NOVEL MACHINE LEARNING ALGORITHM FOR PREVALENT GENE BIOMARKERS FOR EFFECTIVE CANCER TREATMENT BY DETECTING ITS PH

Sahil Sudhakar Patil[1],Darshit Shetty[2],Vaibhav S. Pawar[*3,*4]

[1]Masters Student at Hof University of Applied Science
[2]MBA Student from Mumbai University, JBIMS
[*3]Associate Professor, Mechanical Engineering, Annasaheb Dange College of Engineering & Technology (ADCET), Ashta, Sangli, Maharashtra, India
[*4]PhD (Structures, IIT Bombay) (2013-2019), Graduated in August 2019

ABSTRACT

Patterns discovered from based on collected molecular profiles of patient tumour samples, and also clinical metadata, could be used to provide personalized cancer treatment to patients with similar molecular subtypes. Computational algorithms for cancer diagnosis, prognosis, and therapeutics that can recognize specific functions and aid in classifiers based on a plethora of publicly accessible cancer research outcomes are needed. Machine learning, a branch of artificial intelligence, has a great deal of potential for problem solving in cryptic cancer datasets, as per a literature study. We focus on the new state of machine learning applications in cancer research in this study, illustrating trends and analysing major accomplishments, roadblocks, and challenges along the way to clinic implementation. In the context of non-invasive treating cancer using diet-based and natural biomarkers, we propose a novel machine learning algorithm.

KEYWORDS

Biomarkers, Machine learning, Statistical Models, sequencing, pH sensing.

1. INTRODUCTION

There has been a continuous improvement in Cancer research over the past decades. Scientists used various methods, such as early-stage screening, to detect cancer types before they cause symptoms. They've also developed new strategies for predicting cancer treatment outcomes early on. Due to the introduction of new technologies a large amount of cancer data is available to the medical community. The accurate prediction of a disease outcome, on the other hand, is one of the most interesting and difficult tasks for physicians. As a result, medical researchers are increasingly using machine learning methods.

We present a study that use machine learning methods in cancer prediction and prognosis, in light of the growing trend of applying these methods to cancer prediction and prognosis. Prognostic and predictive features are considered in these studies, which may be independent of a specific treatment or are combined to guide cancer patient therapy [2]. Furthermore, we discuss the types

David C. Wyld et al. (Eds): CCSIT, SIPP, PDCTA, AISC, NLPCL, BIGML, NCWMC - 2021
pp. 155-170, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.110713

of machine learning methods used, the types of data they integrate, and the overall performance of each proposed scheme, as well as their benefits and drawbacks.

Integration of mixed data, such as clinical and genomic data, is a clear trend in the proposed works. However, we noticed a common problem in several works: the lack of external validation or testing of their models' predictive performance. It is clear that using machine learning methods to predict cancer susceptibility, recurrence, and survival could improve accuracy. According to [3,] the accuracy of cancer prediction outcome has improved by 15%–20% in recent years thanks to the use of machine learning techniques.

2. BIOMARKERS

Biomarkers (short for biological markers) are biological indicators of health. "A biomarker is defined as "an objectively measured and examined indicator of normal biochemical functions, pathogenic processes, or pharmacological reactions to a therapeutic treatment."Biomarkers are clinical measurements such as blood pressure or cholesterol levels that are used to monitor and predict health states in individuals and populations in order to plan appropriate therapeutic interventions. Biomarkers can be used individually or in combination to assess a person's health or disease state.

Today, a wide variety of biomarkers are used. Biomarkers are unique to each biological system (for example, the cardiovascular system, metabolic system, or immune system). Many of these biomarkers are simple to measure and are included in routine medical exams. A general health check, for example, might include measurements of blood pressure, heart rate, cholesterol, triglycerides, and fasting glucose. Weight, BMI, and waist-to-hip ratio are all common body measurements used to diagnose obesity and metabolic disorders. An ideal biomarker possesses certain characteristics that make it suitable for assessing a specific disease state. An ideal marker should have the following characteristics: Safe and simple to measure Follow-up is cost-effective. Treatment is modifiable; Gender and ethnic groups are all treated the same.

Biomarkers in disease principles have been applied to cancer detection, screening, diagnosis, treatment, and monitoring. Anti-cancer drugs used to be agents that killed both cancerous and healthy cells. More targeted therapies, on the other hand, have now been developed that can be directed to only kill cancer cells while leaving healthy cells alone. The evaluation of a common cancer biomarker aids in the development of therapies that target the biomarker. This can help to reduce the risk of toxicity while also lowering the cost of treatment. Genetic studies are important in cancer research because genetic abnormalities frequently underpin cancer development. As a result, specific DNA or RNA markers may aid in the detection and treatment of specific cancers.

2.1. Classification of Biomarkers

Biomarkers can be classified as follows:

- Type 0 biomarkers (natural history biomarkers): They aid in determining a disease's natural history and how it correlates with known clinical indicators over time.
- Type 1 biomarkers (drug activity biomarkers): These indicate the effect of drug intervention. They may be further divided[3] into:-

1. Efficacy biomarkers – describing a drug's therapeutic effects
2. Mechanism biomarkers – providing information on a drug's mechanism of action
3. Toxicity biomarkers – a symbol for a drug's toxicological effects

- Type 2 biomarkers (surrogate markers): They can be used to predict the effect of a therapeutic intervention and can also be used to replace a disease's clinical outcome.

Cytokine	Sample	Current purpose as a biomarker	Current well-known function in lung cancer
IL-6	Blood, BALF, and pleural effusion	Diagnostic, prognostic, and predicting the treatment response	Prooncogenic
IL-8	Blood, BALF, and sputum	Diagnostic and prognostic	Prooncogenic
VEGF	Blood, BALF, sputum, pleural effusion, and tissue	Diagnostic and prognostic	Prooncogenic
TNF-α	Blood	Predicting the treatment response	Prooncogenic and antitumor, depending on the context
IL-2	Blood	Prognostic and predicting the treatment response	Not yet determined
IL-18	Blood, BALF, and sputum	Diagnostic	Not yet determined
IL-10	Blood	Diagnostic	Prooncogenic and antitumor, depending on the context
IL-13	Blood	Diagnostic	Not yet determined
IL-22	Blood	Diagnostic and prognostic	Prooncogenic
IFN-γ	Blood	Diagnostic and prognostic	Not yet determined
IL-32	Tissue	Prognostic	Prooncogenic
IL-37	Tissue	Prognostic	Antitumor

Figure 1.Types of Biomarkers

3. MACHINE LEARNING

Machine learning is a branch of artificial intelligence characterized as a software program that can learn quickly by completing a set of tasks. There are three crucial aspects to consider. Describe how machine learning works. Tasks, experience, and performance are three of these factors. Tasks are datasets that are used to train the computer in order to improve its performance. With time and practice, the computer system can refine its model and predict the answer to a topic based on what it has learned from previous attempts. Machine learning employs a variety of algorithms, but they are divided into two categories: supervised and unsupervised learning. The supervised learning group includes any method that uses a set of training data. Each example has an input and output object in the dataset. The algorithm must work on manually entered answers in order to classify the result. This method of working is extremely reliant on the training data. As a result, the set must be correct for the algorithm to understand the data. The algorithm finds undetected patterns in a large amount of data in unsupervised learning. In this method, the computer algorithm is allowed to run and see what patterns will emerge as a result. As a result, there is no clear answer that can be considered correct or incorrect. There are dependent and independent variables in machine learning. The values that control the experiment are stored in the independent variables, which are also known as predictors or control input. The independent variables control the dependent variables, also known as output values.

