Current Journal of Applied Science and Technology

39(6): 11-32, 2020; Article no.CJAST.55851
ISSN: 2457-1024
(Past name: British Journal of Applied Science & Technology, Past ISSN: 2231-0843, NLM ID: 101664541)

Data Mining in Cancer Diagnosis and Prediction: Review about Latest Ten Years

Zahraa Naser Shah Weli[1*]

[1]Al-Nahrain University, Baghdad, Iraq.

Author's contribution

The sole author designed, analyzed and interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/CJAST/2020/v39i630555
Editor(s):
(1) Prof David Coman, The Wesley Hospital, The Lady Cilento Children's Hospital, Australia.
Reviewers:
(1) Elisa Rizzi Health, ASLAT, Italy.
(2) Davide Calandra, University of Turin, Italy.
(3) Awodutire Phillip Oluwatobi, University of Africa, Nigeria.
Complete Peer review History: http://www.sdiarticle4.com/review-history/55851

Review Article

Received 02 February 2020
Accepted 07 April 2020
Published 11 April 2020

ABSTRACT

Data Mining [DM] has exceptional and prodigious potential for examining and analyzing the vague data of the medical domain. Where these data are used in clinical prognosis and diagnosis. Nevertheless, the unprocessed medical data are widely scattered, diverse in nature, and voluminous. These data should be accumulated in a sorted out structure. DM innovation and creativity give a customer a situated way to deal with new fashioned and hidden patterns in the data. The advantages of using DM in medical approach are unbounded and it has abundant applications, the most important: it leads to better medical treatment with a lower cost. Consequently, DM algorithms have the main usage in cancer detection and treatment through providing a learning rich environment which can help to improve the quality of clinical decisions. Multi researches are published about the using of DM in different destinations in the medical field. This paper provides an elaborated study about utilization of DM in cancer prediction and classifying, in addition to the main features and challenges in these researches are introduced in this paper for helping apprentice and youthful scientists and showing for them the key principle issues that are still exist around there.

Keywords: Data mining; machine learning; cancer; classification; diagnosis; prediction.

**Corresponding author: E-mail: stcs-zns16@sc.nahrainuniv.edu.iq, zah2005muh@gmail.com;*

1. INTRODUCTION

Early and precise diagnosis of any disease are core processes in the medical field, especially for fast spreading diseases like cancer. Disease detection is significant and complicated, but nowadays it becomes easier because it is automated.

The automated medical system would improve medical decision and lessen the cost [1]. Cancer is one of such a disease where a precise and accurate determination can diminish the death rate in afflicted patients, especially cancer for being the main source of death in developed nations and the subsequent active driving reason for death in developing nations [2]. Today the massive medicinal data is broad, starting from the symptoms of patients with different diseases and the approaches to help them with the prediction of those diseases, besides information about electronic patient: Records, pictures, datasets, signals, wavelengths, hospitals and so on.

Information extracting and data analyzing by one individual are difficult and sophisticated [complex] mission. In this way, the requirement for a mechanized system to deal with these tremendous databases of medical data is to find valuable examples and pattern learning for them [3]. DM, as the provider of this mechanized system, aids many medicinal advances, especially in the scope of disease diagnosis and prediction and acquiring significant relations among elements in the data [4].

This paper summarizes all of the research that has attempted to predict cancer using data mining algorithms and how well these methods are. This makes it easier for researchers in the coming days to update these algorithms or combine them to achieve better results.

2. OVERVIEW OF DATA MINING

Data mining is a combination of algorithmic systems to extract patterns from data. The colossal data is fundamental to be approached and examined for learning extraction that participates in support for understanding the overall conditions in health care approach. Generally, data mining algorithms are classified in two classes:

- Unsupervised learning [or descriptive model]: engages the relationship between attributes and discover patterns in the data.
- Supervised learning [or predictive model]: foretell the future results from existing conduct [5].

DM procedure integrates [confining a theory → gathering information→ performing pre-handling → assessing the model → understanding the model and make the determination. Dhanya and Tintu [6] Relying on methods and techniques from the approximation of database management, machine learning and statistics. Dealing in DM has devoted their vocations to put better procedures that reach conclusions from the enormous amount of data. Those methods and instruments can help in predicting and prognosis of diseases, classifying, and pulling out medical information.

These techniques are: classification, clustering, regression, association rule, neural network [7], Relationship Mining, Outlier Detections, Text Mining, Social Network Analysis [SNA] [8] and Trend Analysis [9].

2.1 Classification Algorithms

The part of classification algorithms is finding the relationship between attributes in the data and get high accuracy of the data is the main objective of it [10].

Classification algorithms in DM arranging the data into multi classes, where each example has a definite place with a class, which is determined by the value of goal attribute where each of goal attribute is matched to a class [11]. Each example in the database is composed of predictor attribute value and a goal attribute value. A classification procedure condensed in two steps: define the class label for training data which is called a supervised learning then evaluate the classification accuracy in the second step [12].

2.2 Clustering

Clustering algorithms characterize if the object belongs to one of the clusters or outlaw it. Clustering algorithms identified similar objects in the cluster and more compatible to it than other cluster [13]. Clustering techniques have two processes:

- Logical process: depending on the similarities between objects, where objects within a cluster are similar to one another and dissimilar to objects in other clusters.

- Appeal process: depending on the diameter, the most terminal splitting between any two objects in the clusters, in other word: measure the distance instead of defining a similarity measure between the objects under consideration [14].

The difference between clustering and classification is the predefined classes exist in the classification algorithms.

2.3 Regression

In reality regression procedure can be modified for prediction. Regression analysis can be used to manifest the connection between one or more dependent and independent variables. Independent variables are attributes already known and dependent variables are foreseeing them in the future. Reality, many world problems is not just forecasted, like: stock prices, sales volumes and outcome failure rates are hard to forecast because they may depend on sophisticated interactions of multiple predictor variables. Hence, need regression techniques to predict future values.

Regression algorithms have various implementations in time series prediction, business planning, trend analysis, biomedical and drug response, modelling, marketing, environmental modelling and financial prediction [15].

2.4 Association Rule

Association rules are used to figure out rationality relations between variables. The name comes from finding all the rules existing in database that accomplish some minimum support and minimum confidence restrictions. This type of detection causes organizations to settle down on specific selections, such as cross marketing, catalogue design, and customer shopping conduct analysis.

Association rules have been used in many applications in healthcare. IF-THEN rules are the main rules in association rules. However, the resulting rules were diverse. Consequently, rules are selected by the objective and submerge into consideration that are selected rules are strong rules that must have a value more than definite minimum support and minimum confidence. The principal predictive analysis technique is the classification using association rule which targets to find out a little set of rule in the database that structure an exact classifier [16].

2.5 Neural Network

The neural network is connectionists learning composed of a set of input/output units related with arcs where each arc has a weight associated with it. These weights adjusting during the learning phase to predict the right class label of the input tuples. Hence, neural networks have long training terms. The main objective for neural networks is the poor explain ability such as and it is difficult to interpret the symbolic meaning and the learned weights of hidden units.

Neural Networks have plenty of interesting features such as the ability to classify examples not trained ago and high endurance of noisy data. Neural networks are very useful when few information of the relationship between attributes have. They are appropriate for continuous-valued inputs and outputs [17]. Neural Network has little structures as follows:

- Multi-Layer Neural Network [MLNN]: using hidden layers is the principal feature of this type of neural. These hidden layers are usually explained as hyper-planes in solving the classification issues. It is used for classifying different categories of data.
- Polynomial Neural networks [PNN]: Polynomial Neural networks create a multivariate polynomial mapping using neurons as multilayer perception [18].

2.6 Relationship Mining

Relationship mining or Relation Extraction [RE] is the mission of recognizing the various relations that may exist between at least two named elements. These semantic relations are invested for understanding the human language. [RE] frameworks can be broadly detached and separated into three classes: Supervise, unsupervised, and semi- supervised [19]. The two main implementations of RE are: Automatic question -answering and bio -text mining [20].

2.7 Outlier Detections

Outlier Detections or outlier mining is the process of finding objects with practices that are entirely and totally different from desire, such object are called outlier [anomalies]. Discovering anomalies from a compilation of examples is a well -known issue in the field of DM. Detect the outlier is not a well-shaped and command more attention [21].

Outlier detection has many applications like insurance claim, fraud detection, intrusion detection, medical and public health outlier detection, mobile phone and industrial manage detection [22]. The causes for handling the outlets are:

Anomalies significantly affect the aftermaths of databases; outliers are regarded as recorded mistakes. Hence, some of them might be intrigued and helpful for the conclusions and results. In many researches regions, these exceptions can be the key in the revelations of unpredicted information [23].

2.8 Text Mining

Text mining defined as knowledge discovery from structured databases, which has found new information from unidentified textual data through different offprint techniques. Text mining is a multidisciplinary field, particularly text analysis, information extraction, natural language processing, extraction of information and text recapitulation [24]. The Text mining process has two phases:

First: collect information from multiple resources.

Second: return and pre-process it by review format and character sets. At this point the document would go through a text analysis phase.

