IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-6, Issue-8, Aug 2018 E-ISSN: 2347-2693

A Survey on Data Mining Techniques Applied on Cardiovascular Diseases and Cancer, Diagnosis and Prognosis

Deepali Kamath[1*], Anupama Ajith[2], Kavita Pujari[3], Praveena Kumari MK[4]

[1,2,3,4]Dept. of Computer Science, M. G. M.College, Udupi-576102, India

[*]*Corresponding Author: kamath_deepali@yahoo.com, Tel.: +919449388877*

Available online at: www.ijcseonline.org

Accepted: 14/Aug/2018, Published: 31/Aug/2018

***Abstract*—** There has been an exponential growth in the number of cardiovascular diseases and cancer in the present world due to unfavorable environmental factors, faulty food, stress and erroneous lifestyle. These two account for a majority of deaths worldwide. Early detection and prevention plays a remarkable role in preventing deaths. It is not an easy task for medical practitioners to instantly come to a conclusive diagnosis. Hence we can resort to data mining techniques to extract occult, foreseeable information that can be acted upon the large set of medical data. In this survey, we have presented an overview on the symptoms, their aggravating factors in various cardiac illnesses and cancer. We have also enlisted, discussed and analyzed data mining techniques such as Decision Tree, Neural Networks, and Naïve Bayes etc. This paper summarizes various review and technical journals on cardiovascular disease and cancer diagnosis and prognosis.

***Keywords*—** Cardiovascular diseases, Cancer, Diagnosis, Prognosis, Data Mining

I. INTRODUCTION

Heart/Cardiovascular diseases (CVD) and Cancer have accounted for a vast majority of deaths worldwide in the recent past. The causative factors that may be attributed like age, sex, food, living and working conditions, education, quality of water, health care service, environmental and agricultural factors etc are as shown in Figure 1.

Figure1: Causative factors responsible for the worldwide emergence of diseases [1]

CVDs accounted for the largest proportion of NCD (Non Communicable Diseases) deaths (47.9%), second in the list is that by cancers (21%), followed by chronic diseases of respiratory system (11.72%), diseases of the digestive tract (6.1 %), and diabetes (3.5 %). The remaining NCDs were responsible for 9.78% of all deaths. With population explosion, annual NCD deaths have been projected to rise substantially to 52 million by 2030. Annual CVD mortality has been projected to rise to 6 million and that by cancer to 4 million. India is experiencing a fast rise in health evolution with rising NCDs causing major sickness and deaths in urban and rural areas. This has resulted in huge loss in prime life (age 35-64 years). NCDs are estimated to account for about 53 % of all deaths. Figure 2 shows the comparative mortality in the country [2].

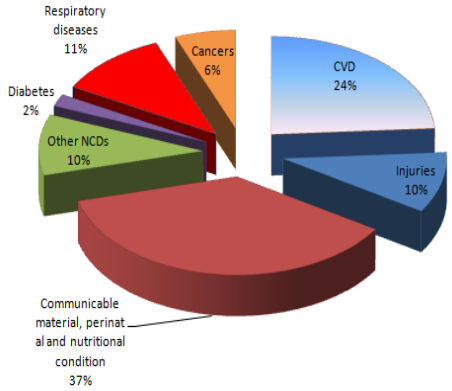

Figure 2: Proportional mortality in India (3 of total deaths, all ages)[2]

Effective treatment and good prognosis are enhanced with early detection. Doctors have been striving hard to find new methods for this purpose. Predicting the outcome of a disease is one of the most interesting and challenging tasks. Most

recently, powerful computers with sophisticated software and automated tools that could store enormous amounts of data on medical history are being scattered on medical research groups. By virtue of the above, Data mining technique in KDD (Knowledge Discovery in Databases) has become a powerful research tool for medical researchers to recognise and exploit patterns and relationships among large number of variables. This has enabled them to predict the outcome of a disease using medical history. In this paper an attempt has been made to carry out a survey on data mining techniques like classification, clustering, association, regression and others to predict and analyse cardiovascular diseases and cancer. The objective of this study is to summarise various reviews and technical articles on diagnosis and prognosis of the diseases in question. It gives an overview of the current researches using data mining techniques to enhance the diagnosis and prognosis.