3.1. Deep learning Architectures

Machine learning includes a subset called deep learning. It's a method of learning that works with multi-level layers and progresses to a more abstract level. The term "deep" refers to the multiple layers of nodes in a neural network. Based on the output from the previous layer variables, each layer in the network was trained on a different feature. The architecture of deep learning is based on neurons and is inspired by the layout of the human brain. A large number of neurons are connected in the human brain, forming a network of communication via signals received. An

artificial neural network is the name given to this idea (ANN). The algorithm in ANN creates layers that pass input values from one layer to the next, eventually resulting in a result.

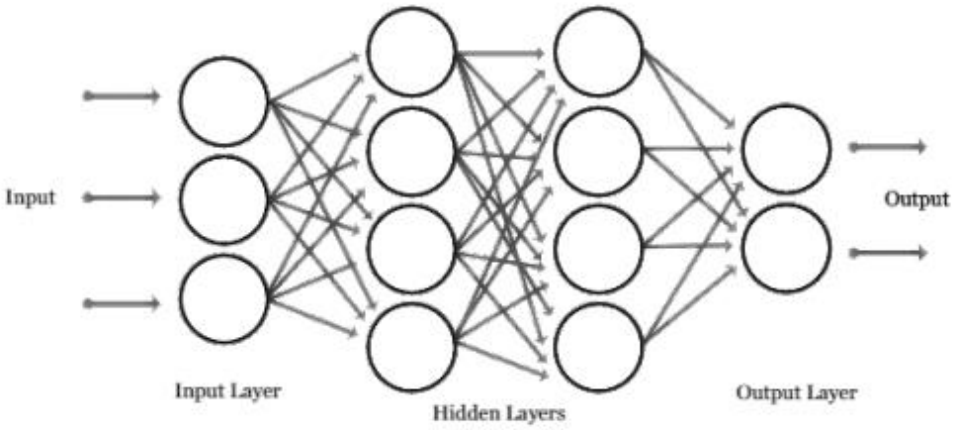

Figure 2.Architectures

Humans do not interfere with the layers of a neural network or the information being processed when using deep learning. Because the system algorithms are trained using data and learning procedures, it does not require human intervention. The method gains the ability to handle data with more dimensions.

3.2. Pre-processing of Genomic Data

Many algorithms can handle vector-matrix data, but converting DNA sequences to matrices is a different story. Values in genomic data are not supposed to be processed as standard text, so the data must be converted into a format suitable for the model. Label encoding and one-hot encoding, which converts nucleotide bases into numerical matrix form with 4-dimensional vectors, are used to accomplish this. Label Encoder converts the input into numerical labels between a value of 0 and N-1 using the Sklearn library ().The one-hot encode method avoids creating a hierarchy problem for the model with the label encode data by using Sklearn's one-hot encoding ( ) function. It changes the sequence by dividing the values into columns and converting them to binary numbers with only 0 and 1 values. This is done because the deep learning algorithm can't work with categorical data or words directly, so transforming input values makes the data more expressive and allows the algorithm to perform logical operations.

3.3. Pre-processing of image data

Image processing is a method of manipulating images to enhance or extract useful information from them so that an AI model can process them. The math function (x,y), where x and y are the image coordinates [40], defines an image as a two-dimensional array of numbers. Pixel values ranging from 0 to 255 are represented by the array numbers. The image height, colour scale, width, and number of levels/pixel are all image input parameters. Red, green, and blue (RGB) colour scales are also known as channels. The first step in pre-processing is to make sure that all of the images are the same size. Cropping the images allows you to change the size. The next step is to resize the photos once all of them have the same aspect ratio. Using a variety of library functions, they can be up scaled or downscaled. They're also normalized to ensure that the data distribution is consistent. The pixel values have been normalized to be between 0 and 1. This is because a network processes inputs using weight values, and smaller values can speed up the

learning process. The size of the image can also be reduced by converting the RGB channel into a grey scale image.

4. MACHINE LEARNING METHODS

The six machine learning techniques of K-nearest neighbour (KNN), Nave Bayes, Ada Boost, Support Vector Machine (SVM), Random Forest, and Neural Network with 10-cross fold technique were used to predict early lung tumours based on metabolomics biomarker features. The classification algorithm SVM aims to create a decision boundary between two categories that allows labels to be predicted from feature vectors [10]. When there is little prior knowledge of data, K-nearest-neighbour (KNN) is the preferred selection method, which is an elementary and straightforward nonparametric classification method [11]. The statistical classifier Nave Bayes was used to predict the probability of class membership [13]. It is hypothesized that all variables contribute independently to classification and that the outcome can be used to predict. [14] the goal of a neural network is to simulate the neuron and the human brain. The Neural Network's artificial neuron uses specific input features to assign mathematical weights that can eventually predict some output object.

4.1. K nearest Neighnors [KNN]

The algorithms for K Nearest Neighbors work. As per the data point's neighbor's similar characteristics. This algorithm used significant positive correlation to forecast the value of a new statistic and allocate the value depending as to how highly correlated we points in the training dataset were. It was used to determine whether or not the patient had cancer. This Algorithm is the best example of implementation.

KNN Is a Nonlinear Learning Algorithm

The ability of machine learning algorithms to estimate nonlinear relationships is a second feature that distinguishes them. Models that predict using lines or hyper planes are known as linear models. The model is shown as a line drawn between the points in the image. The linear model y = ax + b is the most well-known example. In the diagram below, you can see how a linear model could fit the example data.