These steps sometimes repeated until detect profitable information, the resulting information will be set in a management information system to be used by the concerned users [25].

The Text Mining approach is entirely connected to a wide assortment of field, such as scientific discovery in life science and bioinformatics, automated ad allotment in business, e-discovery, passion analysis in social media, national security and competitive intelligence [26]. Text mining has many implementations in medical domain such as, information extraction from electronic health records, extraction of knowledge from biomedical literature, biomakers and related genes associated with disease [27].

2.9 Social Network Analysis [SNA]

Social Network Analysis [SNA] is the significant process for analyzing social network. SNA might be built from relational information and can be categorized as some of social fundament, for example, individuals, gathering, and associations with definite connections or collaborations between them. These systems are normally manifested and illustrated by charts where vertices clarify the social substances and edges justify the ties built up between them [28]. SNA research emerged out of human science, brain research, and chart hypothesis.

SNA are settled in many eliminated of information and in the greater scope of scales. DM of social networks must be employed the diagram mining techniques, for example: modeling, data processing, evolution and structure, prediction, classification / topologies, detection, pattern measurement and metrics, efficiency and communities [29].

SNA has multiple implementations in health research. It can be invested to think about shared systems of health experts, the prevalence of irresistible diseases and risky practices, care pathways and the spread of health protection and promotion programs [30].

2.10 Trend Analysis

Trend analysis is being a typical example of patterns after some time in the information to illustrate the course of progress and can be invested to research vulnerabilities in different time focuses and relationship with different components. There are various time series applications such as: Credit card transactions, offers of an organization`s sales and stock costs. These applications are seen as objects with time characteristics, it is enchanting to find out patterns and regularities in the information along the elements of time. Trend analysis finds these enchanting examples [31]. Trend analysis in medical services employed for quality control of social insurance enhancements, cost of consideration patterns, therapeutic mediation assessment, and so forth [32].

3. DATA MINING PROCESSES

DM process has a re occurred nature. The influence of DM triggers new business addresses, thus can be utilized to grow progressively focussed models [33]. Data mining implies some of the following processing key:

3.1 Problem Definition

The prime step is to determine objectives. In view of the characterized objective, the right arrangement of devices can be applied to the information to make the parallel model.

3.2 data investigation and understanding

If the nature of information isn't fitting for a specific model, then proposals on future information gathering and capacity method-logies can be made at this, for analysis, all information must be solidified With the purpose of that it tends to be dealt with reliably.

3.3 Data Preparation

The motive behind this progression is to clean and change the information with the purpose that absent and invalid values deal with accomplished legal values are made reliable for increasingly powerfully exploration.

3.4 Modelling

Depending on the information and the typical results, an information mining algorithm or combination of algorithms is selected for examination. These algorithms merge traditional systems, for example: clustering, neighborhoods and statistical, in addition to other techniques such as rule based algorithms, decision trees and networks. The definite algorithm is chosen depending on the specific aims to be completed and the nature of the information to be inquired.

3.5 Evaluation and Deployment

Based on the consequences of the DM algorithms, an inquiry is applied to decide key ends from the analysis and make a set of suggestion for thoughts [34].

4. DATA MINING ARCHITECTURE

There are three levels in Data Mining Architecture:

4.1 Data Layer

Data layer can be database and sometimes data distribution center. This layer is an interface for all data sources.

DM results are put away in data layer so it is very well to be demonstrated to the user as reports or other kinds of conception and ideation.

4.2 Data Mining Application Layer

Is invested to retrieve information from the database. Some change routine can be applied here to change information into needing format. At this point, information is prepared utilizing a various DM algorithm.

4.3 Front - End Layer

Gives natural and friendly user interface [UI] for users to activate with DM systems. DM result is shown in representation structure to the user in the Front - end layer [35].

5. IMAGE MINING

The synonymous term to DM is the Image Mining [IM]. Image Mining manages the extraction of image patterns from a huge stack of images.

It is evidently, IM is not really the same as image processing techniques and low- level computer vision in scope of the fact that the focal point of IM is in the pulling out- extraction- of patterns from the huge heap of images, whereas the focal point of image processing and computer vision is in extracting and understanding clear highlights from secluded image. In IM, the purpose is the manifestation of image patterns that are remarkable in a given pile of images [36]. IM procedure consists of few segments and including:

- Image analysis covering image pre-processing, object recognition and feature extraction.
- Image classification.
- Image indexing.
- Image retrieval.
- Data management [37].

Many scientists have an unsuitable impression that IM is only a direct growth of information mining applications while others view IM as another name for pattern distinction [38]. Decide how low- level, pixel portrayal contained in a crude picture or the picture arrangement can be productive and viably tackled to recognize high - level spatial objects and connections, is the main problem in image mining.

The satellite image and medical image are the most well- known uses of IM. The medical databases contain the structural information and non- structural medical image of patients, with the goal that medical data suppose a critical job in the medical domain, since the medicinal analysts consider the efficient satisfaction of the treatment depends on the data in the medical database [39].

6. BRIEF VIEW OF CANCER

Cancer is a broad term. It depicts the sickness that outcomes when the cell changes cause the uncontrolled development and division of cells. A few kinds of cancer cause fast cell development, while others cause cells to develop and separate at a slower rate. Tumors are the name called to certain types of cancer, while others, such as leukemia do not. There are two types of tumors:

- Benign tumor: cells are restricted to one area and are not ready to spread to different parts of the body.
- Malignant tumor: this consists of cancerous cells which can spread by going through the lymphatic system or bloodstream [40].

6.1 Reasons for Cancers

- Physical cancer: causing factors, for example: ionizing radiation and ultraviolet.
- Chemical cancer: causing factors, for example: toxicant, asbestos, arsenic and tobacco smoke.
- Biological cancer: causing factors, for example: diseases from certain infections, microscopic organism, or parasites. It is worth mentioning, the advancement of malignancy in grown-up is the aftermaths of different transformation in multiple qualities connected with controlling the development of cells and modified metabolic changes in tumor cells and the tumor microenvironment that advance and accelerate a definite development of the disease.
- Highest Weight: low food grown from the ground, the lack of physical movement.
- Alcohol use.
- Contamination: causing cancer, for example: hepatitis and human papilloma infection [HPV] particularly in developing countries [41].

Cancers are characterized in two different ways: By the type of tissue wherein the disease begins and by the principal site, or the area in the body where the cancer originally created.

From a histological viewpoint there are many various cancers which are gathered into six noteworthy classifications:

- Carcinoma [Carcinoma refers to a harmful neoplasm of epithelial root or malignant growth of the inside and outside or outer coating of the body].
- Sarcoma [Sarcoma refers to malignancy that begins in strong and connective tissues, for example: bones, ligaments, tendons, muscles, and fat].
- Myeloma [Myeloma is a malignant growth that starts in the plasma cells of the bone marrow].
- Leukemia [Leukemia is the increasing in immature white blood cells in blood].
- Lymphoma [Lymphomas create in the organs or hubs of the lymphatic framework, a system of vessels, hubs, and organs that filter natural liquids and produce disease – fighting white blood cells, or Lymphocytes].
- Mixed Types [The sort parts might be inside one class or from various classifications for example carcinosarcoma] [42].

6.2 Methods of Predicting Cancer

The researchers are depending on numerous methods for predicting cancer [in advance] by using DM methods; such methods are using Medical Imaging and genetic information [DNA assay].

Medical Imaging need turned an essential piece in the analysis, early detection and diagnosis of cancer. The reason is that the important force of medical imaging in giving medical agents with sights and knowledge concerning human malady and physiology.

In this role, imaging may be invested to diagnose cancer stage, screen, and figure out whether a medication will be working, follow up tumor duplication, advance therapeutic research, especially in crucial fields such drug discovery and medical advancement to support patient care [43].

As for genetic information utilized by the gathering of scores of alterations affecting the structure and function of the genome. At the same time it is very important in this process the genetic alterations and epigenetic changes, whereas the former disrupt normal patterns of gene expression, and sometimes leading to the expression of abnormal, fundamentally, active proteins, the latter disrupt the mechanisms such as transcriptional control leading to the improper silencing or activation of cancer - associated genes.