Rest of the paper is systematized as follows, Section I encloses the introduction of causative factors responsible for the world-wide emergence of diseases and also the proportional mortality in India through figures. Section II contains an overview of cardiovascular diseases, its symptoms and also its causative factors. Section III covers the overview of cancer, its symptoms, signs and causes for cancer. Section IV briefly mentions certain paces taken to compose this paper. Section V explains measures of data mining techniques for diagnosis and prognosis of cardiovascular diseases. Section VI describes reviews and results of cancer from a number of inputs; Section VII concludes research work with future directions.

II. OVERVIEW ON CARDIOVASCULAR DISEASES

CVD encompasses diseases of the heart and the vascular system. . In today's world, CVDs are causes for one-third of all deaths [2].

The important ones are ischemic heart disease (IHD), hypertension, cerebrovascular disease (stroke) and congenital cardiac anomalies. Coronary heart disease (ischaemic heart disease) has been defined as "impairment in heart functions due to inadequate blood flow, caused by obstructive changes in the coronary circulation to the heart". Rheumatic heart disease (RHD) continues to be an important health issue in most developing countries. Acute Rheumatic fever is predominantly a disease in children aged 5 to 14 years and generally does not affect children under age 3 or adults. In India, an estimated 46.9 million patients have been reported with cardiovascular disease during the year 2010. Compared to other countries, India suffers the highest loss in potentially productive years of life, due to deaths from CVDs in people aged 35-64 years. Although there is an increase in prevalence of CHD in rural areas, it is not that steep because life-style changes have affected people in urban areas more than rural [2].

Table 1: Age-specific prevalence rate per 1000 for CHDs derived from the selected studies 2004 [2]

Age group (in years)	Urban		Rural	
	Male	Female	Male	Female
20-24	8.00	6.80	17.54	10.47
25-29	19.65	26.24	13.67	14.42
30-34	17.05	22.96	12.39	10.75
35-39	43.18	48.44	18.79	15.99
40-44	47.25	65.85	17.94	23.23
45-49	83.26	105.35	20.72	38.78
50-54	93.07	111.88	31.11	49.86
55-59	162.44	152.75	26.68	50.91
60+	173.65	175.35	71.07	67.44

Following are the symptoms of heart disease [3]:

1. Pain, pressure, or discomfort in the centre of the chest.
2. Pain, tingling, or discomfort in other parts of the upper body. This might include the arms, back, neck, jaw, or stomach.
3. Shortness of breath.
4. Nausea, vomiting, burping, or heartburn.
5. Sweating or having cold, clammy skin.
6. A fast or uneven heartbeat.
7. Loss of appetite.
8. Dizziness or light headedness; and
9. Headache.

The causative factors for heart diseases are many. A few of them are discussed here [4]:

Coronary artery disease: Coronary arteries become damaged or blocked due to accumulated plaque inside arteries due to cholesterol. This causes heart to obtain less oxygen, proteins and minerals.

Enlarged Heart (Cardiomegaly): Size of the heart increases due to some conditions like congestive heart failure, and long standing anemia. It may also be related to high blood pressure.

Angina pectoris: This is due to sudden diminution in blood supply to the heart muscle. It leads to death of, or harm to a component of muscle.

Heart Valve Disease: The valves of heart are flaps of endocardium (innermost layer of heart) that guard blood flow from the upper to the lower chambers. Failure to close at the appropriate moment leads to mixing up of oxygenated and deoxygenated blood resulting in a wide variety of symptoms like breathlessness and cyanosis(bluish discoloration of skin) etc.

Congential Heart Disease: Failure to develop fully or partially of certain components of the heart during embryological life leads to improper functioning of the heart in later life.

III. CANCER OVERVIEW

Cancer is a disease in which cells grow abnormally anywhere in the body by degrading the normal rules of cell division. These abnormal cells are termed cancer cells, malignant cells, or tumor cells [5]. If this harmful growth is allowed to continue, it will eventually interfere with the physiological functions of the body's vital organs causing death.