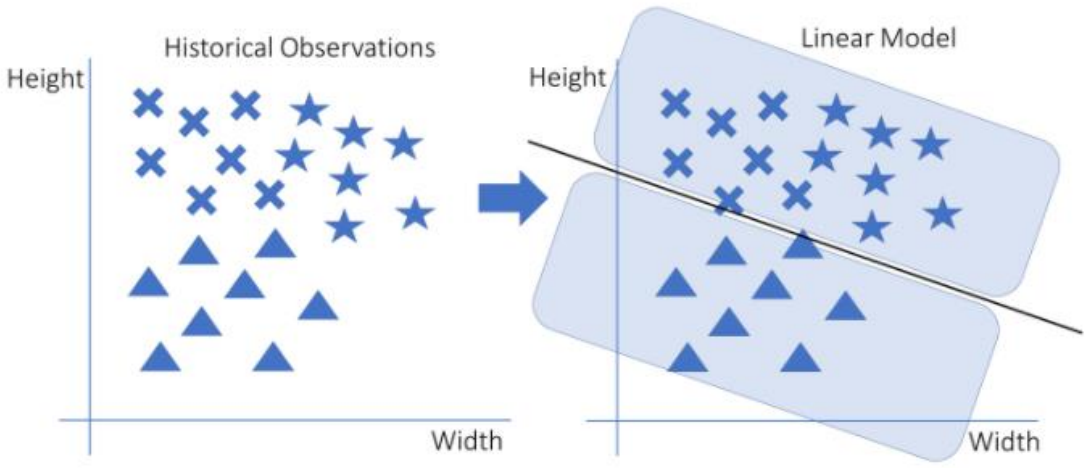

Figure 3. Linear Model

The data points are represented on the left with stars, triangles, and crosses in this illustration. A linear model on the right can distinguish triangles from non-triangles. Every non-triangle point is above the line, and every triangle point is below the line. If you wanted to add another independent variable to the previous graph, you'd have to draw it as a separate dimension, resulting in a cube with the shapes inside. A line, on the other hand, would not be able to split a cube into two parts. The hyper plane is the line's multidimensional counterpart. Nonlinear models are those that separate their cases using a method other than a line. The decision tree, which is essentially a long list of if... else statements, is a well-known example. If...else statements in the nonlinear graph would allow you to draw squares or any other shape you wanted. A nonlinear model is applied to the example data in the graph below.

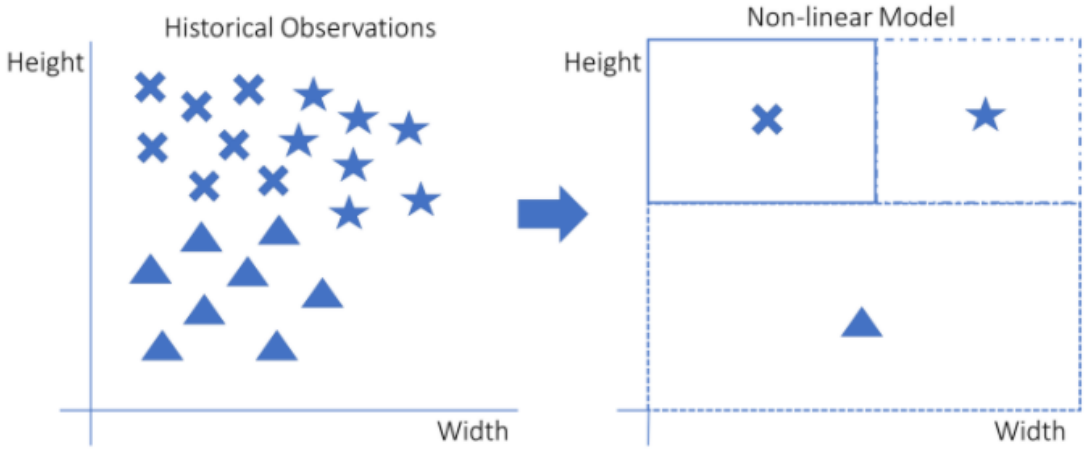

Figure 4. Non-Linear Model

4.2. Support Vector Machine [SVM]

Support Vector Machine (SVM) is a supervised learning classification algorithm broadly used in the development of cancer diagnosis and prognosis. The support vector machine algorithm's goal is to find a hyper plane in an N-dimensional space (N — the number of features) that categorizes data points clearly.

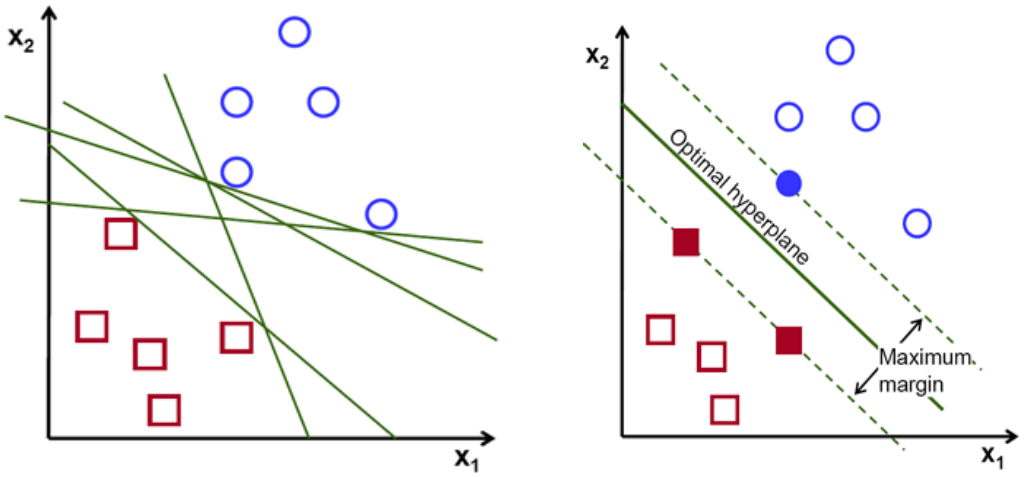

Figure 5. Possible Hyperplanes

There are numerous hyper planes from which to choose to separate the two classes of data points. Our goal is to find a plane with the greatest margin, or the greatest distance between data points from both classes. Trying to maximize the margin distance does provide some reinforcement, making it easier to classify future data points.

Hyper Planes and Support Vectors

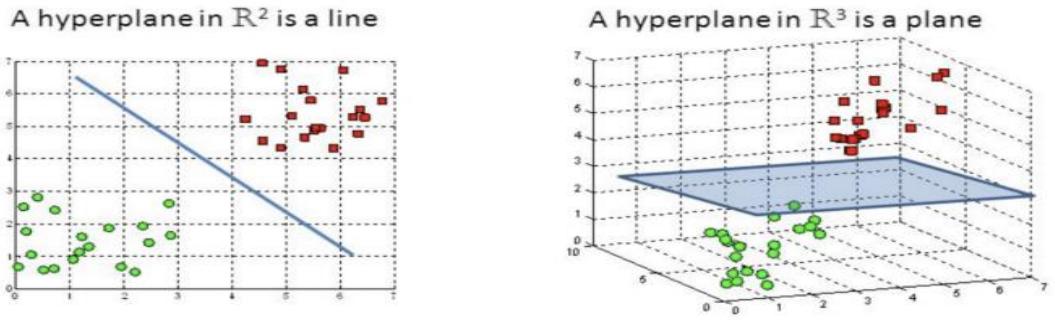

Figure 6. Hyperplanes in 2D and 3D feature space

Hyper planes are operating rules that aid in data classification. Different classes can be assigned to data points on each side of the hyper plane. The Hyper plane's dimension is also determined by the number of features. If there are only two input characteristics, the hyper - plane is just a line.