Table 1. The researchers used DM in predicting and classify cancer

Algorithm's used	Year of publication	Author [s] for publication	Cancer Type
C5 Algorithm	2010	Qi Fan, Chang-jie Zhu, Liu Yin [45]	Breast cancer
SVM, tree forest, tree boost	2010	Medhat Mohamed Ahmed, Hala Abou Sena, Muhamed Wael Farouq, Abdel-Badeeh Mohamed [46]	Breast cancer
Decision Tree	2010	Orlando AnunciaçãoBruno C. GomesSusana VingaJorge GasparArlindo L. OliveiraJosé Rueff [47]	Breast cancer
Bayesian Belief Network	2010	Jyotirmay Gadewadikar, Ognjen Kuljaca, Kwabena Agyepong, Erol Sarigul, Yufeng Zheng, Ping Zhang [48]	Breast cancer
Particle Swarm Optimization	2010	K. Rajiv Gandhi, Marcus Karnan, S. Kannan [49]	Breast cancer
Gaussian Classifier	2010	radu dobrescu, matei dobrescu, stefan mocanu, dan popescu[50]	Skin cancer
Statistical neural network structures, self-organizing map [SOM], radial basis function network [RBF], general regression neural network [GRNN] and probabilistic neural network [PNN]	2010	A. Soltani Sarvestani, A. A. Safavi, N.M. Parandeh ; M. Salehi [51]	Breast cancer
K-means, SOM, HAC	2010	Ritu Chauhan, Harleen Kaur, M.Afshar Alam [52]	Number of cancers
Radial Basis Function [RBF] neural network and multilayer Perceptron [MLP]	2011	J. Padmavathi [53]	Breast cancer
Artificial neural network [ANN], regression tree [CART]	2011	Cheng-Mei Chen, Chien-Yeh Hsu, Hung-Wen Chiu, Hsiao-Hsien Rau [54]	Liver cancer
Decision tree classifier-CART	2011	D. Lavanya , K. Usha Rani [55]	Breast cancer
Support vector machines, Artificial neural network, J48 decision tree, Random forest, LogitBoost, Decision stump, Random subspace, Reduced error pruning tree, Alternating decision tree, Voting	2011	Ankit Agrawal, Sanchit Misra, Ramanathan Narayanan, Lalith Polepeddi, Alok Choudhary [56]	Lung cancer
Naïve Bayesian, K Nearest Neighbors, SVM.	2012	Muhammad Shahbaz, Shoaib Faruq, Muhammad Shaheen, Syed Ather Masood [57]	Leukemia
Latent Class Analysis, LCA-ensemble, Neural Network, NN-ensemble	2012	Shima Ghassem Pour, Peter Mc Leod, Brijesh Verma, Anthony Maeder [58]	breast, lung, blood cancers
Decision tree, Bayesian network, Neural network, Naïve Bayes, support vector machine, logistic regression	2012	Shweta Kharya [59]	Breast cancer

Algorithm's used	Year of publication	Author [s] for publication	Cancer Type
C4.5 algorithm	2012	K. Rajesh, Dr. Sheila Anand [60]	Breast cancer
Multilayer Perceptron [MLP] and Decision tree J48	2012	Gouda I. Salama, M.B.Abdelhalim and Magdy Abd-elghany Zeid [61]	Breast cancer
Multi layer perceptron [MLP] and the radial basis function RBF	2012	Ali Raad, Ali Kalakech, Mohammad Ayache [62]	Breast cancer
EM clustering, bagging, Adaboost	2012	S M Halawani, M Alhaddad, A Ahmad [63]	Breast cancer
Decision tree, SVM, AdaBoost, Bagging and naïve Bayes	2012	Shardul Pandya, Charles A. Edeki [64]	Breast cancer
I^2 - Clustering, k-means	2012	S. Santhosh Kumar, A.Sumathi, Dr. E.Ramaraj [65]	Colon cancer
Decision Tree	2013	A.Priyanga, Dr. S. Prakasam [66]	Breast cancer
Decision tree, J48 algorithm	2013	Shiv Shakti Shrivastava, Anjali Sant, Ramesh Prasad Aharwal [67]	Breast cancer
Neural network and SVM	2013	Ada Rajneet, Kaur [68]	Lung cancer
Naïve Bayes followed by IF-THEN rule, Decision Trees and Neural Network	2013	V.Krishnaiah, G.Narsimha, N.Subhash Chandra[69]	Lung cancer
supervised Artificial Neural Network [ANN],unsupervised Artificial Neural Network ,Statistical and decision tree	2013	H. S. Hota [70]	Breast cancer
Support vector machines [SVMs] and Decision tree	2013	Reeti Yadav, Zubair Khan, Hina Saxena [71]	Breast cancer
Random forest and multivariate adaptive regression	2013	Dengju Yao, Jing Yang, Xiaojuan Zhan [72]	Breast cancer
Back-propagation artificial neural network, k-nearest neighbor	2013	Mahmoud Elgamal [73]	Skin cancer
Decision tree [j48], neural networks, naive bayes [nb], logistic regression [lr], support vector machine [svm], k-nearest neighbor [knn]	2013	G. Ravi Kumar, G. A. Ramachandra, K. Nagamani [74]	Breast cancer
Bayes Net, Naive Bayesian, Simple Logistics, Multilayer Perceptron, Sequential Minimal Optimization [SMO], k-nearest-neighbor[IBK], KStar	2013	Gopala Krishna Murthy Nookala, Bharath Kumar Pottumuthu, Nagaraju Orsu, Suresh B. Mudunuri [75]	Breast Cancer, Lymphoma, Leukemia
Bayesian Network prediction	2013	Ramani RG1, Jacob SG [76]	Lung cancer
Random tree, ID3, CART, C4.5 and Naive Bayes	2013	S. Syed Shajahaan, S. Shanthi,V.ManoChitra[77]	Breast cancer
PCA, Neural network	2013	Ada, Rajneet Kaur [78]	Lung cancer
Decision Tree [DT], Artificial Neural Network [ANN], and Support Vector Machine [SVM]	2013	Sahar A. Mokhtar, Alaa. M. Elsayad [79]	Breast cancer

Algorithm’s used	Year of publication	Author [s] for publication	Cancer Type
Decision tree, radial basis function kernel support vector machine [RBF-SVM]	2013	Alaa Elsayad [80]	Breast cancer
Naïve Bayes and Logistic Regression	2013	Rafaqat Alam Khan, Nasir Ahmad, Nasru Minallah [81]	Breast cancer
semi-supervised learning [SSL]	2013	Juhyeon Kim and Hyunjung Shin [82]	Breast cancer
K-means clustering and rough set	2013	T. Sridevi Annamalai Murugan [83]	Breast cancer
Support vector machines, artificial neural networks, and semi-supervised learning models	2013	Kanghee Park, Amna Ali, Dokyoon Kim, Yeolwoo An, Minkoo Kim, Hyunjung Shin [84]	Breast cancer
ANFIS, artificial neural network, support vector machine and logistic regression	2013	Chang SW, Abdul-Kareem S, Merican AF, Zain RB [85]	Oral cancer
Multivariate Adaptive Regression Splines [MARS], C5.0	2013	Chi-Chang Chang, Sun-Long Cheng, Chi-Jie Lu, and Kuo-Hsiung Liao [86]	Cervical Cancer
Naïve Bayes tree, Radial Basis Function Neural Network, Support Vector Machine	2014	R. Nithya and B. Santhi [87]	Breast cancer
J48, MLP and Rough set	2014	Ahamed Lebbe Sayeth Saabith, Elankovan Sundararajan, Azuraliza Abu Bakar [88]	Breast cancer
Decision tree, K means clustering, support vector machine	2014	P. Ramachandran, N. Girija, T. Bhuvaneswar [89]	More than one cancer
Linear Discriminant Analysis, Multi Layer Perceptron, Decision Trees, Logistic Regression, Support Vector Machines, Naïve Bayes, K-Nearest Neighborhood	2014	Zehra Karapinar Senturk1andResul Kara [90]	Breast cancer
Decision tree [ID3, C4.5, C5, J48, CART and CHAID]	2014	Ronak Sumbaly, N. Vishnusri, S. Jeyalatha [91]	Breast cancer
Classification rules	2014	Miss Jahanvi Joshi Mr. RinalDoshiDr. Jigar Patel[92]	Breast cancer
Algorithms of regression	2014	Ritu Tayal [93]	Breast cancer
Decision Tree, Naïve Bayes and KNN	2014	J. S. Saleema, P. Deepa Shenoy, K. R. Venugopal, L. M. Patnaik [94]	Breast, Colorectal and Respiratory Cancer
Sequential Minimal Optimization [SMO], IBK [K Nearest Neighboursclassifier] , BF Tree	2014	Vikas Chaurasia, Saurabh Pal [95]	Breast cancer
K-means and support vector machine [K-SVM] algorithms	2014	Bichen Zheng, Sang WonYoon, Sarah S.Lam [96]	Breast cancer
Artificial neural network [ANN]	2014	Zakaria Suliman Zubi, Rema Asheibani Saad [97]	Lung cancer
REP tree, radial basis function [RBF] network, simple logistic	2014	Vikas Chaurasia, Saurabh Pal [98]	Breast cancer
semi-supervised learning algorithm based on a graph regularization approach	2014	Park C, Ahn J, Kim H, Park S [99]	Breast,corolectal,colon cancers

