.3. Prediction of Breast Cancer Using Algorithm ANN, Naive Bayes [9]

This research dealt with breast cancer and the prediction of the disease was done through Artificial Neural Network (ANN), Logistic regression, Naive Bayes techniques. The objective of the research aims at giving the following outcomes; firstly, it evaluates medical data set in terms of quality grammatically and secondly, it evaluates data mining methods with respect to their applicability to the data.  Finally, the knowledge extracted from the data set is used for disease prediction by applying Artificial Neural Network (ANN), Logistic Regression, Naïve Bayes. It is found that these techniques had highest lifting factor for most of class values.

### 4.1.4. Predicting Oral cancer using ANN, SIR's Novel Index [10]

The prime focus of this work is to characterize the distribution of colorectal cancer risk using family history of cancer. In colon cancer registries of population cancer were made for 10,066 colorectal cancer cases of families. Family history is analyzed using data mining techniques such as Novel Index ANN and Standardized Incidence Ratios (SIR's). These techniques are used basically to know risk of the disease.. This study identifies 5 big and 66 small classes of risk.

### 4.1.5. Prediction of Different Types of Cancer Using Clustering, Classification, Bio Medicine Genetics. [11]

The advantages of data mining and special characteristics of health data make data mining very important to be considered in health data analysis. This research aims to identify those information which demonstrate the importance of data mining in health care. A critical analysis of the cancer data is carried out in the work and some significant hidden information are extracted out of it. Useful predictive solution are developed using clustering, classification, Bio-medicine and genetics.

### 4.1.6. Diagnosing Cancer via Decision Tree, KNN, Support Vector Machine, NN [12]

Classification based pattern analysis techniques are used in this work for diagnosing the cancer. Several well known classification algorithm such as DT (Decision Tree), SVM (Support Vector Machine), KNN (K-Nearest Neighbor) and NN (Neural Networks) are used for diagnoses of cancer. It is established that the process of classification depends on the value of various features in the collected data. Here, the authors found that medical disease data often have some noise data as well as boundary value data. They suggested techniques to deal with such noisy data. For optimization of accuracy they employed Ant Colony Optimization technique.

### 4.1.7. Predicting Oral Cancer Using NN, Fuzzy Logic, Soft Computing Techniques [13]

Authors in their work used several data mining approaches like classification, classification rule mining, soft-computing techniques, Neural Networks and Fuzzy logic for diagnosis of oral cancer. They specified the effectiveness of each of the above techniques for the classification task in medical domain. Apart from it, they showed the importance of genetic algorithms in optimizing the data mining algorithm in terms of accuracy for prediction.

### 4.1.8. Predicting Cancer by Using ESPO LAZY IBK Classifier [14]

Researchers here addressed the problem of selection of small subset of genes from broad patterns of gene expression data recorded on DNA micro arrays from the available training dataset. In this work, authors proposed method of gene selection by utilizing Elitism Particle Swarm Optimization (ESPO) based on Recursive Feature Reduction (RFR) is select. In this way they discriminated the patterns of gene expression of cancer and normal patients. Further, authors built a classifier suitable for genetic diagnosis.

### 4.1.9. Prediction of Gastric Cancer Using Decision Tree [15]

The main objective of this paper is to find the most important things which are responsible for the death of people suffering from gastric cancer. In addition, researchers introduced Decision Tree model for research by applying classification approach. For this, two categories of patients were made one of dead and the other one of alive patients and selected 20% of data set randomly as test samples and remaining data set were considered as the training sample.

### 4.1.10. Predicting Breast Cancer via K-means Clustering and Decision Tree [16]

The main focus of this work is on the analysis of breast cancer classification and prediction so that preventive measures can be made at early stage before the onset of the breast cancer. Different data mining techniques such as Decision Tree, Clustering Algorithm are employed to achieve the objective. Observation mining technique common

sensual on the data base fly in the ointment relationship and traditions takes are helpful in studying the progression of disease.