Early diagnosis of cancer may cure it or increase the survival rate of the patient. The treatment and prognosis may differ based on the type of cancer, its stage, and age of the patient. Treatment may include surgery, chemotherapy, and radiation. Approximately 10 million new cancer cases are diagnosed every year worldwide. It may double by the year 2020 [6].

Symptoms and signs of cancer [7]:

- Unexplained weight loss, chronic fatigue is the symptoms often accompanied by rapidly progressing cancers.
- A lump or hard area in the breast, swelling or bloody discharge from the nipples and changes in its shape or texture prompt further investigations.
- A change in a wart or mole may be melanoma or squamous carcinoma.
- A persistent change in digestive or bladder habits such as constipation, chronic diarrhea, pain in abdomen, rectal or urinary bleeding, nausea, vomiting indicate gastrointestinal cancer.
- A persistent cough or coughing up blood may indicate lung cancer.
- Unusual bleeding or vaginal discharge may indicate presence of uterine, endometrial or cervical cancer.

Causes of cancer [2]:

1. ENVIRONMENTAL FACTORS: They play a major role in most cancers.
   Tobacco: Tobacco chewing and smoking may affect lungs, larynx, mouth, pharynx, and even lungs.
   Alcohol: Excessive intake of alcohol may lead to liver cancer.
   Parasites: Parasitic infestations may also increase cancer risks.
   Occupational Exposures: These include benzene, arsenic, cadmium, chromium, and vinyl chloride etc.
   Others: Factors like sunlight, radiation, air and water pollution, medications (e.g., hormonal treatments) and pesticides may also be attributed to cancer.
2. **Genetic factors:** Genetic influences have long been suspected. Some people are born with a gene mutation that they inherited from their mother or father. This damaged gene puts them at higher risk for cancer than most people. For example, retinoblastoma occurs in children of the same parents.

IV. METHODOLOGY

Due to resource constraints and nature of the paper itself, the main methodology applied in this paper was through survey of journals and publications in the fields of medicine, computer science and engineering. The research focused on more recent publications.

V. DATA MINING TECHNIQUES FOR DIAGNOSIS AND PROGNOSIS OF CARDIOVASCULAR DISEASES

This section consists of reviews from various research papers and articles on data mining techniques applied in diagnosis and prognosis of cardiovascular diseases.

Three data mining techniques namely, Decision tress, Naïve bayes and Neural Networks were proposed by Sellappan Palaniappan, Rafiah Awang to develop a prototype called "Intelligent Heart Disease Prediction System (IHDPS)". From a large database on history of heart diseases, IHDPS can find out and mine hidden knowledge. The system assists healthcare practitioners make quick and smart decisions based on hidden patterns of data (knowledge) which cannot be carried out by traditional methods. Result of the analysis is visualized and interpreted in tabular and graphical forms. By providing efficient treatments, it aids in reducing treatment costs. It consists of six stages namely: business understanding, data understanding, data preparation, modelling, evaluation, and deployment [8].

Neural network approach for heart disease prediction was proposed in [9]. A list of experiments is conducted on a sample database of 78 patients', input variables 13 are used for training and analysis. He recommended supervised network and training using Black propagation algorithm. From this trained data set, the system can discover unknown data and can predict the imminent diseases. Neural networks facilitate in testing and can be used by professionals to ascertain diagnosis.

In this paper [10], a method is recommended in which heart disease dataset is analyzed on selected user symptoms to pick patients at risk. The algorithm stays away from useless item sets that do not suit support value thus generating efficient item set and saving time by avoiding unwanted comparison. The training dataset contains 1000 patients' records, with 19 clinical features of patients with heart disease.

This [11] describes Data mining techniques that are proven accurate in predicting diagnosis of heart diseases. STATISTICA Data Miner (SDM) provides Association Rules to carry out research work. Three predefined threshold values are defined viz. Minimum Support, Minimum

Confidence and Minimum Correlation. A priori algorithm is applied on Medical Data to discover frequent item sets using candidate generation, which in turn produce Association rules to identify key factors behind disease. The major risk factors for heart disease were found for both men and women using rule based analysis carried out by classifying data based on gender.