Algorithm's used	Year of publication	Author [s] for publication	Cancer Type
J48, NB, SVM, Bayesian network	2014	Aniket Bochare, Aryya Gangopadhyay, Yelena Yesha, Anupam Joshi and Yaacov Yesha [100]	Breast cancer
Decision tree	2015	Joana Diz, Goreti Marreiros, Alberto Freitas [101]	Breast cancer
NB, J48, BK, multilayer perceptron	2015	Er.Tapas Ranjan Baitharu, Subhendu KumarPani [102]	Lung cancer
naïve bayes', J48, decision trees	2015	Peter Adebayo Idowu, Kehinde Oladipo Williams, Jeremiah Ademola Balogun, Adeniran Ishola Oluwaranti [103]	Breast cancer
Decision tree	2015	K.Arutchelvan,Dr.R.Periyasamy [104]	breast, skin, and lung cancers
FP-Growth algorithm and decision tree	2015	Jaimini Majali, Rishikesh Niranjan,Vinamra Phatak, Omkar Tadakhe [105]	Breast cancer
Artificial neural network [ANN]	2015	Htet Thazin Tike Thein, Khin Mo Mo Tun [106]	Breast cancer
Rough set with backpropagation neural network [RS-BPNN]	2015	Kindie Biredagn Nahato, Khanna Nehemiah Harichandran, Kannan Arputharaj [107]	Hepatitis, heart and breast cancer
K-means clustering, AprioriTid Algorithm, Decision Tree	2016	Neelam Singh,Santosh Kumar Singh Bhadauria[108]	Lung cancer
J48, CART, ADTree	2016	B.Padmapriya, T.Velmurugan [109]	Breast cancer
Naive Bayes, Random Forests	2016	Diz J., Marreiros G., Freitas A [110]	Breast cancer
PCA and LDA	2016	Divya Chauhan ,Varun Jaiswal [111]	Lung cancer
Naive Bayes, Bayesian network and J48 algorithm	2016	T.Christopher, J.Jamera banu [112]	Lung cancer
SVM, C4.5, NB, K-NN	2016	Hiba Asri, Hajar Mousannif,Hassan AlMoatassime,Thomas Noel [113]	Breast cancer
Artificial neural network [ANN], PS-classifier, genetic algorithm	2016	Shokoufeh Aalaei, Hadi Shahraki, Alireza Rowhanimanesh, Saeid Eslami [114]	Breast cancer
Threshold and morphological operation, GLCM, SVM	2016	Mona nasr, Amr Atif Abd El-Mageed [115]	Lung cancer
Neural network and relief feature selection	2017	Zahraa N. Shahweli, Ban N. Dhannoon [116]	Breast cancer
Back propagation neural network	2017	Zahraa N. Shahweli, Ban N. Dhannoon [117]	Lung cancer
Back propagation neural system	2017	P. Mohamed Sajid, A. Rajesh, Abdul Hakeem [118]	Skin cancer
Expectation Maximization [EM] and Classification and Regression Trees CART]	2017	Mehrbakhsh Nilashi, Othman Ibrahim, Hossein Ahmadi, Leila Shahmoradi [119]	Breast cancer
Information Gain and Support Vector Machine	2017	Lingyun Gao, Mingquan Ye, Lu, Daobin Xiaojie Huangd [120]	Lung, colon and prostate cancer

Algorithm’s used	Year of publication	Author [s] for publication	Cancer Type
Gauss-Newton representation based algorithm [GNRBA]	2017	Lingraj Dora, Sanjay Agrawal, Rutuparna Panda, Ajith Abraham [121]	Breast cancer
Naïve Bayes[NB], Logistic Regression[LR], Decision Tree[DT]	2017	Subrata Kumar Mandal [122]	Breast cancer
Information gain [IG], Genetic Algorithm [GA] and Genetic Programming [GP]	2017	Hanaa Salem,Gamal Attiyab,Nawal El-Fishawy [123]	seven cancers
J48, Function Tree, Random Forest Tree, AD Alternating Decision Tree, Decision stump and Best First	2017	Nusaibah Kh. Al-Salihy, Turgay Ibrikci [124]	Breast cancer
Backpropagation and Support vector machine system [SVM]	2017	Gomathi N, Sandhya P [125]	Breast cancer
ZeroR, J48	2017	Sumalatha.G, Archana.S [126]	Breast cancer
Naive Bayes, logistic regression, decision trees,	2017	Elham Sagheb Hossein Pour, Rohit J. Kate [127]	bladder,breast, cervix uteri, colorectal, corpus uteri, esophagus, liver, lung, prostate and stomach
Support vector machine [SVM], Regularized Least Squares [RLS], multi-layer perceptron [MLP] with back propagation and deep neural network [DNN]	2017	Wafaa K. Shams Zaw Z. Htike [128]	Oral cancer
Instance-based k-nearest Neighbors, Naive Bayesian, logistic model tree	2018	Safae Sossi Alaoui, Yousef Farhaoui Labsiv, B. Aksasse [129]	Breast cancer
J48,Naïve Bayes,K Nearest Neighbor	2018	Yomna Omar, Abdullah Tasleem, MichelPasquier,Assim Sagahyroon [130]	Lung cancer
K-means	2018	Noor Kadhim Ayoob [131]	Breast cancer
SVM, NB and C4.5	2018	Pradeep K R, Naveen N C [132]	Lung cancer
Decision Tree [DT], Random Forest [RF], Support Vector Machine [SVM], Neural Network [NN] and Logistics Regression	2018	Yixuan Li, Zixuan Chen [133]	Breast cancer
Naı¨ve Bayes, RBF Network, and J48	2018	Vikas Chaurasia, Saurabh Pal a, BB Tiwari [134]	Breast cancer
Probabilistic neural network, perceptron-based neural network, random forest, one rule, decision tree	2019	Davide Chicco, Cristina Rovelli [135]	Lung cancer
Back propagation neural network	2019	Ayad Ghany Ismaeel [136]	B-Thalassemia

Algorithm's used	Year of publication	Author [s] for publication	Cancer Type
GRU-SVM[4], Linear Regression, Multilayer Perceptron [MLP], Nearest Neighbor [NN] search, Softmax Regression, and Support Vector Machine [SVM]	2019	Abien Fred M. Agarap [137]	Breast cancer
Decision support system [DSS] and random optimization [RO]	2019	Patrizia Ferroni, Fabio M. Zanzotto, Silvia Riondino, Noemi Scarpato, Fiorella Guadagni,Mario Roselli [138]	Breast cancer
Random Forests	2019	Nguyen Phuoc Long, Kyung Hee Jung, Nguyen Hoang Anh, Hong Hua Yan, Tran Diem Nghi, Seongoh Park and et al [139]	Pancreatic cancer
Decision Tree, Naïve Bayes, Association rule, Multilayer Perceptron [MLP], Random Forest, and Support Vector Machines [SVM]	2019	Keerthana Rajendran, Manoj Jayabalan, Vinesh Thiruchelvam, and V. Sivakumar [140]	Breast cancer
genetic algorithm	2019	F. Leena Vinmalar, Dr. A. Kumar Kombaiya [141]	Lung cancer
Wrapper Subset method and Random Forest	2019	Esraa H. Abdelaziz, Sanaa M. Kamal, Khaled El-Bhanasy, Rasha Ismail [142]	Liver Cancer
Boosted decision tree, decision forest and decision jungle algorithms	2019	Talha Mahboob Alam, Muhammad Milhan Afzal Khan, Muhammad Atif Iqbal, Abdul Wahab,Mubbashar Mushtaq [143]	Cervical cancer
Gray Wolf Optimization [GWO] and support vector machine [SVM]	2019	Kamel, S.R., YaghoubZadeh, R. & Kheirabadi, M [144]	Breast cancer
Decision Tree, Naïve Bayes, Association rule, Multilayer Perceptron [MLP], Random Forest, and Support Vector Machines [SVM]	2019	Keerthana Rajendran, Manoj Jayabalan, Vinesh Thiruchelvam, and V. Sivakumar [145]	Breast cancer
SVM, DT and k-NN	2019	Saadaldeen Rashid Ahmed Ahmed, Israa Al_Barazanchi, Ammar Mhana , Haider Rasheed Abdulshaheed [146]	Lung Cancer
XG Boost and logistic regression	2019	Taninaga, J., Nishiyama, Y., Fujibayashi, Toshiaki Gunji, Noriko Sasabe, Kimiko Iijima & Toshio Naito [147]	Gastric Cancer
Deep Belief Network and Restricted Boltzmann Machines	2020	Zahraa Naser Shahweli [148]	Lung cancer

Both types of changes are inheritable at the cellular level, thus contributing and participating to the clonal expansion of cancer cells.

So, important knowledge on how genetic alterations in oncogenes or tumor suppressor genes, as well as epigenetic changes, which can be utilized in the clinics as biomarkers for cancer detection, diagnosis and prognosis [44].

7. METHODOLOGY

This research summarized most of the research that dealt with a prediction or classification of cancer types based on data mining for the years between 2010 and 2020. The researcher relied on his research on several sites, including: google scholar, Scopus, web of science, Elsevier. So the number of research was large and for different types of cancer.

8. LITERATURE REVIEW

In Table 1 most researches which use data mining for diagnosis, predict or classify cancer in latest ten years.

9. RESULTS AND DISCUSSION

By reviewing the researches above, found the effectiveness of data mining and machine education in the medical field, especially the classification and detection of cancer. Through the analysis of these researches, the benefits of machine learning algorithms have emerged, in addition to the challenges facing the use of these algorithms.

9.1 Benefits

DM uses to distinguish helpful and reasonable patterns by breaking down abundant arrangements of information.

DM can be invested to decrease costs by expanding efficiencies, improve understanding personal conviction and may be in particular spare the lives of more patients.