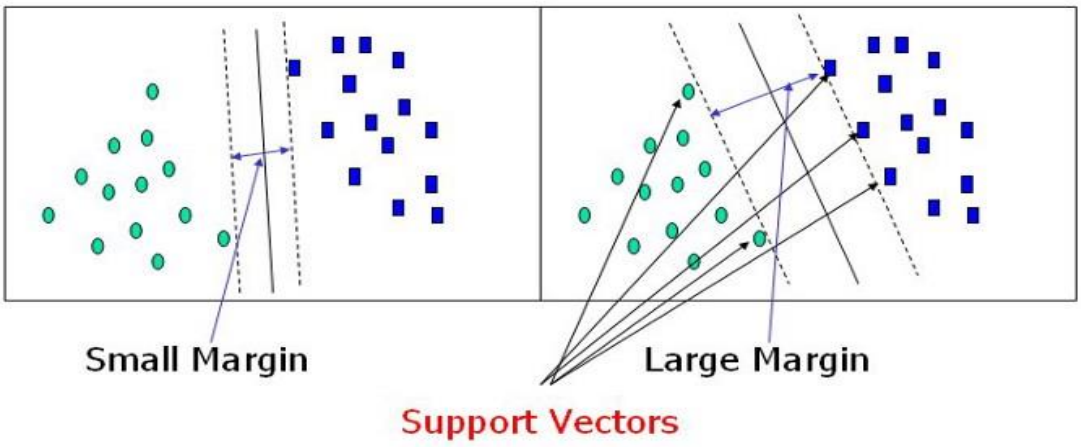

Figure 7. Support Vectors

Support vectors are data points that are nearer to the hyper plane and have an impact on the hyper plane's coordinate system. We significantly increase the classifier's margin by using these support vectors. The hyper plane's position will be altered if the support vectors are deleted. These are all the points that will assist us in constructing our SVM.

4.3. Naïve Bayes [NB]

Navie Bayes is a classification technique that relies on the Bayes theorem with independence among predictor variables. This particular feature in a class of features in a particular class is not

linked to the occurrence of any other features. If all these characteristics are reliant on each other, then these properties add value to the possibility of the class individually, which is the main reason for calling this “Naïve”. Navie Bayes is called "Nave" because it assumes that the characteristics of a measurement are independent of one another. Bayes is naive because he is almost never correct. It's simple to bold, and it's especially useful for large data sets. Naive Bayes is a sophisticated classification method that is known for its simplicity. The Naive Bayes model is simple to construct and is especially useful for large data sets. Naive Bayes is known to outperform even the most sophisticated classification methods due to its simplicity. The Bayes theorem allows you to calculate posterior probability P(c|x) from P(c), P(x), and P(x|c) using P(c), P(x), and P(x|c). Take a look at the following equation.

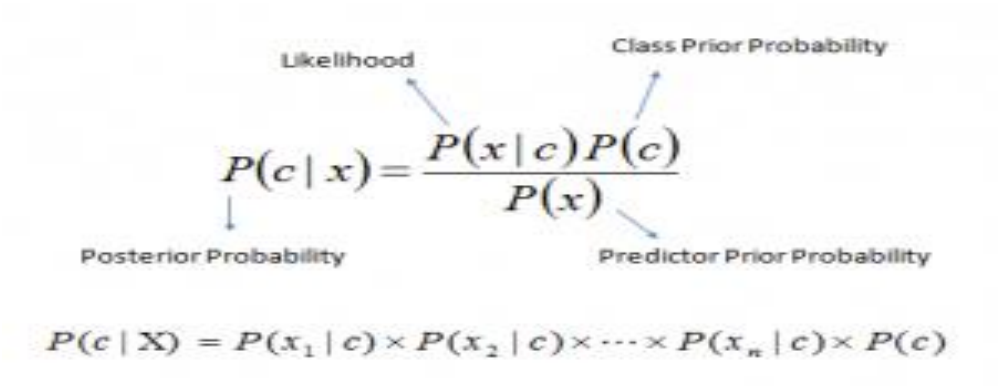

Figure 8. Bayes Theorem

4.4. Logistic Regression [LR]

It's a categorical dependent variable that's used in a classification algorithm. The goal of Logistic Regression is to discover a link between features and the likelihood of a specific outcome. Binomial Logistic Regression is a type of problem in which the response variable has two values: 0 and 1, or pass and fail, or true and false. When the response variable can have three or more possible values, Multinomial Logistic Regression is used. The basic components of a machine learning method are data samples. Every sample has a number of features, out of which each has a different set of values. Furthermore, knowing what type of data will be used ahead of time allows for proper tool and technique selection for their analysis. Some data-related issues concern the data's quality and the steps taken to prepare it for machine learning. Noise, outliers, missing or duplicate data, and biased-unrepresentative data are all examples of data quality issues. When data quality is improved, the quality of the resulting analysis is usually improved as well. Additionally, pre-processing steps focusing on data modification should be used to improve the raw data's suitability for further analysis. There are a variety of data pre-processing techniques and strategies that focus on modifying the data for better fit in a specific ML method. Some of the most important techniques are I dimensionality reduction, (ii) feature selection, and (iii) feature extraction. Dimensionality reduction has numerous advantages when datasets have a large number of features. When the dimensionality of the data is low, ML algorithms perform better [15]. Reduced dimensionality can also eliminate irrelevant features, reduce noise, and produce more robust learning models because fewer features are involved. The process of reducing dimensionality by selecting new features that are a subset of the old ones is known as feature selection. For feature selection, there are three main approaches: embedded, filter, and wrapper [15]. In the case of feature extraction, the initial set of features can be used to create a new set of features that captures all of the important information in a dataset. The creation of new sets of features allows the benefits of dimensionality reduction to be gathered.