### 4.1.11. Prediction of Acute Myeloid Leukemia Cancer Using ANN- Regression Tree Algorithm [17]

The study addressed the issue of applying medical data mining and using various data mining techniques for diagnosis of Acute Myeloid Leukemia cancer. Various techniques employed in this work are clustering, regression, classification and a survival prediction model is constructed out of them. This paper discusses three important aspects; firstly it presents significance of data mining approach in this regard, secondly it provides a comprehensive survey related to the selected task and finally it compares the correct accuracy level of various models.

### 4.1.12. Predicting Breast Cancer Using KNN Imputation, EM-imputation, Logistic Regression [18]

In this research work in which the data set are taken from Institute of Portuguese for breast cancer, the approach proposed by the researchers deal effectively with a dataset having high percent of unknown categorical information. Patients suffering from breast cancer are mainly emphasized in this study and their survival as well as to help in improving the treatment of disease. Different model such as EM-implementation, KNN-implementation, classification trees, logistic regression are constructed.

### 4.1.13. Predicting Lung Cancer Using Ant Colony Optimization Technique [19]

The work proposed here emphasizes on determining the factors which are responsible for lung cancer. It categorizes persons as the smokers and non smokers. Further, aims to classify people affected with lung cancer based on the smoking habit via applying data mining techniques. Multiple algorithms used to achieve the goal are Decision Tree and Ant Colony Optimization. Through this paper, we find that data mining techniques have pivot role in finding out the hidden information in medical data and the quality if data is improved by data processing

### 4.1.14. Detection of Cancer Using Partitioned Clustering Classification [20]

Central theme of this work is woven for nurses and other health care professionals who care for and educate cancer patients and their family about lung cancer symptoms, path physiology and treatment.  In order to achieve better accuracy in the prediction of disease and improving survivability rate various significant techniques like Partitioned clustering classification are used. They facilitates better medicos information system and help to list type of lung cancer, treatment options for non small cell lung cancer.

### 4.1.15. Prognosis of Oral Cancer Using ANN Logistic Regression. Decision Tree [21]

Authors of the work proposed a procedure of combination of two data mining technology namely clustering and classification features to predict the difference in symptoms of past cases where patients survived or died of oral cancer. For diagnosis, they adopted data mining techniques like Decision Tree, ANN, Logistic Regression and these techniques were effectively used to practically analyze oral cancer. In addition, it provides information to health physicians to take preventive measures

**4.1.16. Predicting Cancer Using KNN, SMO, KNN, Naive Bayes in Abu Dhabi [22]**

Keeping in mind the need of prediction of health care services in Abu Dhabi mainly four models were built by using data mining techniques that help the planners to take appropriate decision in health authority and Abu Dhabi government that which type of health care services should be carried out either in the form of hospital or clinic. To achieve objective of the work, KNN, SMO (Sequential Minimum Optimization and, Naïve Bayes are used.

# 5. Results and Discussion

Based on in depth study and analysis of various literatures, it is found that data mining approach is a boon to health care systems. These approaches are capable enough in finding critical hidden patterns for the identification of disease and interrelationships among various parameters causing the deadly disease such as cancer. Moreover, data mining techniques can be significantly used to construct several predictive models which are extremely useful for disease diagnosis at an early stage, even before its onset. Also, these prediction models provides a deep insight to decision makers who have to take critical decision regarding proper medication, change of medicine to patient and escalating line of treatment, *etc.* A detailed composition of various research developments in this domain is summarized in Table 1. This table gives a deep insight of several techniques employed by various researchers in their work. Also, it provides a detailed comparison of various advantages and disadvantages of different approaches in this regard.

**Table 1. Comparison Table of Various Methodologies**

| Year | Title | Technique | Pros | Cons |
|------|-------|-----------|------|------|
| 2013 [7] | Ensemble based optimal classification model for pre-diagnosis of lung cancer | KNN, Random Forest, Random Tree Classifier | These techniques give better performance and reduce dimension of attributes | ANN can be better and give better results than KNN |
| 2014 [8] | Reversal DNA methylation pattern for cancer diagnosis | Rank based Method | Reversal pairs are real and can easily be transported to independent cohorts of samples which are helpful in cancer diagnosis. | It serve as only bio marker based cancer diagnosis |
| 2015 [9] | Implementation of Data mining Algorithm to Analysis Breast cancer | Neural, Network, Logistic Regression, Naive Bayes | Naive Bayes can handle discrete and numeric attribute value both and reduce cost of medical test | Neural Network can bear noisy data but may mislead and can cause health complication |
| 2015 [10] | Determining the familial risk distribution of colorectal | Neural Network, SIR's | Effective for hidden Cancer aggregation pattern & for classifying familiar risk | The results of the clustering of map are not easily identified every time. |
| 2014 [11] | Predicts survivability of | Ensemble of classifiers | Provides better accuracy | Relatively higher |