DM methods are useful for disclosing the distinction in complex protemic designs.

DM strategies can be used to set up a computational procedure to predict the class of tumors from the auxiliary and physicochemical properties of protein arrangements.

Prediction of clinical result of patients after cancer medical procedure.

Patients checking, remote observing, tele medicine and home care [149].

9.2 Challenges

Data produced is voluminous and heterogeneous and from different sources which affect determination, all these parts can basically affect on extracting conclusion, guess, and treatment of the patient.

Medical information is difficult to be accurate and decisive.

Privacy is another test to the medical information in the light of that information must be shared.

Building and keeping up an information warehouse for the most skilled and exact mining can be a costly challenge.

While DM is a significant tool for discovering patterns in vast databases, Noteworthy to say that the affectability and clearness of DM tools will affect the prescient assessment of the piled data. affectability and clearness are particularly so significant in medical DM.

Another stumble stone which is that: all outcomes, findings and medications in drug are uncertain and exposed to mistake rate. At this point here the examination of affectability and clearness are being employed for the estimation of these mistakes [150].

10. CONCLUSION

This paper reviewed several research works which are done for diagnosis, predicting and classifying cancers. The DM is used in the scope of medical foreseeing which is discussed. The main focus is on using different algorithms for cancer prediction by using data mining. Depending on the analysis of their outcomes, it is getting clearer that the compact of multidimensional diverse data, affiliated with the applicant of different techniques for feature selection and classification can provide promising tools for inference in the cancer domain. The use of these methods by researchers helps physicians and specialists to provide ideas that help in the accuracy of the results and the development of methods to reach better results. So it is possible in the future to combine more than one way to achieve the goals.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Gayathri V, Chanda Mona M, Banu Chitra S. A survey of data mining techniques on medical diagnosis and research. International Journal of Data Engineering. 2014;6(6):301-310.
   Avialable:https://pdfs.semanticscholar.org/0ae0/ed6e36950fc216cf4504d00eaf9246a5fb8f.pdf
2. Jemal A1, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA:* A cancer journal for clinicians. 2011;61(2):69-90.
   DOI: 10.3322/caac.20107.
3. Vijiyarani S, Sudha S. Disease prediction in data mining technique – A survey. International Journal of Computer Applications & Information Technology. 2013;2(1):17-21.
   Available:https://www.ijcait.com/IJCAIT/21/213.pdf
4. Elma Kolçe, Neki Frasheri. A Literature Review of Data Mining Techniques Used in Healthcare Databases. ICT Innovations 2012 Web Proceedings. 2012; 577-582.
   Available:https://shomatrans.com/wp-content/uploads/2017/04/zzz.pdf.
5. Pradnya P. Sondwale. Overview of predictive and descriptive data mining techniques. International Journal of Advanced Research in Computer Science and Software Engineering. 2015;5(4): 262-265.
6. Dhanya PV, Tintu PB. A survey on health data using data mining techniques. International Research Journal of Engineering and Technology (IRJET). 2015;2(7):713-720.
7. Mansi Gera, Shivani Goel. Data mining - techniques, methods and algorithms: A review on tools and their validity. International Journal of Computer Applications. 2015;113(18):22-29.
8. Hussain AM, Zahid A, Munam AS. Data mining techniques and applications – A decade review. 23rd International Conference on Automation and Computing (ICAC). IEEE; 2013.
   Available:https://www.researchgate.net/publication/320822551_Data_mining_techniques_and_applications_-_A_decade_review.
9. Shital HB, Nirav B. Data mining techniques and trends – a review. Global journal for research analysis (Graz). 2016;5(5):252-254.
10. Lourdu Caroline A, Manikandan S, Kanniamma D. Comparative study of Classification algorithms for Data Mining. International Journal of Engineering Science Invention (IJESI). 2018;3:7-11.
11. Safae Sossi A, Yousef Farhaoui, Brahim Aksasse. Classification algorithms in data mining. ACM Computing Classification System. Information system-Information systems applications- Data mining; 2018.
12. Neelamegam S, Ramaraj E. Classification algorithm in data mining: An overview. International Journal of P2P Network Trends and Technology (IJPTT). 2015; 3(5):1-5.
13. Sharmila K, Vethamanickam SA. Survey on data mining algorithm and its application in healthcare sector using hadoop platform. International Journal of Emerging Technology and Advanced Engineering. 2015;5(1):567-571.
14. Anupama YK, Amutha S, Ramesh BD. Survey on data mining techniques for diagnosis and prognosis of breast cancer. International Journal on Recent and Innovation Trends in Computing and Communication. 2017;5(2):33-37.
15. Nanhay Singh, Ram Shringar R, Chauhan RK. Data mining with regression technique. Journal of Information Systems and Communication. 2012;3(1):199-202.
16. Alaa ME, Asem HS. Blood tumor prediction using data mining techniques. Health Informatics - An International Journal (HIIJ). 2017;6(2):23-30.
17. Jiawei Han, Micheline Kamber, Jian Pei. Data Mining Concepts and Techniques. 3rd ed. British Library Cataloguing. Elsevier; 2012.
18. Simon Hakin. Neural Network: A Comprehensive foundation. 2nd ed. Prentice Hall PTR Upper Saddle River, NJ, USA; 1998.

19. Ioannis PK, Suraj P, Suresh M. Graph-based relation mining. Multimedia communications, services and security. MCSS.Communications in Computer and Information Science, Springer, Berlin, Heidelberg. 2011;149:100-112.
Available:https://doi.org/10.1007/978-3-642-21512-4_12
20. Natalia Konstantinova. Review of relation extraction methods: What Is new out there? Analysis of images, social networks and texts. AIST 2014. Communications in Computer and Information Science, Springer, Cham, 2014;436:15-28.
Available:https://doi.org/10.1007/978-3-319-12580-0_2
21. Krishna Modi, Bhavesh Oza. Outlier analysis approaches in data mining. International Journal of Innovative Research In Technology. 2016;3(7):6-12.
22. Rashi Bansal, Nishant Gaur, Shailendra Narayan S. Outlier detection: Applications and techniques in Data Mining. 2016 6th International Conference - Cloud System and Big Data Engineering (Confluence). Noida, India. IEEE; 2016. DOI: 10.1109/CONFLUENCE.2016.7508146.
23. Deepti Mishra, Devpriya Soni. Outliers in Data Mining: Approaches and detection. International Journal of Engineering & Technology. 2018;7(4.39):189-198.
24. Shilpa Dang, Peerzada Hamid A. Text mining: Techniques and its application. IJETI International Journal of Engineering & Technology Innovations. 2014; (4):22-25.
25. Sonali Vijay G, Archana Chaugule, Pramod Patil. Text mining methods and techniques. International Journal of Computer Applications. 2014;85(17):42-45.
26. Varsha DB, Petras JK, Ilya AV. Text Mining for Protein Docking. PLoS Comput Biol. 2015;11(12).
DOI: 10.1371/journal.pcbi.1004630
27. Vinaitheerthan Renganathan. Text mining in biomedical domain with emphasis on document clustering. Healthcare Informatics Research. 2017;23(3):141-146.
DOI: 10.4258/hir.2017.23.3.141.
28. Shazia Tabassum, Fabiola SF Pereira, Sofia Fernandes, João Gama. Social network analysis: An overview. Wiley interdisciplinary reviews: Data Mining and Knowledge Discovery. 2018;8(5).
DOI: 10.1002/widm.1256.
29. Peter Mika. Ontologies are Us: A unified model of social networks and semantics. Journal of Web Semantics. 2007;5(1):5-15.
DOI:10.2139/ssrn.3199347.
30. Abhinav Choudhury, Shruti Kaushik, Varun Dutt. Social-network analysis in healthcare: analysing the effect of weighted influence in physician networks. Network Modeling Analysis in Health Informatics and Bioinformatics. 2018; 7(17).
Available:https://doi.org/10.1007/s13721-018-0176-y
31. Shital H. Bhojani, Nirav Bhatt. Data mining techniques and trends – A review. Gjra - Global Journal for Research Analysis. 2016;5(5):252-254.
32. Yin L, Wu Q, Hong D. Statistical methods and software package for medical trend analysis in health rate review process. Journal of Health & Medical Informatics. 2016;7(2).
33. Roseline Jecintha I, Poonguzhali V. Study on Data mining techniques for cancer prediction system. International Journal of Data Mining Techniques and Applications. 2018;7(1):60-63.
34. Shelly Gupta, Dharminder Kumar, Anand Sharma. Data mining classification techniques applied for breast cancer diagnosis and prognosis. Indian Journal of Computer Science and Engineering (IJCSE). 2011;2(2):188-195.
35. Hemlata Sahu, Shalini Shrma, Seema Gondhalakar. A brief overview on data mining survey. International Journal of Computer Technology and Electronics Engineering (IJCTEE). 2011;1(3):114-121.
36. Khasim Syed1, Srinivasu SVN. A review of web image mining tools, techniques and applications. International Journal of Computer Trends and Technology (IJCTT). 2017;49(1):36-43.
37. Barbora Zahradnikova, Sona Duchovicova, Peter Schreiber. Image mining: Review and new challenges. (IJACSA) International Journal of Advanced Computer Science and Applications. 2015; 6(7):242-246.
38. Yasodha KR, Yuvaraj KS. A study on image mining techniques. Innovation:

International Journal of Applied Research. 2013;1(1).