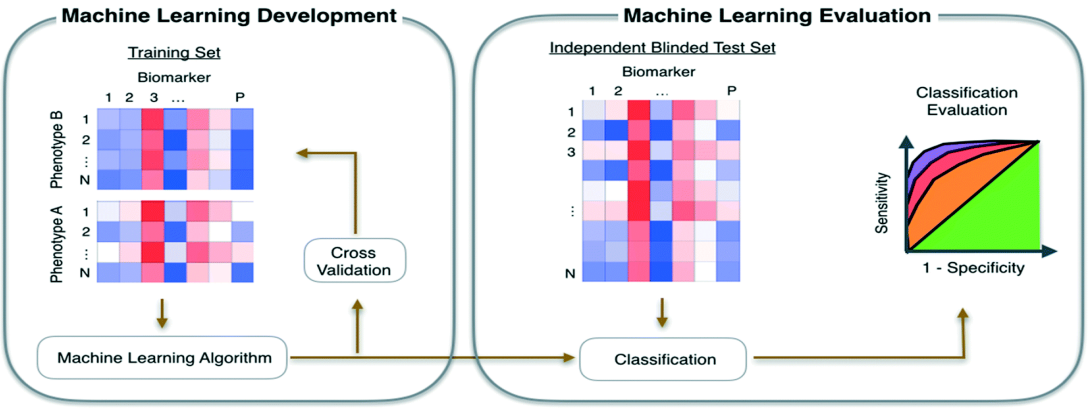

Figure 9.Machine Learning Evaluation

The use of feature selection techniques, on the other hand, may cause specific fluctuations in the creation of predictive feature lists. Several studies have been published that discuss the lack of agreement between different groups' predictive gene lists, the need for thousands of samples to achieve desired results, the dangers of information leaks and the lack of biological explanation of forecasting signatures.

The primary goal of machine learning techniques is to create a model that can be used for classification, prediction, estimation, or any other task. Classification is the most common task in the learning process. This learning function, as previously stated, classifies the data item into one of several predefined classes. Training and generalization errors can occur when ML techniques are used to create a classification model. The former refers to training data misclassification errors, while the latter refers to expected testing data errors. A good classification model should be able to accurately classify all of the instances in the training set. The phenomenon of model over fitting occurs when a model's test error rates begin to rise while its training error rates fall. This situation is related to model complexity, which means that as the model complexity increases, the training errors of the model will decrease. Obviously, the ideal complexity of a non-over fitting model is the one that produces the smallest generalization error. The bias–variance decomposition is a formal method for analysing a learning algorithm's expected generalization error. The error rate of a learning algorithm is measured by the bias component of that algorithm. Variation in the learning method is a second source of error that affects all possible training sets and test sets of a given size. The sum of bias and variance, referred to as the bias–variance decomposition, is the overall expected error of a classification model.

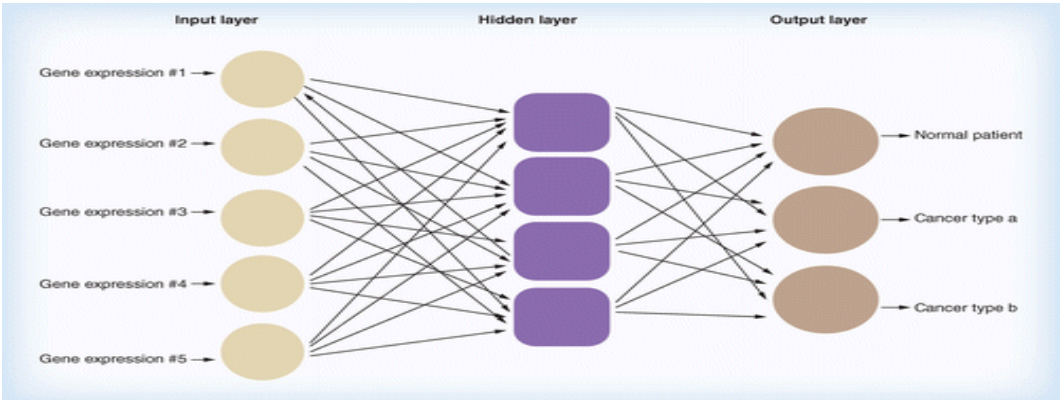

Figure 10. Classification of Gene Expression

5. PATIENT DATA

- Image
- What is age
- Previous background of illness /heredity
- Previous medication for disease
- Immediate history
- Whether any disorder since long
- Significant change in diet
- Different habits such as smoking
- Typical symptoms from start
- Severity of symptoms
- Medication once symptoms initiated
- Symptoms improved or as they were
- Which medical treatment doctor as prescribed
- After medication new symptoms

6. MACHINE LEARNING APPROACH BY DETECTING ITS PH

Cancerous cells differ from healthy cells in several ways that help distinguish them as dangerous. For example, the pH (acidity level) of a cancerous cell differs from the pH of a healthy cell.

Over the last few decades, immunofluorescence has been widely used in a variety of biological and biomedical applications to visualize specific biological phenomena at the cellular and subcellular levels. Despite the fact that it has a number of disadvantages To begin with, fluorophores have the ability to Phototoxic effects are caused by the generation of reactive oxygen species, which have been shown to have harmful effects, negative consequences for cell physiology and health.

Phototoxic damage can be measured and reduced, but it cannot be eliminated. Furthermore, because antibodies cannot cross the cell membrane, immunofluorescence necessitates a cell fixation step. As a result, any downstream analysis that necessitates the presence of living cells is no longer possible. In addition, some research areas, such as in vitro stem cell and drug discovery studies, require very little cell manipulation. To enable scientists to extract valuable information from living cells, new efficient and sensitive alternative methods are required. Intracellular acidity has been shown to be a useful tool for studying single cells, among other things. Intracellular acidity, in particular, is linked to a variety of physiological processes, including cell migration, division and apoptosis and influences how the entire cellular environment functions by regulating events ranging from enzymatic activity to cytoskeletal structure dynamics. Ten to twelve Physiological pH ranges from 4.7 to 8.0 and abnormal intracellular acidity has been linked to the development of diseases like Alzheimer's and even heat stroke.

White and colleagues recently highlighted the importance of deregulated pH dynamics in cancer initiation, progression, and adaptation. In cancer cells, the intracellular pH is higher than in normal cells, while the extracellular pH is lower. This phenomenon has been seen in the early stages of cancer, with pH differences between intracellular and extracellular environments increasing as the cancer progresses. Increased intracellular pH has been linked to the epithelial-to-mesenchyme transition, which is linked to the initiation of metastatic disease.

To study cellular pH, a variety of methods have been developed, most of which rely on fluorescence indicators and decorated nanoparticles. However, they have drawbacks, such as

complex multi-step protocols for nanoparticle synthesis and functionalization. Photo bleaching, which is known to affect cell physiology, is also a factor that affects fluorescence imaging methods. In 2017, Hou et al. published the first paper on a novel single-cell pH-based imaging method, in which the authors were able to rapidly identify cancer cells using a combination of UV-vis micro-spectroscopy and common pH indicators. Innovative approaches to extracting valuable information from biological and medical images have been enabled by numerous advancements in the field of computer vision. Specifically, ML-based algorithms have been developed to extract multiple features from single cells and even subcellular components, which can then be used to identify complex phenotypes and diagnose diseases.