Supervised machine learning algorithms are used for data classification in [12]. This paper discusses K-mean, Nearest Neighbour, and Entropy Based Mean Clustering algorithm. The experiment is undertaken based on training data sets of 14 different attributes. The Tanagra software is employed to implement the above said algorithms. This software is based on Mean clustering algorithm and is ideal for processing datasets and to predict accuracy in finding risk factors.

In the study [14], three features of vessel stenosis have been employed namely Left Anterior Descending (LAD), Left Circumflex (LCX) and Right Coronary Artery (RCA). The ones whom LAD, LCX or RCA vessel is clogged are classified as CAD patients, others as healthy. Table 6 summarises the results obtained from different classification algorithms for diagnosis of CAD specifically.

In [14], neural networks algorithm was used for prediction of stenosis of each vessel separately. A multi-layered perceptron neural network was employed for the classification. The accuracy reached 73%, 64.85% and 69.39% for LAD, LCX and RCA vessels, respectively. Another study was by applying Navïe Bayes, C4.5, and KNN classification algorithms. The best accuracy was obtained by C4.5 wherein achieved accuracy was 74.20%, 63.76%, and 68.33%, respectively.

Table 7 shows the results obtained from different studies to diagnose CVD and CHD diseases. Initially, CVD and CHD were predicted using three different supervised machine learning algorithms namely Navïe Bayes, KNN, decision list techniques. Navie Bayes algorithm showed slightly better performance with accuracy of 52.33%.
Moreover, CVD predictions were further analysed using RIPPER classifier, decision tree, artificial neural network, and SVM. Their analysis shows that, SVM predicts cardiovascular disease with the highest accuracy of 84.12%. The study then applied Navïe Bayes, decision tree, and classification via clustering out of which decision tree obtained the best accuracy of 99.2%. Another study used Navïe Bayes and decision trees to detect CVD and achieved 62.03%, and 60.40% of accuracy, respectively. A recent study used a hybrid genetic neural network method to detect CVD and achieved 89% accuracy. A study applied on C4.5, multilayer neural network (MLP), Bayesian classifier, and SVM, where MLP obtained the highest accuracy (89.7%) compared with the others. Another three prediction models for CHD were implemented namely C5 classifier, MLP and SVM. The SVM was the best predictor gaining an accuracy of 92.1%.

Table 6: Accuracy of Neural Network, Naïve Bayes, C4.5 and KNN Classification algorithms for diagnosis of coronary arteries disease(CAD) [13]

Data Mining Techniques	Accuracy (LAD)	Accuracy (LCX)	Accuracy (RCA)
Neural Network	73%	**64.85%**	**69.39%**
Naïve Bayes	51.81%	62.73%	67.29%
C4.5	**74.20%**	63.76%	68.33%
KNN	59.65%	61.39%	59.11%

Table 7: Accuracy of Decision Tree, Naïve Bayes and Classification via clustering algorithms for diagnosis of cardiovascular disease (CVD) [13]

Data Mining Techniques	Accuracy
Naïve Bayes	52.33%
KNN	52%
Decision List	45.67%
RIPPER Classifier	81.08%
Decision Tree	79.05%
Artificial Neural Network	80.06%
SVM	84.12%
Naïve Bayes	96.5%
Decision Tree	99.2%
Classification via Clustring	88.3%
Naïve Bayes	62.03%
Decision Tree	60.40%
Hybrid Genetic Neural Network	89%
C4.5	82.5%
Multilayer Neural Network (MLP)	89.7%
Bayesian Classifier	82%
SVM	82.5%
C5 Classifier	89.6%
MLP	91.0%
SVM	92.1%

VI. DATA MINING TECHNIQUES FOR DIAGNOSIS AND PROGNOSIS OF CANCER

This section consists of reviews of various research papers and articles on data mining techniques applied in breast cancer, oral cancer, lung cancer, skin cancer and brain tumors, its diagnosis and prognosis.

Breast cancer

Breast cancer is the most common disease for female mortality in India. Breast cancer threat in India shows that 1 in 28 women develop breast cancer during her life time. The rate is higher in urban women being 1 in 22 in her lifetime compared to rural where this risk is relatively much lower (1 in 60 women). In India the average age of the high risk group is 43-46 years unlike in the west where women aged 53-57 years are more prone to breast cancer [14].

