

Machine Learning Frameworks in Cancer Detection

Sabyasachi Pramanik^{1*}, K. Martin Sagayam², and Om Prakash Jena³

¹Department of Computer Science & Engineering, Haldia Institute of Technology, India.

²Department of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India

³Department of Computer Science, Ravenshaw University, India.

Abstract. Cancer has been described as a diverse illness with several distinct subtypes that may occur simultaneously. As a result, early detection and forecast of cancer types have gained essentially in cancer fact-finding methods since they may help to improve the clinical treatment of cancer survivors. The significance of categorizing cancer suffers into higher or lower-threat categories has prompted numerous fact-finding associates from the bioscience and genomics field to investigate the utilization of machine learning (ML) algorithms in cancer diagnosis and treatment. Because of this, these methods have been used with the goal of simulating the development and treatment of malignant diseases in humans. Furthermore, the capacity of machine learning techniques to identify important characteristics from complicated datasets demonstrates the significance of these technologies. These technologies include Bayesian networks and artificial neural networks, along with a number of other approaches. Decision Trees and Support Vector Machines which have already been extensively used in cancer research for the creation of predictive models, also lead to accurate decision making. The application of machine learning techniques may undoubtedly enhance our knowledge of cancer development; nevertheless, a sufficient degree of validation is required before these approaches can be considered for use in daily clinical practice. An overview of current machine learning approaches utilized in the simulation of cancer development is presented in this paper. All of the supervised machine learning approaches described here, along with a variety of input characteristics and data samples, are used to build the prediction models. In light of the increasing trend towards the use of machine learning methods in biomedical research, we offer the most current papers that have used these approaches to predict risk of cancer or patient outcomes in order to better understand cancer.

1 Introduction

The field of cancer research has seen a constant development over the last few decades. Scientists used a variety of techniques, such as early stage screening, to identify various kinds of cancer once they manifest themselves with symptoms. Furthermore, they have created novel methods for the early detection of the outcome of cancer therapy. As a result of the technological advances in the area of panacea, huge quantities of cancer datasets are being gathered and made accessible to the medical scientific community for analysis. The precise prognosis of a disease outcome, on the other hand, is one of the most fascinating and difficult problems that doctors face today. As a consequence, machine learning techniques have grown more popular among medical researchers. These methods are capable of discovering and identifying patterns and connections between them in large datasets, and they are also capable of accurately predicting the future results of a cancer type in humans.

In these research, prognostic and predictive characteristics are taken into consideration that may be independent of a specific treatment or that may be incorporated in order to advise treatment for people with

cancer. In addition, we examine the kinds of machine learning techniques that are being utilized, the kinds of data that are being integrated, the effectiveness and efficiency of each suggested scheme, as well as the advantages and disadvantages of each approach.

The integration of various datasets, like clinical and genomic data, is a clear trend in the suggested works, and this is reflected in the proposed works. However, we have observed a similar issue in many works, which is the absence of constant confirmation or examining of their models' prognostic achievement. This is a problem that we believe is widespread. Evidently, the use of machine learning techniques to cancer responsiveness, repetition, and durability forecast may enhance the precision of cancer risk and survival predictions. According to [1], the reliability of cancer's realization has increased by 15 percent to 20 percent in the past few years as a result of the use of machine learning [2-3] methods.

Numerous researches have been published in the survey, each of which is depending on a distinct strategy that may aid in the diagnosis and prognosis of cancer in its early stages. In particular, these papers discuss methods linked to the screening of moving miRNAs that was shown to be an encouraging category of molecules for

* Corresponding author: author@email.org

the forecast and diagnosis of cancer. The described techniques, on the other hand, have poor sensitivity when it comes to early-stage screening, and they have trouble distinguishing between benign and malignant tumors. Several elements of the prognosis of cancer prognosis depending on gene expression profiles are addressed in [4]. As a result of these research, we now have a better understanding of both the promise and the limits of microarrays in the prognosis of cancer outcome. Despite the fact that gene signatures have the potential to substantially enhance our capacity to predict cancer patient outcomes, little progress has been achieved in their use in clinical performances. Nevertheless, until gene appearance monitoring will be utilized in medical practice, big data specimens and enhanced thorough authorization are required in research with larger data samples. Researches which used ML approaches to forecast cancer prognosis along with treatment are the only ones that are discussed here.