39. Saranya P, Satheeskumar B. A survey on feature selection of cancer disease using data mining techniques. International Journal of Computer Science and Mobile Computing. 2016;5(5):713-719.
40. Cancer: an overview. Cancer Council Western Australia; 2017.
    Available: https://bit.ly/2sUvQX7
41. Cancer. World Health Organization (WHO); 2018.
    Available: https://bit.ly/34V6F3Q
42. Anupama YK, Amutha S, Ramesh Babu DR. Survey on data mining techniques for diagnosis and prognosis of breast cancer. International Journal on Recent and Innovation Trends in Computing and Communication. 2017;5(2):33-37.
43. Neelaja K, Anand Kumar M, Minavathi. Medical image classification techniques for early cancer diagnosis: A comprehensive study. International Journal of Advanced Research in Computer and Communication Engineering. 2017;6(5):778-782.
44. Zdenko Herceg, Pierre Hainaut. Genetic and epigenetic alterations as biomarkers for cancer detection, diagnosis and prognosis. Molecular Oncology. 2007;1(1):26-41.
    DOI: 10.1016/j.molonc.2007.01.004
45. Qi Fan, Chang-jie Zhu, Liu Yin. Predicting breast cancer recurrence using data mining techniques. 2010 International Conference on Bioinformatics and Biomedical Technology, Chengdu, China. IEEE; 2010.
    DOI: 10.1109/ICBBT.2010.5478952
46. Medhat Mohamed A A, Hala Abou Sena, Muhamed Wael F, Abdel-Badeeh Mohamed. Using data mining for assessing diagnosis of breast cancer. Proceedings of the International Multi-conference on Computer Science and Information Technology, Wisla, Poland. IEEE; 2010.
    DOI: 10.1109/IMCSIT.2010.5679647.
47. Orlando Anunciação, Bruno C. Gomes, Susana Vinga, Jorge Gaspar, Arlindo L. Oliveira, José Rueff. A data mining approach for the detection of high-risk breast cancer groups. Advances in Bioinformatics. Advances in Intelligent and Soft Computing, 74. Springer, Berlin, Heidelberg; 2010.
    Available:https://doi.org/10.1007/978-3-642-13214-8_6.
48. Jyotirmay Gadewadikar, Ognjen Kuljaca, Kwabena Agyepong, Erol Sarigul, Yufeng Zheng, Ping Zhang. Exploring bayesian networks for medical decision support in breast cancer detection. African Journal of Mathematics and Computer Science Research. 2010;3(10):225-231.
49. Rajiv Gandhi K, Marcus Karnan S. Kannan. Classification rule construction using particle swarm optimization algorithm for Breast cancer data sets. International Conference on Signal Acquisition and Processing. Bangalore, India. IEEE; 2010. DOI: 10.1109/ICSAP.2010.58
50. Radu Dobrescu, Matei Dobrescu, Stefan Mocanu, Dan Popescu. Medical images classification for skin cancer diagnosis based on combined texture and fractal analysis. Wseas Transactions On Biology And Biomedicine. 210; 7(3);223-232.
51. Soltani Sarvestani A, Safavi AA, Parandeh NM, Salehi M. Predicting breast cancer survivability using data mining techniques. 2nd International Conference on Software Technology and Engineering. San Juan, PR, USA. IEEE; 2010.
    DOI: 10.1109/ICSTE.2010.5608818
52. Ritu Chauhan, Harleen Kaur, M.Afshar Alam. Data clustering method for discovering clusters in spatial cancer databases. International Journal of Computer Applications. 2010;10(6):9-14.
53. Padmavathi J. A comparative study on breast cancer prediction using RBF and MLP. International Journal of Scientific & Engineering Research. 2011;2(1):1-5.
54. Cheng-Mei Chen, Chien-Yeh Hsu, Hung-Wen Chiu, Hsiao-Hsien Rau. Prediction of survival in patients with liver cancer using artificial neural networks and classification and regression trees. 2011 Seventh International Conference on Natural Computation. Shanghai, China. IEEE; 2011.
    DOI: 10.1109/ICNC.2011.6022187
55. Lavanya D, Usha Rani K. Analysis of feature selection with classfication: Breast cancer datasets. Indian Journal of Computer Science and Engineering (IJCSE). 2011;2(5):756-763.

56. Ankit Agrawal, Sanchit Misra, Ramanathan Narayanan, Lalith Polepeddi, Alok Choudhary. A lung cancer outcome calculator using ensemble data mining on Seer data. BIOKDD '11 Proceedings of the Tenth International Workshop on Data Mining in Bioinformatics. San Diego, California; 2011.
57. Muhammad Shahbaz, Shoaib Faruq, Muhammad Shaheen, Syed Ather Masood. Cancer diagnosis using data mining technology. Life Science Journal. 2012;9(1):308-313.
58. Shima Ghassem Pour, Peter Mc Leod, Brijesh Verma, Anthony Maeder. Comparing Data Mining with Ensemble Classification of Breast Cancer Masses in Digital Mammograms. Proceedings of the Second Australian Workshop on Artificial Intelligence in Health. Aachen, Germany: CEUR-WS. AIH. Sydney. 2012;55-63.
59. Shweta Kharya. Using data mining techniques for diagnosis and prognosis of cancer disease. International Journal of Computer Science, Engineering and Information Technology (IJCSEIT). 2012; 2(2): 55-66.
60. Rajesh K, Sheila Anand. Analysis of SEER dataset for breast cancer diagnosis using C4.5 classification algorithm. International Journal of Advanced Research in Computer and Communication Engineering. 2012;1(2):72-77.
61. Gouda I. Salama, M.B.Abdelhalim, and Magdy Abd-elghany Zeid. Breast Cancer Diagnosis on Three Different Datasets Using Multi-Classifiers. International Journal of Computer and Information Technology (2277 – 0764). 2012;1(1):36-43.
62. Ali Raad, Ali Kalakech, Mohammad Ayache. Breast cancer classification using neural network approach: mlp and rbf. The 13th International Arab Conference on Information Technology ACIT. 2012;15-19.
63. Halawani SM, Alhaddad M, Ahmad A. A study of digital mammograms by using clustering algorithms. Journal of Scientific and Industrial Research (JSIR). 201; 71: 594-600.
64. Tapas Ranjan Baitharu, Subhendu Kumar Pani. A Comparative Study of Data Mining Classification Techniques using Lung Cancer Data. International Journal of Computer Trends and Technology (IJCTT). 2012;22(2):91-95.
65. Santhosh Kumar S, Sumathi A, Ramaraj E. Development of an efficient clustering technique for colon dataset. International Journal of Engineering and Innovative Technology (IJEIT). 2012;1(5): 83-86.
66. Priyanga A, Prakasam S. The role of data mining-based cancer prediction system (DMBCPS) in cancer awareness. International Journal of Computer Science and Engineering Communications-IJCSEC. 2013;1:54-61.
67. Shiv Shakti Shrivastava1, Anjali Sant2, Ramesh Prasad Aharwal3. An overview on data mining approach on breast cancer data. International Journal of Advanced Computer Research. 2013;3(4):256-262.
68. Ada, Rajneet Kaur. A study of detection of lung cancer using data mining classification techniques. International Journal of Advanced Research in Computer Science and Software Engineering. 2013;3(3):131-134.
69. Krishnaiah V, Narsimha G, Subhash Chandra N. Diagnosis of lung cancer prediction system using data mining classification techniques. International Journal of Computer Science and Information Technologies. 2013;4(1):39-45.
70. Hota HS. Diagnosis of breast cancer using intelligent techniques. International Journal of Emerging Science and Engineering (IJESE). 2013;1(3):45-53.
71. Reeti Yadav, Zubair Khan, Hina Saxena. Chemotherapy prediction of cancer patient by using data mining techniques. International Journal of Computer Applications. 2013;76(10):28-31.
72. Dengju Yao, Jing Yang, Xiaojuan Zhan. A Novel method for disease prediction: hybrid of random forest and multivariate Adaptive Regression Splines. Journal Of Computers. 2013;8(1):170-177.
DOI:10.4304/jcp.8.1.170-177.
73. Mahmoud Elgamal. Automatic skin cancer images classification. International Journal of Advanced Computer Science and Applications. 2013;4(3):287-294.
74. Ravi Kumar G, Ramachandra GA, Nagamani K. An efficient prediction of breast cancer data using data mining techniques. International Journal of

Innovations in Engineering and Technology (IJIET). 2013;2(4):139-144.

75. Gopala Krishna MN, Bharath Kumar P, Nagaraju Orsu, Suresh BM. Performance analysis and evaluation of different data mining algorithms used for cancer classification. International Journal of Advanced Research in Artificial Intelligence. 2013;2(5):49-55.