In the study [15], multiple prediction models were developed and compared for breast cancer survivability using

a large dataset. They used two popular algorithms, artificial neural networks and decision trees along with logistic regression. The comparison results in this paper indicate decision tree (C5) is the best predictor with 93.6% accuracy, artificial neural networks is the second with 91.2% and the logistic regression model with 89.2% accuracy.

This study [16] has presented an analysis of the prediction of survivability rate of breast cancer patients using data mining techniques. The data used by them is the SEER Public-Use Data. The pre-processed data set consists of 151,886 records, which have all the available 16 fields from the SEER database. They have investigated three data mining techniques: the Naïve Bayes, the back-propagated neural network, and the C4.5 decision tree algorithms. Several experiments were conducted by them using these algorithms which show that C4.5 algorithm has a much better performance.

In [17], they have presented an automatic diagnosis system for detecting breast cancer based on association rules (AR) and neural network (NN). They have used AR for reducing the dimension of breast cancer database and NN is used for intelligent classification. A 3-fold cross validation method was applied to the Wisconsin breast cancer database to evaluate the proposed AR + NN system performances. This research demonstrated that the AR + NN model can be used to obtain fast automatic diagnostic systems for other diseases.

Oral cancer

Globally, about 575000 new cases and 320000 deaths occur every year from oral cancer [18]. The symptoms for an oral cancer at an earlier stage are [19]: 1) Patches inside the mouth or on lips that are white, red or a mixture of white and red. 2) Bleeding in the mouth. 3) Difficulty or pain while swallowing; and 4) A lump in the neck. Treatments for Oral Cancer include surgery, radiation, and chemotherapy.

In this [20], the clustering and classification features were compared to determine the differences. Two data mining tools, namely decision tree and artificial neural network, were used and compared with the performance of logistic regression. The comparative results show that, the trees created by the decision tree models are relatively easier to interpret compared to that of the artificial neural network models.

In another study [21], it was compiled and enlarged the oral cancer gene database to include 374 genes, by adding 132 gene entries to enable fast retrieval of updated information.

A study [22], has predicted oral cancer survivability using classification algorithms like CART, Random Forest, LMT, and Naïve Bayesian. The algorithms classify the cancer survival using 10 fold cross validation and training data set. The Random Forest technique correctly classified the cancer survival data set.

Lung cancer

Smoking is the major cause for lung cancer whose early diagnosis is vital in treatment success. Risk factors such as smoking, alcohol consumption, obesity etc had statistically played an important role in pre-diagnosis stage [23].

In [23], they have examined the potential use of classification based data mining techniques such as Rule based, Decision tree, Naïve Bayes and Artificial Neural Network. For data pre-processing and effective decision making One Dependency Augmented Naïve Bayes classifier (ODANB) and naive creedal classifier 2 (NCC2) are used. The most effective model to predict patients with Lung cancer disease appears to be Naïve Bayes followed by IF-THEN rule, Neural Network and Decision Trees. They found overall only 14% of people diagnosed with lung cancer survive five years after the diagnosis.

In [24], data mining classification, clustering, association techniques were used to predict diseases in humans. They have summarized the medical data mining techniques used in various diseases and have mentioned that lung cancer can be predicted using distinguished disease sub types with the aid of Ensemble approach.

Skin cancer

Skin cancer is the abnormal growth of skin cells. Approximately 65% to 90% of them are caused by UV radiation [25]. Skin cancers are curable if treated early.

In paper [26], they have considered risk factors to collect 200 people's data from different diagnostic centers. After pre-processing, data is clustered using K-means clustering algorithm for separating relevant and non-relevant data and then MAximal Frequent Itemset Algorithm (MAFIA) is used to mine the frequent patterns. MAFIA algorithm is treated as better algorithm than others because it combines diverse old and new algorithmic ideas to form a practical algorithm [27] [28]. Finally they implemented a system using Lotus Notes to predict Skin Cancer risk level with suggestions which are easier, cost reducible and time savable.

In study [24], have summarized the medical data mining techniques used for various diseases among which data mining techniques used in terms of prediction and decision making of skin diseases is Categorization of skin disease using integrated tree model with neural network classification methods.