2 Machine Learning Techniques

ML, a subfield of AI, is concerned with the relationship between the issues of understanding from datasets. In general, the learning procedure is divided into two domains: i) estimate of undiscovered relationships in a platform from a supplied dataset, and (ii) application of the approximated needs to anticipate future outcomes of the framework. Additionally, machine learning (ML) has shown to be a fascinating domain of biomedicine exploration with numerous utilizations, in which a tolerable generalization is acquired by finding across an n-dimensional room in accordance with a specified deposit of genetic specimens, utilizing a variety of methods and algorithms. Unsupervised learning [5] techniques, on the other hand, do not offer labelled examples and do not provide any indication of the output throughout the learning process. This means that finding patterns or discovering groupings of data in the input data is the responsibility of the learning scheme/model. This method may be viewed of as a classification issue in the context of supervised learning. The job of categorization relates to a studying experience which divides datasets into a collection of defined categories once it has been classified. Regression and clustering are two additional popular machine learning tasks. Within the case of regression issues, a learning function is used to map the data into a categorical number that can be calculated. On the basis of this method, it is possible to estimate the quantity of a predictor factor for every new sample that is collected. Clustering is a typical unsupervised activity where a framework attempts to identify aggregation of data items for explaining the data elements in the dataset. Depending on the results of the indicated procedure, every newer specimen may be allocated to one of recognized collections based on the similarities in features which they have in common.

Consider the following scenario: we have gathered medical data pertaining to breast cancer and are attempting to determine whether a tumor is benign or malignant depending on its size. It is expected that the ML inquiry would be related to an estimate of whether

the tumor is malignant or not (1 = Yes, 0 = No). When using a machine learning technique, data samples serve as the fundamental building blocks. Every sample is characterized by a number of characteristics, each of which is composed of a variety of various kinds of values. Furthermore, knowing the particular kind of data that will be utilized in advance enables for the judicious choice of tools and methods that may be used to analyze the data in advance. Some data-related problems include the data quality as well as the procedures used to prepare the data to make it more appropriate for machine learning. Data quality problems also include inclusion of noise, incomplete or redundant data, as well as data that is anomalous of the population. The efficiency of the resultant analysis usually improves as a consequence of the improvement in the data quality, as well. Additionally, for creating the raw data further, appropriate for future research, pre-processing procedures must be implemented which are primarily concerned with the alteration of the information. There are a variety of different methods and tactics available that are related to data pre-processing and which are focused on altering the data records in order to better fit it into a particular machine learning approach. There are many significant methods among these approaches, with the most key techniques being (i) data minimization (ii) feature detection along with (iii) feature extraction being among the most important. When dealing with datasets that include a high number of characteristics, there are many advantages to reducing their dimensionality. When the dimensionality of the data is minimal, machine learning algorithms perform better. Additional benefits of decreasing the number of features include that they remove unnecessary features, decrease background noise, and create more robust learning models since fewer features are being used in the model. Feature selection is a broad term that refers to the process of reducing dimensionality by choosing new characteristics that are subsets of the existing ones. There are three primary methods to feature selection, which are as follows: embedding, filter, and wrapper approaches. If we use feature extraction, we may generate a lot of new features from the beginning edition which apprehends various important details in a dataset by starting with the initial set of features. The development of newer group of characteristics permits for the collection of the advantages of dimensionality reduction that have been discussed.