7. METHODOLOGY

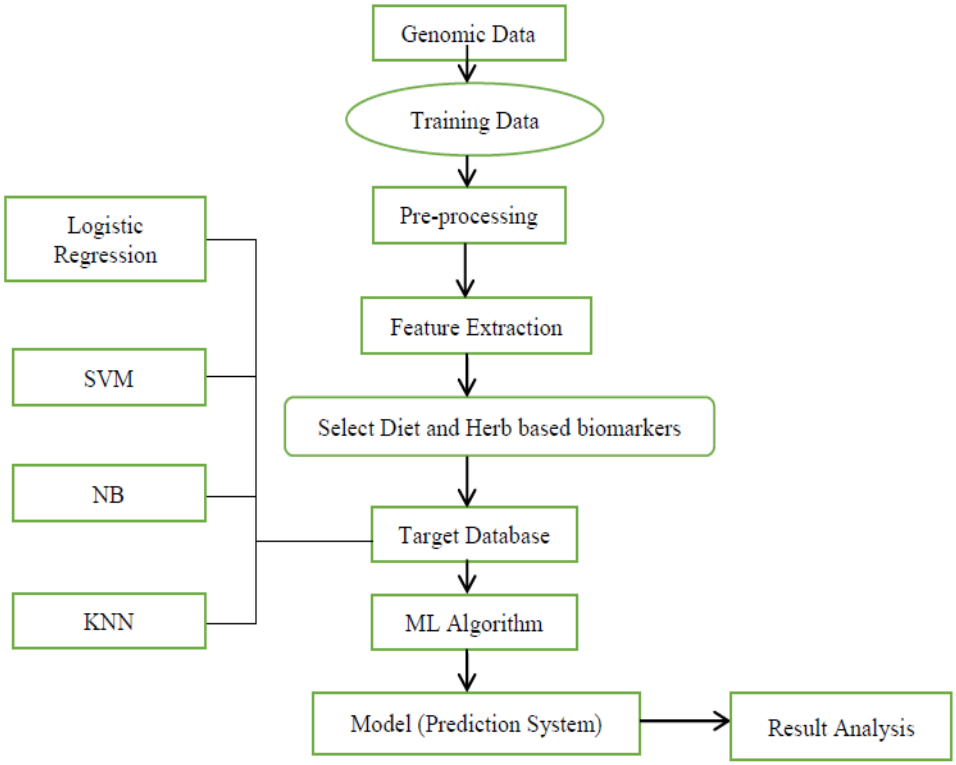

Figure 11. Methodology

8. ALGORITHM STEPS

- Defining the problem and fixing the parameters highly robust and flexible machine learning model
- Deciding severity index and defining mathematical preliminaries
- Assembling of data set
- Choosing a measure of success
- Deciding on an evaluation protocol
- Preparing the data
- Data sorting and cleaning
- Developing a model that does better than a baseline
- Developing a model that over fits and regularizing the model and tuning the parameters

9. CONCLUSIONS

In this study, we looked at machine learning concepts and how they can be used to predict and prognosis cancer. The majority of recent studies have centred on the development of predictive models using supervised machine learning methods and classification algorithms with the goal of accurately predicting disease outcomes. Based on their findings, it's clear that combining multidimensional heterogeneous data with various feature selection and classification techniques can result in promising inference tools in the cancer domain.

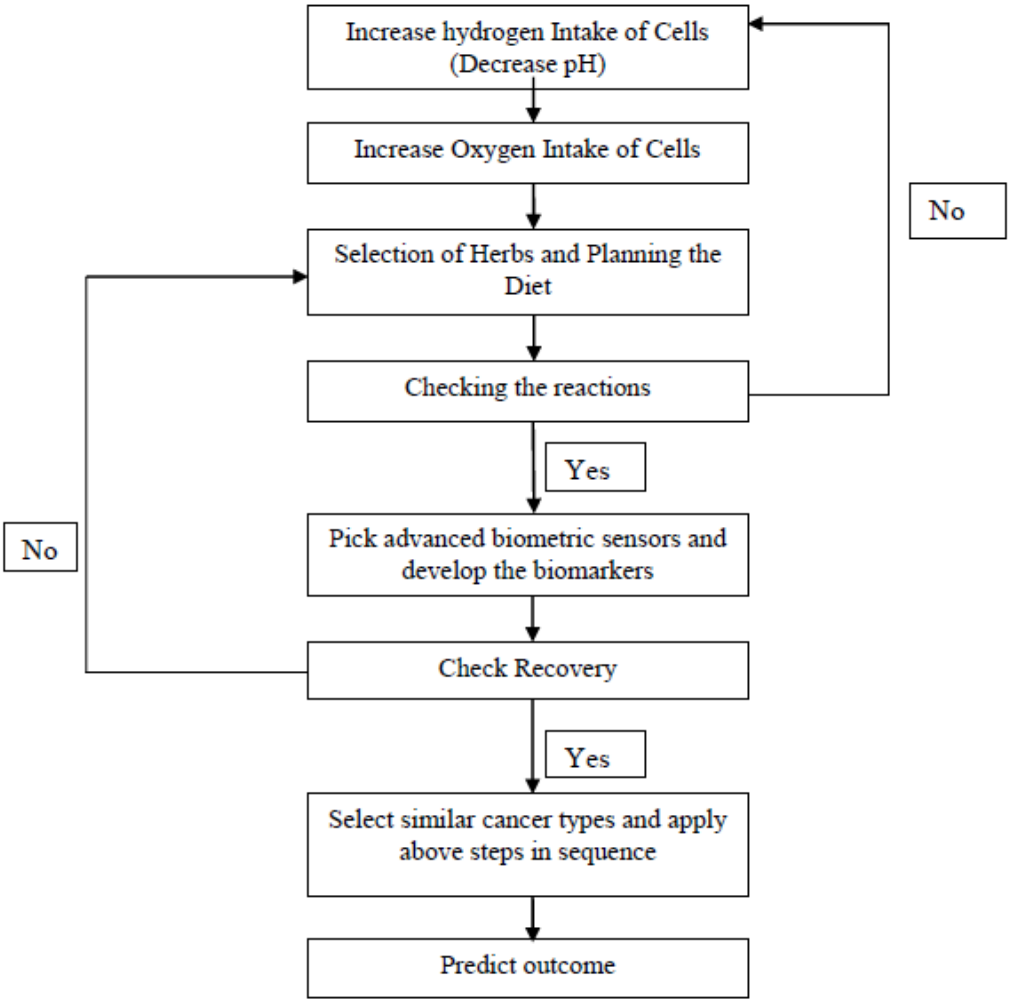

Figure 12.Flowchart for Machine Learning Algorithm

We proposed novel machine learning algorithm as portrayed in the Flowchart above. Idea that is being executed is based upon herb based diagnostics for cancer treatment. Accordingly, deep learning from the infected cells, multi-regression based machine learning model is being developed. Simple biomarkers with non-invasive treatment are being attempted along with appropriate genome sequencing.