| | breast cancer | | | computation cost |
|---|---|---|---|---|
| 2016 [12] | Classification based pattern analysis on the Medical Data in Healthcare environment | Decision Tree, KNN, Neural Network | Naive Bayes can handle both discrete and numeric values | KNN is slow when the dataset is large |
| 2016 [13] | A New feature selection Method for oral cancer using Data mining techniques | Rule mining, NN, Fuzzy Logic | Extracts meaningful technique for efficient classification | This is not done with real time database |
| 2015 [14] | Cancer Classification using Elitism PSO Based LAZY IBK on Gene Expression Data | ESPO LAZY IBK Classifier | Helpful tool for gene selection | Not applicable for small Data |
| 2015 [15] | Predicting the probability of mortality of gastric cancer patients using decision tree | Decision Tree | Perform well with large dataset | Does not cover generalization aspect of fuzzy logic |
| 2015 [16] | Mining Based optimization for breast cancer analysis- A review | Decision tree K-mean clustering | Fruitful in decision making | SVM, NN can be better in making decision |
| 2015 [17] | Prediction of acute myeloid leukemia cancer using Data min- A survey | ANN, Classification Regression Tree, Survival Prediction-model | Indicates accurate solutions in many field apart from disease diagnosis | The correct accuracy of the classification algorithm are not stable and differ from one another |
| 2015 [18] | Missing date imputation on the 5-year survival prediction of Breast cancer with an unknown discrete values | KNN, Classifiers, Logistic Regression | Prediction model gives best result based on KNN . | Discards the variable with high value of MD which does not improve the survivor performance in more dataset |
| 2015 [19] | A study of mining lung cancer data for | Ant Colony Optimization | Helps in changing disease prediction value | Attributes cannot be increased much |

| | increasing or decreasing disease prediction value by using Ant colony optimization Techniques | | | because of computational complexity |
|---|---|---|---|---|
| 2015 [20] | Analysis of cancer detection system using data mining approach | Clustering Classification | Useful in improving the treatment effectiveness services | Neither patient nor healthcare organization are interested in sharing information |
| 2015 [21] | The Application of Data mining Technique to oral cancer prognosis | ANN, Logistic regression | The techniques used are much more better than traditional statistical model | Decision tree are easier to understand as compared to ANN |
| 2014 [22] | Data mining approaches for predicting demand for health care services in Abu Dhabi | KNN, Classification, Algorithm Naive Bayes | Learning Technique is fast stable and accurate | When data set are large the KNN is slow |

After careful analysis research works specifying the application of data mining techniques in medical domain for extracting different kinds of knowledge and patterns, it is evident that some of the data mining techniques have been extensively used in the literature as well as experiments. Each of these techniques has their own advantages and limitations. Moreover, different techniques have shown different accuracy level. Table 2 lists the techniques applied in cancer disease along with their prediction accuracy as reported in literatures. The relative accuracy comparison of accuracy is also depicted in Figure 2.