Obviously, the optimum intricacy of a model that is not vulnerable to curse of dimensionality is the one which generates the smallest generalization error in the data set. The bias–variance decomposition is a mathematical technique for evaluating the anticipated generalization error of a learning system. In a specific learning algorithm, the bias component is a way of measuring how often the system makes mistakes. Additionally, the variance of a learning technique is a secondary source of error that is spread across all potential teaching sets of a specific dimension as well as various realizable possible test sets, and it is defined as follows: Overall predicted errors of a classifiers is composed of the total of bias and variance, referred to as the bias–variance decomposition. Once each classification model has been

created using one or more machine learning methods, it is necessary to estimate the performance of the classifier. The accuracy and area under the curve of each proposed model are all evaluated as part of the performance evaluation process (AUC).

The percentage of genuine positives that are properly detected by the classifier is described as sensitivity, while the percentage of true negatives which are correctly recognized is defined as specificity. Accuracy and AUC are two quantitative measures that are used to analyze the performance of a classification algorithm. Precision is a measure of how accurate a forecast is in relation to the number of accurate classification.

The Random Sampling, Bootstrap, Holdout Methodology and Cross-Validation are some of the most frequently used techniques for assessing the effectiveness of a classifier by dividing the original labelled data into subsets. Other approaches include: (i) Cross-validation, and (ii) Bootstrap. For Holdout technique, the data sets are divided into 2 distinct sets, which are referred to as the training and test sets, respectively. The training set is used to build a classification model, which is subsequently used to estimate the model's performance on the test set. When compared to the Holdout technique, randomization is a comparable strategy. In this example, the Holdout technique is performed many times, with the training and test cases being chosen at random, in order to get a more accurate estimate of the accuracy. Using a third method, called cross-validation, each item is used for testing the same amount of times as it is used for training.

Depending on the purpose of this scrutiny article, we may only mention to ML techniques which are extensively utilized in the research for the research analysis of cancer diagnosis in order to avoid duplication of effort. We highlight trends in the kinds of machine learning techniques that are being utilized, the kinds of data that are being incorporated, and the assessment methods that are being used to evaluate the complete effectiveness of the approaches that are utilized for cancer forecast or illness outcome prediction.

ANNs are efficient of controlling a broad span of classification and pattern realization issues. When they are trained, they will produce an output that is a combination of the input variables. This procedure is usually carried out using several hidden layers, each of which represents a mathematical representation of the neural connections. ANNs, despite the fact that they are the standard technique for a variety of classification tasks, they have a number of disadvantages. Their general layered structure proved to be time-consuming, and it may also result in extremely poor performance when used in large quantities. The fact that this particular method is referred to as a "black-box" technology is also noteworthy. It is almost difficult to determine how it conducts the categorization process or even why an ANN failed to perform as expected. In Decision Trees (DT), the input factors are represented by nodes, and the decision outcomes are represented by leaves, which are arranged in a tree-structured classification scheme. DTs are among the oldest and

most famous machine learning techniques, and they have been extensively used for categorization purposes since their inception. In accordance with their architectural design, they are easy to understand and "fast" to master. When we traverse the tree during the categorization of a newer specimen, we will be capable in making educated guesses regarding the sample's categorization. The choices made as a consequence of their unique design allow for sufficient deliberation, which renders them a desirable method to use. Statistical learning models (Support Vector Machines) are a more modern approach to machine learning techniques that have been used in the area of cancer diagnosis and prognosis. First, SVMs transfer an input vector to a higher-dimensional feature space, where they can then find the hyper plane which divides the pieces of information into 2 distinct groups. Increasing the minimal separation connecting the selection hyper plane along with the examples which are nearest to the border will result in the greatest possible benefit. A significant amount of generalizability is achieved by the resultant classifier, which may be utilized for the reliable categorization of fresh data. Notably, SVMs may also provide probabilistic outputs [20], which is something to keep in mind. Support vector machine could be used to categorize tumors as benign or malignant depending on their size and the age of the patients. The hyper plane that has been discovered may be perceived of as a selection border linking 2 different clusters. It goes without saying that the presence of a decision boundary enables for the identification of any misclassifications that may be caused due to the algorithm. Bayes Naive classifiers provide likelihood estimates in preference to prognosis, as opposed to traditional classifiers. As their name implies, directed acyclic graphs are used to express information combined with possible relationships amongst the parameters. Many classification problems, as well as information processing and reasoning challenges, have seen extensive use of Bayesian networks (BNs).