10. INITIAL RESULTS

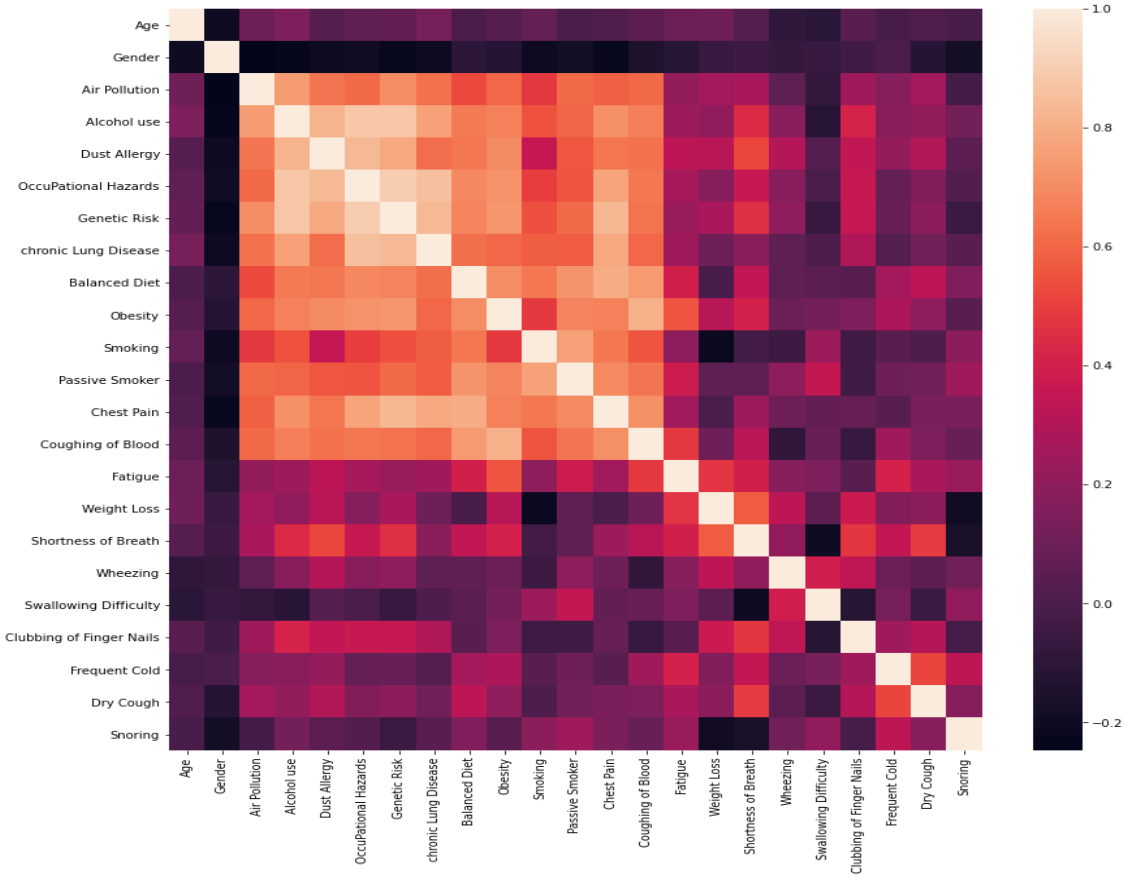

Figure 13. Visual Understanding of Correlation between Patient Characteristics

Data was taken from the publicly available data world for the initial observation of the results. Their sociodemographic characteristics, lifestyle characteristics, and external and internal residential area characteristics were all included in the data. Based on data collected, regression models were created to investigate the links between lung cancer and urban spatial factors.

Individual level factors serve as control variables, attempting to capture the socioeconomic status and lifestyle of surveyed residents, both of which have an impact on their health outcomes. Age, gender, workplace, tobacco use (smoking history and family/colleague smoking status), cooking fume exposure, duration of outdoor exercise, and chronic medical history are all factors to consider.

The findings back up the hypothesis that both indoor and outdoor spatial factors are linked to lung cancer incidence. To revise the criteria for lung cancer screening of high-risk individuals, certain principles based on modeling results are proposed. This will help in the further study to investigate the herbs necessary for the treatment, that will help in maintaining the pH level.