**Table 2. Comparison of Data Mining Approaches**

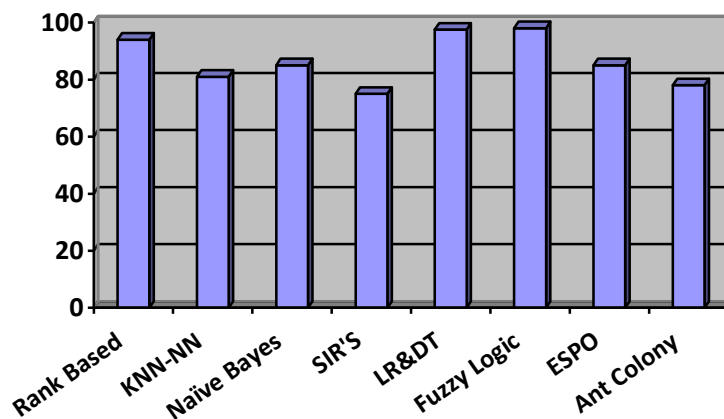| S. No. | Data Mining Technique | Accuracy in % |
|---|---|---|
| 1. | Rank - based | 94% |
| 2. | KNN-NN | 81% |
| 3. | Naive Bayes | 85% |
| 4. | SIR'S | 75% |
| 5. | Logistic regression & D.T. | 97.5% |
| 6. | Fuzzy logic | 98% |
| 7. | ESPO-Lazy IBK Classifies. | 85% |
| 8. | ANT Colony | 78% |

**Figure 2. Accuracy Chart of Data Mining Techniques in Health Care**

## 6. Conclusion and Future Work

Data mining techniques play an important role in medical research. Data mining approaches are valuable for research because it helps to gain the clinical data which helps in diagnosis and improving health of patient. After studying several papers, as discussed above, we came to know that by applying different data mining techniques timely diagnosis and treatment of the disease can be ensured. In various situations, such mined knowledge can even save the lives.

The most important aspect in cancer disease is its timely diagnosis and decreasing the risk of disease and this can be done with the help of data mining techniques. The techniques are effective for identifying hidden cancer aggregation pattern and for classification of familiar risk by the help of Neural Network and SIR. The ensemble classifier provides better accuracy in many cases as compared to other techniques.

By applying data mining techniques like Rank Based method in which reversal pairs are real, it can be easily transported to independent cohorts of samples by the help of which cancer diagnosis becomes easy. There are many more techniques which are used and discussed by the authors in different papers. As mentioned in Table 1, which lists some other techniques like clustering which is found to be useful in improving the treatment effectiveness. Various techniques such as Fuzzy logic, Decision Tree, Logistic Regression, *etc.* have been extensively used in prediction of cancer disease related patterns. The techniques ANN, Fuzzy based classifiers and logistic regression gives better accuracy than other traditional models, thus they are better models for health physician in predicting disease and taking preventive measures.

We can conclude that by applying different data mining techniques timely diagnosis and prventive measures can be taken which reduce the fatality of disease and if proper management in application is used it can bring a revolution in health care services.

## References

[1]  D. Tomar and S. Agarwal, "A survey on Data Mining approaches for Healthcare", International Journal of Bio-Science and Bio-Technology, vol. 5, no. 5, **(2013)**, pp. 241-266.
[2]  D. Tomar and S. Agarwal, "Hybrid feature selection based weighted least squares twin support vector machine approach for diagnosing breast cancer, hepatitis, and diabetes", Advances in Artificial Neural Systems, vol. 2015, no. 1.
[3]  B. R. Prasad and S. Agarwal, "Modeling risk prediction of diabetes - A preventive measure", 9th International Conference on Industrial and Information Systems (ICIIS), Gwalior, IEEE, **(2014)**, pp. 1-6.

[4]  D. Tomar, B. R. Prasad and S. Agarwal, "An efficient Parkinson disease diagnosis system based on Least Squares Twin Support Vector Machine and Particle Swarm Optimization", 9th International Conference on Industrial and Information Systems (ICIIS), Gwalior, IEEE, **(2014)**, pp. 1-6.

[5]  A. K. Yadav, D. Tomar and S. Agarwal, "Clustering of lung cancer data using Foggy K-means", 2013 International Conference on Recent Trends in Information Technology (ICRTIT), IEEE, **(2013)**.

[6]  WHO. http://www.who.int/mediacentre/factsheets/fs297/en/ Retrieved on May 20, **(2016)**.

[7]  K. Balachandran and R. Anitha, "Ensemble based optimal classification model for pre-diagnosis of lung cancer", 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), IEEE, **(2013)**.

[8]  H. Li, G. Hong and Z. Guo, "Reversal DNA methylation patterns for cancer diagnosis", 2014 8th International Conference on Systems Biology (ISB), IEEE, **(2014)**.