Brain tumors

Brain tumors are considered as the most lethal, difficult to identify and treat. Pathologists evaluate the aggressiveness of brain tumors by visually examining tissue sections (biopsies) based on guidelines determined by the World Health Organization (WHO). According to the WHO grading system, the appearance of certain histopathological features, such as cellularity, pleomorphism, mitosis, necrosis, vascular proliferation, and apoptosis, classify tumors on the basis of

their aggressiveness as low or high-grade tumors. Low-grade tumors are less insistent and are associated generally with good prognosis. High-grade tumors are more aggressive, and are characterized by rapid growth with tendency to invade nearby tissues [29].

The paper [30], presented an automated data mining system that allows public health decision makers to access analytical information on brain tumors. The emphasis in their study is laid on use of ontology in an automated data mining process.

The study [31], has proposed a method for characterizing brain tumors, which models the human thinking approach. The classified results are compared with other computational intelligent techniques. The novelty of the method is based on the fuzzy cognitive maps (FCMs) to represent and model experts' knowledge (experience, expertise, heuristic). Main advantage of the proposed FCM grading model is sufficient interpretability and transparency in decision process, which makes it a convenient consulting tool in characterizing tumor aggressiveness for everyday clinical practice.

VII. CONCLUSION

Cardiovascular diseases and cancer are the leading causes for death across the world. In this survey we have studied different technical and review papers and found that Data mining techniques help to provide effective treatment, early diagnosis, and easy prognosis at the right moment.

Heart disease is the leading cause of death in the worldwide. This paper reviews state-of-the-art by data mining techniques applied in diagnosing three heart diseases, namely CAD, CVD, and CHD. Among them CAD can be diagnosed by understanding stenosis of blood vessels. Such disease is data rich but unfortunately the obtained accuracy from CAD classifiers is poor. It is found that decision trees and Naïve Bayes classifiers are recommended for CVD diagnosis with an accuracy reaching more than 95%. Further, C5, SVM, and neural networks are the best recommended classifiers for CHD prediction. The empirical results show that we can produce short but accurate prediction list in heart patients by applying predictive models on the medical records of newer patients.

In breast cancer prediction, C4.5 algorithm gives a much better performance. Decision tree (C5) plays the best predictor with 93.6% accuracy. AR + NN model can be adopted for fast automatic diagnostic systems in detecting breast cancers. Among the Classification algorithms, Random Forest techniques present with less errors in predicting oral cancer. In Lung cancer prediction, the most effective model appears to be Naïve Bayes. MAFIA algorithm is found efficient and easier for collecting patterns in skin cancer detection and this system is implemented with Lotus Notes. The study reveals that use of ontology in an automated data mining process gives useful information about the treatment of brain tumors. We may conclude that India is lagging behind in research collaborations in spite of its less stringent privacy norms. Data mining provides many benefits to the healthcare industry in every instance in that it benefits doctors, the management, hospital staff and organizations, as well as it provides effective treatment and diagnosis in minimum time, plus the healthcare services become cheaper and manipulations are kept at bay.

REFERENCES

[1] D. Tomar, S. Agarwal, "A survey on Data Mining approaches for Healthcare", International Journal of Bio-Science and Bio-Technology, Vol 5(5),pp.241-266, 2013.

[2] K. Park , "Park's Textbook of PREVENTIVE AND SOCIAL MEDICINE", 23rd Edition, Bhanot publishers, India, pp 363-385,2015.

[3] Colin Ogilvie, Christopher C. Evans, "An Introduction to Medical Diagnosis", 11th Edition,Butterworth-Heinemann, Oxford, pp 201-205.

[4] S.Vijayarani, S.Sudha, "A Study of Heart Disease Prediction in Data Mining" International Journal of Computer Science and Information Technology & Security ", (IJCSITS), ISSN: 2249-9555 Vol. 2, No.5, October 2012 .

[5] M. Hejmadi, "Introduction to Cancer Biology",ebook 2nd Edition, Bookboon publishing, pp- 7, 2010, ISBN 978-87-7681-478-6

[6] J.F. McCARTHY, K.A. Marx, P.E.Hoffman, A.G. GEE, P O'neil, M.L.Ujwal, and J Hotchkiss, "Applications of Machine Learning and High-Dimensional Visualization in Cancer Detection, Diagnosis, and Management." Annals of the New York Academy of Sciences1020, no. 1 (2004): pp. 239-262.