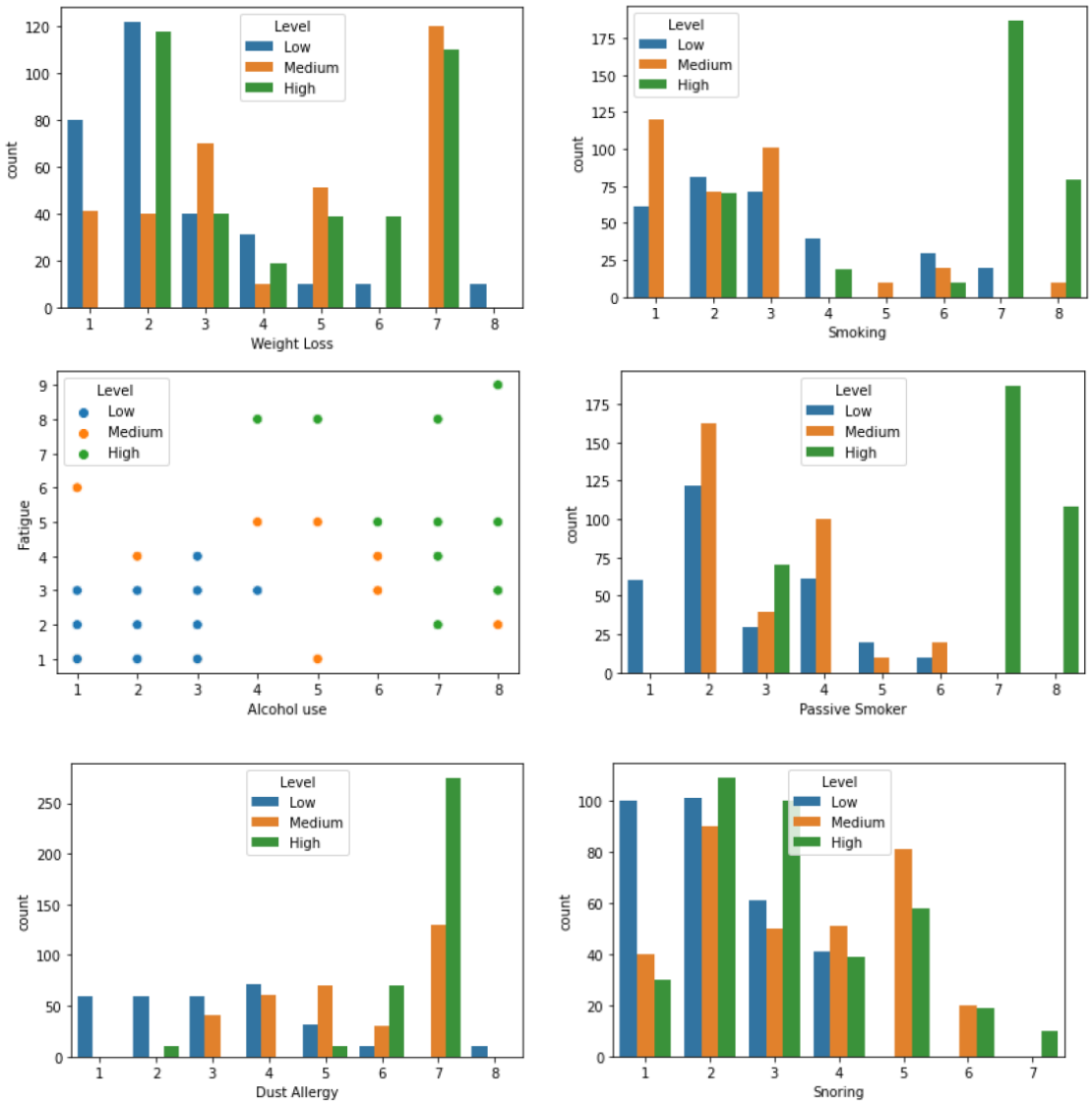

Figure 14. Classification on the basis of lifestyle Characteristics

- Following results were also obtained for various lifestyle characteristics which were classified into low, medium and high level.
- Under Smoking the bar chart illustrates the number of low, medium and high levels of smokers which are divided into 8 groups. It can be seen that from group 1 to 4 there is an undulated trend between low and medium smokers. A sudden increase is seen in high level smokers from group 6 to 7. A fall in medium level smokers is seen in group 5. Overall there is a fluctuating trend in all groups except for group 7 and 8.
- Under Dust allergy the bar chart illustrates the number of low, medium and high levels of people with dust allergy which are divided into 8 groups. A constant trend of low level allergic people is seen between groups 1- 3. There is rise in people with medium level dust allergy between groups 3 - 5. Though only few people with low dust allergy are seen in group 8. A sudden increase in group 7 with high level allergic people is striking. Overall low level allergic people are less as compared to high and medium levels.

- Under Snoring the bar chart illustrates the number of low, medium and high levels of people with snoring disorder which are divided into 8 groups. A constant decrease in low and high level of snoring disorder is seen between group 1-4 and 2-4 respectively. There is a dip in medium level people from group 2 - 5. Overall group 5, 6 and 7 have nil low level people and least high level people with snoring disorder.
- Overall it is observed that, each parameter has fluctuating effect over the patient number count. It is indeed essential to have the combine effect of all the parameters via multi-regression study. Furthermore, diet based and herb based studies need to be attempted in the context of pH and overall health and outcome of Cancer patients. These things are being accommodated in the proposed model suggesting the importance of Novel Machine learning algorithm with herb and diet based biomarkers along with the existing scheme.

REFERENCES

[1] D. Hanahan, R.A. Weinberg, Hallmarks of Cancer: the next generation Cell, 144(2011), pp. 646-674
[2] M.Y.C. Polley, B. Freidlin, E.L. Korn, B.A. Conley, J.S. Abrams, L.M. Mc hane, Statistical and practical considerations for clinical evaluation of predictive biomarkers, J Natl Cancer Inst, 105 (2013), pp. 1677-1683.
[3] J.A. Cruz, D.S. WishartApplications of machine learning in cancer prediction and prognosis, Cancer Informat, 2 (2006), p. 59
[4] O. Fortunato, M. Boeri, C. Verri, D. Conte, M. Mensah, P. Suatoni, et al. Assessment of circulating microRNAs in plasma of lung cancer patients, Molecules, 19 (2014), pp. 3038-3054
[5] H.M. Heneghan, N. Miller, M.J. Kerin, MiRNAs as biomarkers and therapeutic targets in cancer, CurrOpinPharmacol, 10 (2010), pp. 543-550
[6] D. Madhavan, K. Cuk, B. Burwinkel, R. Yang, Cancer diagnosis and prognosis decoded by blood-based circulating microRNA signatures, Front Genet, 4 (2013)
[7] K. Zen, C.Y. Zhang, Circulating microRNAs: a novel class of biomarkers to diagnose and monitor human cancers, Med Res Rev, 32 (2012), pp. 326-348
[8] S. Koscielny, Why most gene expression signatures of tumors have not been useful in the clinic, SciTransl Med, 2 (2010) [14 ps12-14 ps12]
[9] S. Michiels, S. Koscielny, C. Hill, Prediction of cancer outcome with microarrays: a multiple random validation strategy, Lancet, 365 (2005), pp. 488-492
[10] Y. Sun, S. Goodison, J. Li, L. Liu, W. Farmerie, Improved breast cancer prognosis through the combination of clinical and genetic markers, Bioinformatics, 23 (2007), pp. 30-37
[11] L. Bottaci, P.J. Drew, J.E. Hartley, M.B. Hadfield, R. Farouk, P.WR. Lee, et al., Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions, Lancet, 350 (1997), pp. 469-472
[12] P.S. Maclin, J. Dempsey, J. Brooks, J. Rand, Using neural networks to diagnose cancer, J Med Syst, 15 (1991), pp. 11-19
[13] R.J. Simes, Treatment selection for cancer patients: application of statistical decision theory to the treatment of advanced ovarian cancer, J Chronic Dis, 38 (1985), pp. 171-186
[14] M.F. Akay, Support vector machines combined with feature selection for breast cancer diagnosis, Expert SystAppl, 36 (2009), pp. 3240-3247
[15] T. Pang-Ning, M. Steinbach, V. Kumar, Introduction to data mining (2006)

AUTHORS

Sahil Sudhakar Patil Bachelors in Mechanical Engineering 2012- 2016, Pursuing Master in Operational excellence at Hof University of Applied Science (2020-2022)

Darshit Shetty Bachelors in Mechanical Engineering 2008-2012; Pursuing Masters in Marketing Management from Mumbai University, JBIMS.

Dr. Vaibhav S. Pawar*, Associate Professor, Mechanical Engineering, Annasaheb Dange College of Engineering & Technology (ADCET), Ashta, Sangli, Maharashtra, India; PhD (Structures, IIT Bombay) (2013-2019), Graduated in August 2019

© 2021 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.