[9]  K. Shiny, "Implementation of Data Mining Algorithm to Analysis Breast Cancer", International Journal for Innovative Research in Science and Technology, vol. 1, no. 9, **(2015)**, pp. 207-212.

[10] R. Chau, "Determining the familial risk distribution of colorectal cancer: a data mining approach", Familial cancer, **(2015)**, pp. 1-11.

[11] N. Rathore, D. Tomar and S. Agarwal, "Predicting the survivability of breast cancer patients using ensemble approach", 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), IEEE, **(2014)**.

[12] S. S. Shrivastava, V. K. Choubey and A. Sant, "Classification Based Pattern Analysis on the Medical Data in Health Care Environment", International Journal of Scientific Research in Science, Engineering and Technology, vol. 2, no. 1, **(2016)**.

[13] R. Vidhu and S. Kiruthika, "A New Feature Selection Method for Oral Cancer Using Data Mining Techniques", International Journal of Advanced Research in Computer and Communication Engineering, vol. 5, no. 1, **(2016)**.

[14] R. Nagpal and R. Shrivastava, "Cancer Classification Using Elitism PSO Based Lezy IBK on Gene Expression Data", Journal of Scientific and Technical Advancements, vol. 1, no. 4, **(2015)**, pp. 19-23.

[15] F. Mohammadzadeh, "Predicting the probability of mortality of gastric cancer patients using decision tree", Irish Journal of Medical Science, vol. 4, no. 2, **(2015)**, pp. 277-284.

[16] M. Kumar, S. S. Tomar and B. Gaur, "Mining based Optimization for Breast Cancer Analysis: A Review", International Journal of Computer Applications, vol. 19, no. 13, **(2015)**.

[17] M. Duraira and R. Deepika, "Prediction of Acute Myeloid Leukemia Cancer Using Datamining-A Survey", International Journal of Emerging Technology and Innovative Engineering, vo1. 2, **(2015)**, pp. 94-98.

[18] P. J. García-Laencina, "Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values", Computers in biology and medicine, vol. 59, no. 1, **(2015)**, pp. 125-133.

[19] T. Christopher, "A Study on Mining Lung Cancer Data for Increasing or Decreasing Disease Prediction Value by Using Ant Colony Optimization Techniques", Proceedings of the UGC Sponsored National Conference on Advanced Networking and Applications, **(2015)**.

[20] K. Arutchelvan and R. Periasamy, "Analysis of Cancer Detection System Using Datamining Approach", International Journal of Innovative Research in Advanced Engineering, vol. 2, no. 11, **(2015)**.

[21] W. Tseng, "The Application of Data Mining Techniques to Oral Cancer Prognosis", Journal of medical systems, vol. 39, no. 5, **(2015)**, pp. 1-7.

[22] A. N., Noura, "Data mining approaches for predicting demand for healthcare services in Abu Dhabi", 2014 10th International Conference on Innovations in Information Technology (INNOVATIONS), IEEE, **(2014)**.

## Authors

**Satyam Shukla**, is an M. Tech. scholar at Kamla Nehru Institute of Technology (KNIT) Sultanpur (U.P.), in Computer Science & Engineering department under the supervision of Prof. Dharmendra Lal Gupta. His is an active researcher in the field of data mining and its applications in healthcare domain and other domains.

**Dharmendra Lal Gupta**, is currently working as an Associate Professor in the Department of Computer Science & Engineering at KNIT, Sultanpur (U.P.) India. He received B. Tech. (1999) from Kamla Nehru Institute of Technology (KNIT) Sultanpur (U.P.), in Computer Science & Engineering, M. Tech. Hon's (2003) in Digital Electronics and Systems from Kamla Nehru Institute of Technology (KNIT) Sultanpur (U.P.) India. He has been a member of IEEE Computer Society. He has published about 24 papers in International/National Journals/workshops/conferences and seminars. His research interests are Software Quality Engineering, Software Engineering, Cryptography &Network Security.

**Bakshi Rohit Prasad**, is a research scholar in Information Technology Division of Indian Institute of Information Technology, Allahabad. His primary research interests are Data Mining, Machine Learning, Big Data Storage, Computing and Algorithms to deal with related issues. Also, he has significant publications in high speed streaming data mining, analytics and optimizations and their applications in several domains.