[7] Mahajan & Gupta, "Textbook of PREVENTIVE AND SOCIAL MEDICINE",4th Edition,Jaypee brothers medical publishers, India, PP 359,2013, ISBN 978-93-5090-187-8 978-

[8] S. Palaniappan, R. Awang, "Intelligent heart disease prediction system using data mining techniques", IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.8, 343-349,August 2008

[9] N. Guru, A. Dahiya, N. Rajpal, "Decision Support System for Heart Disease Diagnosis Using Neural Network", Delhi Business Review, Vol. 8, No. 1 (January - June 2007).

[10] Ramyasri M M (14CSR151), Renuka P (14CSR158) and Rajeshkumar R (14CSL275) and Kumaravel T, "Prediction of Heart Diseases Risk through Frequent Itemsets in Data Mining "3 rd National Conference on Innovative Research Trends in Computer Science and Technology (NCIRCST) ISSN: 2454-4248 Volume: 4 Issue: 3 115 – 120, 2018

[11] I. U. Said , A. H. Adam , Dr. A. B. Garko, "ASSOCIATION RULE MINING ON MEDICAL DATA TO PREDICT HEART DISEASE", International Journal Science Technology and Managaement Vol.No.4, Issue 08, Aug 2015

[12] D. C.Sekar,& K. R. H.Rao, Predicting the heart attack symptoms using biomedical data mining techniques. International Journal of Computer Science & Applications (TIJCSA), 1(3),(2012).

[13] S. M. Alzahani, A. Althopity, A. Alghamdi, B. Alshehri, & S. Aljuaid, (2014). An overview of data mining techniques applied for heart disease diagnosis and prediction. Lecture Notes on Information Theory Vol, 2(4),December 2014

[14] V. Chaurasia , S. Pal, "A Novel Approach For Breast Cancer Detection Using Data Mining Techniques" , International Journal Of Innovative Research in Computer and Communication