3 Machine Learning and Cancer Prognosis

A number of different machine learning methods along with feature selection approaches is extensively used to illness prognosis during the past two decades [6-10]. The majority of these studies use ML techniques to predict the development of cancer and to find informative variables that may then be used in a categorization later on. More to the point, in nearly all of the research, gene expression profiles are combined with clinical factors and histological characteristics in a complimentary way in order to be given as inputs to the prediction process. Various query search in the Scopus healthcare repository were used to compile the data that was used to compile this report. More specifically, queries such as "cancer risk assessment" AND "Machine Learning," "cancer recurrence" AND "Machine Learning", "cancer survival" AND "Machine Learning", and "cancer prediction" AND "Machine Learning," including "cancer risk assessment" AND

"Machine Learning", produced various number of papers. Except for the rejection of publications before 2015, there were no restrictions placed on the results of the search. While the accuracy of a clinical diagnosis is unquestionably reliant on the accuracy of the clinical investigation, a prognosis forecast must be considered further to the choice to perform a diagnostic procedure. When it comes to cancer prognosis/prediction, there are 3 anticipating jobs to consider: (i) the forecast of clinical outcomes (assessment of risks), (ii) the prognosis of cancer appearance/localized mastery, and (iii) the forecast of cancer durability (survival prediction). For the initial two instances, the goal is in determining (i) the probability of acquiring a certain kind of cancer along with (ii) the chance of redesigning a specific category of cancer following a full or part suspension of the disease. It is the primary goal in this instance to predict a survival result following cancer diagnosis or therapy, like disease specificity or survival rates after cancer diagnosis and treatment. It is customary to refer to the prognosis of cancer result in the context of (i) survival anticipation, (ii) survival, (iii) continuance and (iv) medication awareness.

4 A review of Machine Learning Utilizations in Cancer

A thorough exploration was carried out for information on the utilization of ML approaches in cancer vulnerability, repetition, and survival prognosis. PubMed and Scopus are two of the electronic databases that were visited. Because of the large amount of publications produced by the seeking results, additional scrutiny was required to save the frequently applicable papers. Fortunately, this was accomplished. According to the keywords of the three prognostic jobs identified in the topic titles along with abstracts of each publication, the relevance of each article was determined. After actually examining their titles and abstracts, we narrowed the field down to just those articles that focused on an individual strategies out of the three frameworks of cancer prognosis and mentioned that focus in their topic titles. For the most part, these investigations make utilization of a variety of various kinds of input data, such as genetic, healthcare, diagnosis, radiological and disinfection information, or a fusion of these. Normal analytical approaches (e.g., Cox regression or chi-square) were used to predict cancer development; however, papers which used approaches for tumor categorization or the discovery of predictive variables were removed from consideration. Based on the findings of various researches and their study on machine learning applications in cancer diagnosis, we have seen a significant rise in the number of articles published in the past decade. Despite the fact that it is difficult to include all relevant articles in a single review, we think that a substantial number of important articles were retrieved and are provided in this one. Based on a thorough review of more recent research, it was discovered that there is a rising trend in threat assessment and prediction of recurrence for cancer types, independent of the machine learning

method employed. In recent years, several research organizations have attempted to forecast the likelihood of recurrence of cancer following remission, with the results seeming to be more accurate than predictions made using other statistical methods. Furthermore, the overwhelming majority of these papers made predictions based on genetic and clinical data, which was utilized to support their findings. As a result of the introduction of HTTs, the usage of such quantifiable characteristics as input data is becoming more popular.