76. Ramani RG1, Jacob SG. Improved classification of lung cancer tumors based on structural and physicochemical properties of proteins using data mining models. PLOS (Public Library of Science). San Francisco, California. 2013; 8(3).
DOI: 10.1371/journal.pone.0058772

77. Syed Shajahaan S, Shanthi2 S, Mano Chitra V. Application of data mining techniques to model breast cancer data. International Journal of Emerging Technology and Advanced Engineering. 2013;3(11):362-369.

78. Ada, Rajneet Kaur. Using some data mining techniques to predict the survival year of lung cancer patient. International Journal of Computer Science and Mobile Computing. 2013;2(4):1-6.

79. Sahar A. Mokhtar, Alaa. M. Elsayad. Predicting the severity of breast masses with data mining methods. International Journal of Computer Science Issues (IJCSI). arXiv:1305.7057. 2013;10(2).

80. Alaa M. Elsayad, Elsalamony HA. Diagnosis of breast cancer using decision tree models and SVM. International Journal of Computer Applications. 2013; 83(5):19-29.

81. Rafaqat Alam Khan, Nasir Ahmad, Nasru Minallah. Classification and regression analysis of the prognostic breast cancer using generation optimizing algorithms. International Journal of Computer Applications. 2013;68(25):42-47.

82. Juhyeon Kim, Hyunjung Shin. Breast cancer survivability prediction with labeled, unlabeled and pseudo-labeled patient data. Journal of the American Medical Informatics Association (JAMIA). 2013; 20(4):613-618.

83. Sridevi T, A.Murugan. An intelligent classifier for breast cancer diagnosis based on K-means clustering and rough set. International Journal of Computer Applications. 2013;85(11);38-42.

84. Kanghee Park, Amna ali, Dokyoon Kim, Yeolwoo An, Minkoo kim, Hyunjung shin. Robust predictive model for evaluating breast cancer survivability. Engineering Applications of Artificial Intelligence. 2013; 26(9): 2194-2205.
Available:https://doi.org/10.1016/j.engappai.2013.06.013

85. Chang SW, Abdul-Kareem S, Merican AF, Zain RB. Oral cancer prognosis based on clinicopathologic and genomic markers using a hybrid of feature selection and machine learning methods. BMC bioinformatics. 2013;170.
DOI: 10.1186/1471-2105-14-170.

86. Chi-Chang Chang, Sun-Long Cheng, Chi-Jie Lu, and Kuo-Hsiung Liao. Prediction of recurrence in patients with cervical cancer using MARS and classification. International Journal of Machine Learning and Computing. 2013;3(1):75-78.

87. Nithya R, Santhi B. A data mining techniques for diagnosis of breast cancer disease. World Applied Sciences Journal 29 (Data Mining and Soft Computing Techniques. 2014;18-23.
DOI: 10.7763/IJMLC.2013.V3.276

88. Ahamed Lebbe Sayeth Saabith, Elankovan Sundararajan, Azuraliza Abu Bakar. Comparative study on different classification techniques for breast cancer dataset. International Journal of Computer Science and Mobile Computing. 2014; 3(10):185-191.
DOI: 10.5829/idosi.wasj.2014.29.dmsct.4

89. Ramachandran P, Girija N, Bhuvaneswari T. Early detection and prevention of cancer using data mining techniques. International Journal of Computer Applications. 2014; 97(13):48-53.

90. Zehra Karapinar Senturk 1, Resul Kara1. Breast cancer diagnosis via data mining: Performance analysis of seven different Algorithms. Computer Science & Engineering: An International Journal (CSEIJ). 2014;4(1):35-46.

91. Ronak Sumbaly N. Vishnusri S. Jeyalatha. Diagnosis of breast cancer using decision tree data mining technique. International Journal of Computer Applications. 2014; 98(10):16-24.

92. Jahanvi Joshi, RinalDoshi, Jigar Patel. Diagnosis and prognosis breast cancer using classification rules. International

Journal of Engineering Research and General Science. 2014;2(6):315-323.

93. Ritu Tayal, Anshul Tickoo. Performance analysis of regression data mining techniques implemented on breast cancer dataset. International Journal of Latest Trends in Engineering and Technology (IJLTET). 2014;4(1):188-195.
94. Saleema JS, Deepa Shenoy P, Venugopal KR, Patnaik LM. Cancer prognosis prediction model using data mining techniques. Data Mining and Knowledge Engeneering. 2014;6(1).
95. Vikas Chaurasia, Saurabh Pal. A novel approach for breast cancer detection Using data mining techniques. International Journal of Innovative Research in Computer and Communication Engineering. 2014; 2(1).
96. Bichen Zheng, Sang WonYoon, Sarah S. Lam. Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. Expert Systems with Applications. 2014;41(4):1476-1482.
DOI:org/10.1016/j.eswa.2013.08.044.
97. Zakaria Suliman Zubi, Rema Asheibani Saad. Improves treatment programs of lung cancer using data mining techniques. Journal of Software Engineering and Applications. 2014;7:69-77.
98. Vikas Chaurasia, Saurabh Pal. Data mining techniques: To predict and resolve breast cancer survivability. International Journal of Computer Science and Mobile Computing IJCSMC. 2014;3(1):10-22. DOI:org/10.4236/jsea.2014.72008
99. Chihyun Park, Jaegyoon Ahn, Hyunjin Kim, Sanghyun Park. Integrative gene network construction to analyze cancer recurrence using semi-supervised learning. PLoS ONE. 2014;9(1).
100. Aniket Bochare, Aryya Gangopadhyay, Yelena Yesha, Anupam Joshi,Yaacov Yesha. Integrating domain knowledge in supervised machine learning to assess the risk of breast cancer. International Journal of Medical Engineering and Informatics. 2014;6(2).
DOI:10.1371/journal.pone.0086309
101. Joana Diz, Goreti Marreiros, Alberto Freitas. Using data mining techniques to support breast cancer diagnosis. new contributions in information systems and technologies. Advances in Intelligent Systems and Computing. Springer, Cham. 2014;353.
DOI:org/10.1007/978-3-319-16486-1_68
102. Tapas Ranjan Baitharu, Subhendu Kumar Pani. A comparative study of data mining classification techniques using lung cancer data. International Journal of Computer Trends and Technology (IJCTT). 2015; 22(2):91-95.
103. Peter Adebayo Idowu, Kehinde Oladipo Williams, Jeremiah Ademola Balogun, Adeniran Ishola Oluwaranti. Breast cancer risk prediction using data mining classification techniques. Transactions on Networks and Communications. 2015; 3(2):1-11.
104. Arutchelvan K, Periyasamy R. Cancer prediction system using datamining techniques. International Research Journal of Engineering and Technology (IRJET). 2015;2(8):1179-1183.
DOI: 10.14738/tnc.32.662
105. Jaimini Majali, Rishikesh Niranjan, Vinamra Phatak, Omkar Tadakhe. Data mining techniques for diagnosis And prognosis of cancer. International Journal of Advanced Research in Computer and Communication Engineering. 2015;4(3): 613-616.
106. Htet Thazin Tike Thein, Khin Mo Mo Tun. An approach for breast cancer diagnosis classification using neural network. Advanced Computing: An International Journal (ACIJ). 2015;6(1):1-11.
DOI 10.17148/IJARCCE.2015.43147
107. Kindie Biredagn Nahato, Khanna Nehemiah Hari chandran, Kannan Arputharaj. Knowledge mining from clinical datasets using rough sets and back propagation neural network. Computational and Mathematical Methods in Medicine, Hindawi Publishing Corpora tion; 2015.
DOI:10.5121/acij.2015.6101
108. Neelam Singh, Santosh Kumar Singh Bhadauria. Early detection of cancer using data mining. International Journal of Applied Mathematical Sciences. 2016; 9(1): 47-5.
DOI.org/10.1155/2015/460189.
109. Padmapriya B, Velmurugan T. Classification algorithm based analysis of breast cancer data. International Journal of Data Mining Techniques and Applications. 2016;5(1):43-49.
110. Diz J, Marreiros G, Freitas A. Applying data mining techniques to improve breast

cancer diagnosis. Journal of Medical Systems. 2016;9.

111. Divya Chauhan, Varun Jaiswal. An efficient data mining classification approach for detecting lung cancer disease. International Conference on Communication and Electronics Systems (ICCES). IEEE. Coimbatore, India; 2016. DOI:10.1109/CESYS.2016.7889872

112. Christopher T, Jamera Banu J. Study of classification algorithm for lung cancer prediction. International Journal of Innovative Science, Engineering & Technology. 2016;3(2):42-49.

113. Hiba Asria, Hajar Mous annif, Hassan Al Moatassime, Thomas Noel. Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Computer Science. 2016;83:1064-1069.
DOI:org/10.1016/j.procs.2016.04.224

114. Shokoufeh Aalaei, Hadi Shahraki, Alireza Rowhanimanesh, Saeid Eslami. Feature selection using genetic algorithm for breast cancer diagnosis: experiment on three different datasets. Iranian Journal of Basic Medical Sciences. 2016;19:476-482.