Engineering (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 1, 2014
[15] D. Delen, G. Walker, A. Kadam, "Predicting Breast Cancer Survivability: A Comparison Of Three Data Mining Methods", Elsevier: Artificial Intelligence in Medicine , Vol. 34, Issue 2, pp: 113-127, 2005
[16] A Bellaachia & E. Guven, "Predicting Breast Cancer Survivability Using Data Mining Techniques", Age, Vol. 58, Issue 13, pp: 10-110, 2006
[17] M. Karabatak & M. C. Ince, "An Expert System For Detection Of Breast Cancer Based On Association Rules And Neural Network", Elsevier: Expert Systems With Applications, Vol. 36, Issue 2, Part 2, pp: 3465-3469, 2009
[18] N. Sharma and H. Om, "Framework For Early Detection And Prevention Of Oral Cancer Using Data Mining", International Journal of Advances in Engineering & Technology, Vol. 4, Issue 2, pp: 302-310, 2012.
[19] C. Scully, J.V. Bagan, C. Hopper, J.B. Epstein, "Oral Cancer: Current And Future DiagnosticsTechniques – A Review Article", American Journal of Dentistry, Vol. 21, Issue 4, pp: 199-209, 2008
[20] W. T. Tseng, W. F. Chiang, S. Y. Liu, J. Roan, & C. N. Lin, "The Application Of DataMining Techniques To Oral Cancer Prognosis", Journal Of Medical Systems, Vol.39, Issue 5, pp: 59,2015
[21] Gadewal & Zingde, "Database And Interaction Network Of Genes Involved In Oral Cancer: Version II", Online Journal of Bioinformatics, Vol. 6, Issue 4, pp:169-170, 2011
[22] D. Kaladhar, B. Chandana and P. B. Kumar, "Predicting Cancer Survivability Using Classification Algorithms", International Journal Of Research And Reviews In Computer Science (IJRRCS), Vol 2, Issue 2, pp: 340 – 343, 2011
[23] V.Krishnaiah, Dr.G.Narsimha, Dr.N.Subhash Chandra,"Diagnosis of lung cancer prediction system using data mining classification techniques". International Journal of Computer Science and Information Technologies, Vol. 4, Issue 1, pp.39-45, 2013.
[24] S. Patel and H. Patel, "Survey of data Mining techniques used in healthcare Domain", International Journal of Information Sciences and Techniques, Vol. 6,pp.53-60 , 2016.
[25] National Library of Australia Cataloguing–in–Publication data: Lifestyle and cancer: knowledge, attitudes and behavior in NSW 2009 SHPN (CI) 120203, Published by the Cancer Institute NSW, pp.1-29, 2012, ISBN 978-1-74187-760-1
[26] K. Ahmed, T.Jesmin, & M.Z. Rahman,"Early prevention and detection of skin cancer risk using data mining", International Journal of Computer Applications, Vol. 62, Issue 4,pp.1-6, 2013.
[27] D. Burdick, M. Calimlim and J. Gehrke, "MAFIA: A Maximal Frequent Itemset Algorithm for Transactional Databases", Proceedings of the 17th International Conference on Data Engineering, pp.443-452, April 02-06, 2001.
[28] D. Burdick, M. Calimlim, J. Flannick, J. Gehrke and T. Yiu, MAFIA: "A Performance Study of Mining Maximal Frequent Itemsets", Proceedings of the 17th International Conference on Data Engineering, pp.443-452, April 02-06, 2001.
[29] E.I.Papageorgiou, P.P.Spyridonos, D. Th.Glotsos, C.D.Stylios, P.Ravazoula, G.N.Nikiforidis, P.P.Groumpo, "Brain Tumor Characterization Using The Soft Computing Technique Of Fuzzy Cognitive Maps, Elsevier: Applied Soft Computing, Vol. 8, Issue 1, pp: 820-828, 2008.
[30] R.S.Santos, S.M.F.Malheiros, S.Cavalheiro, J.M.P. Oliveira," A Data Mining System For Providing Analytical Information On Brain Tumors To Public Health Decision Makers, Elsevier: Computer Methods And Programs In Biomedicine, Vol. 109, Issue 3, pp: 269-282, 2013
[31] E. I. Papageorgiou, P. P. Spyridonos, D. T. Glotsos, C. D. Stylios, P. Ravazoula, G. N. Nikiforidis, & P. P. Groumpos, Brain Tumor Characterization Using The Soft Computing Technique Of Fuzzy Cognitive Maps, Elsevier:Applied Soft Computing, Vol. 8, Issue 1, pp: 820-828, 2008

Authors Profile

Mrs. Deepali Kamath pursued her Masters degree in Computer Applications from Indira Gandhi National Open University Delhi in the year 2004. Currently she is an Assistant Professor in the Department of Computer Science, Mahatma Gandhi Memorial College, Udupi, which comes under University of Mangalore, KonajeShe has presented 2 research papers in reputed state level conferences. She has 13 years of teaching experience.

Mrs. Anupama Ajith is a post graduate, with the degree MCA. She is working as an Assistant professor in Mahatma Gandhi Memorial College, Udupi, Karnataka, India. She has submitted an UGC sponsored minor research project on the topic "Security Issues in Distributed Processing" and has published a paper in the International journals. Also to her credit, she has a paper in national conference.

Mrs. Kavita Pujari born in Udupi, Karnataka, India on 3rd August 1986 received her BCA and M.Sc degrees in Computer Science from Karnataka University Dharwad, in 2007 and 2010, respectively. She cleared the UGC NET Exam in 2012 and has 7 years of teaching experience. She has attended few conferences and presented papers at the National and International levels. Currently she is working as an Assistant Professor in the Department of Computer Science at Mahatma Gandhi Memorial College, Udupi, Karnataka, India.

Mrs. Praveena Kumari MK, was born in Moodbidri, Karnataka, India on 19 December 1980. She has completed her Masters in Computer Applications from Mangalore University, India in 2004, M.Phil in Computer Science from Alagappa University, India in 2012. To her credit she has papers in international and national conferences. With 15 years of experience in teaching, she is an Assistant Professor in Department of Computer Science at Mahatma Gandhi Memorial College, Udupi, Karnataka, India.