4.1 Cancer Susceptibility Testing and Prediction

Using Scopus and PubMed advanced search, we were only able to find articles published within the past five years. One of the papers based on these findings makes use of ML techniques to diagnose cancer threat in a specific cancer type. According to Learning Classifying Systems, the researchers conducted a genetic hygienic research in bladder cancer vulnerability in parameters for Learning Categorizing Systems. Given the fact that it works with genetic data and investigates additional genetic issues, we chose to remove this research from the current case study.

Depending on the restrictions, we narrowed our exploration to biomedical repositories that were relevant to our needs. The majority of these titles made no reference to the specific keywords that were stated in the relevant poll and did not make use of machine learning methods to make their predictions. A recent and highly interesting research in accordance with the breast cancer threat estimate by the techniques of ANNs was chosen from amongst the most recent articles which emerged from our restricted literature search on the cancer threat evaluation prognosis [11-12]. This research differs from the others presented in this journal article in that it uses a different kind of data than the others. Our work, despite the fact that it does not correspond to our general statement about our search criteria, was included in this case study since no other search term fulfilled our requirements. The authors discuss the significant effort that has gone into creating judgment tools which can distinguish between malignant and benign signs in breast cancer, as well as the challenges that remain. Moreover, mentioned is the fact that, while building predictive model, risk stratification is of particular importance. Existing research centered on using computer models, according to their expertise, have also used particular machine learning methods, such as artificial neural networks, in order to evaluate the breast cancer risk patients. In this research, artificial neural networks (ANNs) are used to create a prediction model that can distinguish between malignant and benign mammographic results. A huge number of hidden layers were used in the construction of their model. With regard to the information gathered in this research, 52874 mammographic results, and also demographic risk variables and tumor characteristics, were taken into consideration. A panel of radiologists examined and gathered many of the mammographic data. The dataset was subsequently used as inputs for the

ANN framework, which produced various results. The effectiveness of the system was determined via the use of cross-validation. Additionally, the authors utilized the Elastic Search (ES) method in order to avoid the possibility of over fitting the data. Over fitting is prevented by using this method, which regulates the error message during training and terminates it if it happens. After training and testing their model using k-fold cross validation, they were able to determine the AUC of their model, which was 0.974. The researchers said that their approach, which incorporates a huge data sample, can correctly predict the risk analysis of breast cancer patients. They also said that their model is distinct from others when we consider that the mammography results combined with tumor registry outcomes were the most significant variables they utilized to build the ANN model.

4.2 Estimation of Cancer Recurrence

The most significant and current papers which suggested the utilization of ML techniques in cancer recurrence prognosis are included in this section, depending on our survey results.

It was discovered that researchers could anticipate a potential relapse of oral squamous cell carcinoma (OSCC) by combining data from several sources (clinical, imaging, and genomic) in effort to stop the disease from recurring. In this research, a total of 86 individuals were examined, 13 of them were found to have had a recurrence while the other patients were found to be disease free. Two feature selection algorithms, notably wrapper algorithm and Completely Fair Scheduler (CFS), were used in a particular feature selection process, which was succeeded by the use of the two feature selection approaches. Due to the consequence, any bias can be bypassed while choosing the utmost instructive characteristics of their reference diverse dataset. Then the relevant parameters that have been identified may be utilized as input vectors for particular classifiers. In each category, the overall quantity of medical, imaging [13-14], and genomic characteristics was 65, 17, and 40 respectively before the use of feature selection methods. Following the use of the CFS method, the final outcome of medical, imaging, along with genomic data points utilized in every classification was 9, 7, and 8, respectively, as a result of the usage of the method. More precisely, among the clinical characteristics, the smoking, tumor thickness, and the P53 stain were shown to be the most informative for each classification method, according to the results.