115. Mona Nasr, Amr Atif Abd El-Mageed. Using Image mining techniques For optimizing the treatment methods of lung cancer. Journal of Multidisciplinary Engineering Science and Technology (JMEST). 2016;3(1):3613-3620.

116. Zahraa N. Shahweli, Ban N. Dhannoon. Neural network with new relief feature selection for predicting breast cancer based on tp53 mutation. International Research Journal of Computer Science (IRJCS). 2017;12(4):7-11.

117. Zahraa Naser Shahweli, Ban Nadeem Dhannoon. In Silico Model for Lung Cancer Prediction Based on TP53 mutations using Neural Network.1st Scientific International Conference, College of Science, Al-Nahrain University. 2017;195-201.
DOI: 10.26562/IRJCS.2017.DCCS10080.

118. Mohamed Sajid P, Rajesh A. Automatic early detection of skin cancer using Neural Network. Journal of Adv Research in Dynamical & Control Systems. 2017; 4:286-293.
DOI:10.22401/SIC.I.26

119. Mehrbakhsh Nilashi, Othman Ibrahim, Hossein Ahmadi, Leila Shahmoradi. A knowledge-based system for breast cancer classification using fuzzy logic method. Telematics and Informatics. 2017;34(4): 133-144.
DOI:org/10.1016/j.tele.2017.01.007

120. Lingyun Gao, Mingquan Ye, Xiaojie Lu, Daobin Huang. Hybrid method based on information gain and support vector machine for gene selection in cancer classification. Genomics Proteomics Bioinformatics. 2017;15(6):389-395.
DOI: 10.1016/j.gpb.2017.08.002.

121. Lingraj Dora, Sanjay Agrawal, Rutuparna Panda, Ajith Abraham. Optimal breast cancer classification using Gauss–Newton representation based algorithm. Expert Systems with Applications. 2017;85(1): 134-145.
DOI:org/10.1016/j.eswa.2017.05.035

122. Subrata Kumar Mandal. Performance analysis of data mining algorithms for breast cancer cell detection using Naïve Bayes, Logistic regression and decision tree. International Journal of Engineering and Computer Science. 2017;6(2):20388-20391.

123. Hanaa Salem, Gamal Attiya, Nawal El-Fishawy. Classification of human cancer diseases by gene expression profiles. Applied Soft Computing. 2017;50:124-134.
DOI:org/10.1016/j.asoc.2016.11.026

124. Nusaibah Kh. Al-Salihy, Turgay Ibrikci. Classifying breast cancer by using decision tree algorithms. : ICSCA '17: Proceedings of the 6th International Conference on Software and Computer Applications. 2017;144-148.
DOI:10.1145/3056662.3056716.

125. Gomathi N1 and Sandhya. Prognosis and diagnosis of breast cancer using interactive dashboard through big data analytics. Biotechnol Indian Journal. 2017; 13(1):128.
DOI: 10.1145/3056662.3056716

126. Sumalatha G, Archana S. A study on early prevention and detection of breast cancer using data mining techniques. International Journal of Innovative Research in Computer and Communication Engineering. 2017;5(6):11045- 11050.

127. Elham Sagheb Hossein Pour, MS, Rohit J. Kate. Stage-specific survivability prediction models across different cancer types. AMIA. Annual Symposium proceedings. 2017;1421–1429.

DOI: 10.15680/IJIRCCE.2017. 0506006.

128. Wafaa K. Shams, Zaw Z. Htike. Oral cancer prediction using gene expression profiling and machine learning. International Journal of Applied Engineering Research. 2017;12(15):4893-4898.
129. Safae Sossi Alaoui, Yousef Farhaoui, Brahim Aksasse. Classification algorithms in Data mining. International Journal of Tomography & Simulation. 2018;31(4).
130. Yomna Omar, Abdullah Tasleem, Michel Pasquier, Assim Sagahyroon. Lung cancer prognosis system using data mining technique. Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies. 2018;5:361-368.
131. Noor Kadhim Ayoob. Breast cancer diagnosis using K-means methodology. Journal of University of Babylon For Pure And Applied Sciences. 2018;26(1):9-16. DOI: 10.5220/0006553703610368
132. PradeepK R, NaveenN C. Lung cancer Survivability prediction based on performance using classification techniques of support vector machines, C4.5 and Naive Bayes algorithms for healthcare analytics. Procedia Computer Science. 2018;132:412-420.
133. Yixuan Li, Zixuan Chen. Performance evaluation of machine learning methods for breast cancer prediction. Applied and Computational Mathematics. 2018;7(4): 212-216.
DOI:org/10.1016/j.procs.2018.05.162
134. Vikas Chaurasia, Saurabh Pal, BB Tiwari. Prediction of benign and malignant breast cancer using data mining techniques. Journal of Algorithms & Computational Technology. 2018;12(2):119-126.
DOI:10.11648/j.acm.20180704.15
135. Davide ChiccoID, Cristina Rovelli. Computational prediction of diagnosis and feature selection on mesothelioma patient health records. PLoS ONE. 2019;14(1). DOI: 10.1177/1748301818756225
136. Ayad Ghany Ismaeel. Diagnose Mutations Causes B-Thalassemia: Biomining method using an optimal neural learning algorithm. International Journal of Engineering and Technology. 2019;8(1-11):1-8.
DOI:org/10.1371/journal.pone.0208737
137. Abien Fred M. Agarap. On breast cancer detection: An application of machine learning algorithms on the Wisconsin Diagnostic Dataset. arXiv:1711.07831v4.2019.
138. Patrizia Ferroni, Fabio M. Zanzotto, Silvia Riondino, Noemi Scarpato, Fiorella Guadagni, Mario Roselli. Breast cancer prognosis using a machine learning approach. Cancers. 2019;11:328.
139. Nguyen Phuoc Long, Kyung Hee Jung, Nguyen Hoang Anh, Hong Hua Yan, Tran Diem Nghi, Seongoh Park, et al. An Integrative data mining and omics-based translational model for the identification and validation of oncogenic biomarkers of pancreatic cancer. Cancers. 2019;11(2). DOI:10.3390/cancers11030328
140. Keerthana Rajendran, Manoj Jayabalan, Vinesh Thiruchelvam, Sivakumar V. Feasibility study on data mining techniques in diagnosis of breast cancer. International Journal of Machine Learning and Computing. 2019;9(3):328-333.
DOI:10.3390/cancers11020155
141. Leena Vinmalar F, Kumar Kombaiya A. Prediction of lung cancer using data mining techniques. International Journal of Engineering Research & Technology. 2019;7(1):1-4.
DOI:10.18178/ijmlc.2019.9.3.806
142. Esraa H. Abdelaziz, Sanaa M. Kamal, Khaled El-Bhanasy, Rasha Ismail. The application of data mining techniques and feature selection methods in the risk classification of Egyptian liver cancer patients using clinical and genetic data. International Conference Proceeding Series (ICPS). 2019;200-205.
143. Talha Mahboob Alam, Muhammad Milhan Afzal Khan, Muhammad Atif Iqbal, Abdul Wahab, Mubbashar Mushtaq. Cervical cancer prediction through different screening methods using data mining. International Journal of Advanced Computer Science and Applications. 2019; 10(2):388-396.
144. Kamel SR, Yaghoub Zadeh R, Kheirabadi M. Improving the performance of support-vector machine by selecting the best features by Gray Wolf algorithm to increase the accuracy of diagnosis of breast cancer. J Big Data 6. 2019;90.
145. Keerthana Rajendran, Manoj Jayabalan, Vinesh Thiruchelvam, Sivakumar V. Feasibility study on data mining techniques in diagnosis of breast cancer. International

Journal of Machine Learning and Computing. 2019;9(3):328-333. DOI:org/10.1007/s10586-018-1702-5
146. Saadaldeen Rashid Ahmed Ahmed, Israa Al_Barazanchi, Ammar Mhana, Haider Rasheed Abdulshaheed. Lung cancer classification using data mining and supervised learning algorithms on multi-dimensional data set. Periodicals of Engineering and Natural Sciences. 2019; 7(2):438-447.
147. Taninaga J, Nishiyama Y, Fujibayashi K, Toshiaki Gunji, Noriko Sasabe, Kimiko Iijima, Toshio Naito. Prediction of future gastric cancer risk using a machine learning algorithm and comprehensive medical check-up data: A case-control study. Scientific Reports 9. 2019;12384.
148. Zahraa Naser Shahweli. Deep belief network for predicting the predisposition to lung cancer in TP53 Gene. Iraqi Journal of Science. 2020;61(1):171-177. DOI: 10.24996/ijs.2020. 61.1.19.
149. Niloofar Mohammad zadeha, Reza Safdarib, Farshid Mohammadzadehc. Using intelligent data analysis in cancer care: Benefits and Challenges. Journal of Health Informatics in Developing Countries. 2014;8(2):66-72.
150. Wenya Linda Bi, Ahmed Hosny, Matthew B. Schabath, Maryellen L. Giger, Nicolai J. Birkbak, Alireza Mehrtash, et al. Artificial Intelligence in Cancer Imaging: Clinical Challenges and Applications. CA: A Cancer Journal for Clinicians. 2019;69: 127-157.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/55851