4.3 Forecast of Cancer Survival

Researchers describes the development of a prediction model in assessing the survival of women who are detected with breast cancer, and the significance of soundness in the face of framework parameter dissimilarity. It was determined that three classification models were superior to one another [15] using the SEER cancer database: the SVM, the ANN, and the

SSL. The collection consists of 162,500 entries, each of which has one of 16 essential characteristics. Another statistic was taken into consideration, namely survivability, which referred to the number of patients who had died against the number of patients who had survived. The following characteristics are amongst the most explanatory ones: (i) the size of the tumor, (ii) quantity of sensor nodes, and (iii) the lifetime of diagnosis. They discovered that the estimated accuracy of the SVM, ANN, and SSL approaches was 65 percent, 51 percent, and 71 percent, respectively, when they compared the best performance for each of the three models. The accuracy of the prediction models was evaluated using a five-fold cross validation procedure. Taking note of these results, the researchers recommended that the SSL model be used for survival data by clinical professionals in the future. It is important to notice that the researchers did not specify any preprocessing procedures prior to the collection of the most relevant characteristics in their paper. They then went on to utilize the full SEER datasets, while the box-whisper plot was utilized to estimate the variance in accomplishing over 24 various amalgamations of parameter values. A tiny box area inside a particular approach implies more resilience and stability when variable combinations are used. With its tiny boxes, the SSL model demonstrated that it was more accurate than other models in terms of accuracy. The following year, a noteworthy research was published in which the authors attempted to evaluate the survival prognosis of non-small cell lung cancer (NSCLC) sufferers via the utilization of artificial neural networks [16]. The raw data from NSCLC patients' expression of genes was used to create their dataset. They also tested many different kinds of artificial neural network designs in order to identify the most effective one for cancer survival prediction. The predicted performance of the categorization scheme was found to be 83 percent accurate on average, according to the data given.

5 Results and Discussion

The most current work related to cancer prediction/prognosis using machine learning methods is given in various review articles. After providing a brief overview of the machine learning branch, as well as the techniques of data preprocessing approaches, feature selection methods, along with classification approaches, we presented 3 particular instances involving the forecasting of various cancers, cancer recurrence, and cancer survival using familiar machine learning tools. Clearly, there has been a significant increase in the number of ML published studies during the last ten years which have provided correct outcomes regarding specified forecasting cancer possibilities. But for the retrieval of clinical choices, the identification of possible disadvantages, such as the research setup, the collecting of suitable data samples, and the confirmation of the categorized findings, is essential. Also worth noting is that, despite promises that various ML classification methods may lead to sufficient and effective decision making, only a small number of them

have really made their way into clinical practice. A number of illnesses are now better understood because to recent advancements in omics technology; nevertheless, more precise validation findings are required before gene expression profiles can be used in the clinic to guide treatment decisions. In the research published in the past two years that used semi-supervised machine learning methods for predicting cancer survival, a rising trend has been seen. In [17] it has been shown that using this kind of algorithms, when compared to current supervised methods, the predicted performance is increased. SSL may be seen of as a fantastic replacement to other 2 types of ML approaches (supervised learning along with un-supervised learning), which both rely on a small number of labelled samples in general. The most often mentioned drawbacks in the research included in this evaluation is the limited number of data specimens available. While utilizing classification frameworks to model an infection, one of the most important requirements is that the training datasets be big enough to accommodate the classification scheme. A reasonably big dataset allows for adequate splitting into training and test sets, resulting in acceptable verification of the estimators when the dataset is categorized into training and testing sets. An insufficiently large training specimen as correlated to the data's dimension may effect in misapprehension, while insufficiently large calculations can result in unreliable and prejudiced frameworks. It goes without saying that increasing the number of sufferers who are included in the survival forecast may improve the generalization of the predictive model.

6 Conclusions

In this study, we addressed the principles of machine learning (ML) while also highlighting their use in cancer prediction and prognosis research. In the past few years, several research have been suggested, with the majority of them focusing on the creation of prediction models utilizing supervised machine learning techniques and classification algorithms with the goal of predicting accurate illness consequences. Depending on the examination of their findings, it is clear that the incorporation of multidimensional diverse data, coupled with the use of various methods for feature selection, may result in the development of powerful techniques for inference in the field of cancer.

References

1. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform.* 2007;2:59-77. (2007).
2. A. Mandal, S. Dutta and S. Pramanik, "Machine Intelligence of Pi from Geometrical Figures with Variable Parameters using SCILab", in Methodologies and Applications of Computational Statistics for Machine Intelligence, D. Samanta, R. R. Althar, S. Pramanik and S. Dutta, Eds, IGI Global, DOI: 10.4018/978-1-7998-7701-1.ch003.

3. Y. Meslie, W. Enbeyle, B. K. Pandey, S. Pramanik, D. Pandey, P. Dadeech, A. Belay and A. Saini, "Machine Intelligence-Based Trend Analysis of COVID-19 for Total Daily Confirmed Cases in Asia and Africa", in Methodologies and Applications of Computational Statistics for Machine Intelligence, D. Samanta, R. R. Althar, S. Pramanik and S. Dutta, Eds, IGI Global, DOI: 10.4018/978-1-7998-7701-1.ch009.
4. Koscielny S. Why most gene expression signatures of tumors have not been useful in the clinic. *Sci Transl Med.* (2010) Jan 13;2(14):14ps2. doi: 10.1126/scitranslmed.3000313. PMID: 20371465.
5. A. Bhattacharya, A. Ghosal, A. A. Obaid, S. Krit, V. K. Shukla, K. Mandal and S. Pramanik, "Unsupervised Summarization Approach with Computational Statistics of Microblog Data", in Methodologies and Applications of Computational Statistics for Machine Intelligence, D. Samanta, R. R. Althar, S. Pramanik and S. Dutta, Eds, IGI Global, DOI: 10.4018/978-1-7998-7701-1.ch002, (2021).
6. Cicchetti D. Neural networks and diagnosis in the clinical laboratory: state of the art. *Clin Chem* (1992);38:9-10.
7. Cochran AJ. Prediction of outcome for patients with cutaneous melanoma. *Pigment Cell Res.* (1997) Jun;10(3):162-7. doi: 10.1111/j.1600-0749.1997.tb00479.x. PMID: 9266604.
8. Exarchos K P, Goletsis Y, Fotiadis DI. Multiparametric decision support system for the prediction of oral cancer reoccurrence. *IEEE Trans Inf Technol Biomed* (2012);16: 1127-34.
9. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med.* (2001) Aug;23(1):89-109. doi: 10.1016/s0933-3657(01)00077-x. PMID: 11470218.
10. Sun Y, Goodison S, Li J, Liu L, Farmerie W. Improved breast cancer prognosis through the combination of clinical and genetic markers. *Bioinformatics.* (2007) Jan 1;23(1):30-7. doi: 10.1093/bioinformatics/btl543. Epub 2006 Nov 26. PMID: 17130137; PMCID: PMC3431620.
11. Gilmore S, Hofmann-Wellenhof R, Soyer HP. A support vector machine for decision support in melanoma recognition. *Exp Dermatol.* (2010) Sep;19(9):830-5. doi: 10.1111/j.1600-0625.2010.01112.x. Epub 2010 Jul 11. PMID: 20629732.
12. Mac Parthaláin N., Zwiggelaar R. (2010) Machine Learning Techniques and Mammographic Risk Assessment. In: Martí J., Oliver A., Freixenet J., Martí R. (eds) Digital Mammography. IWDM 2010. Lecture Notes in Computer Science, vol 6136. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-13666-5_90
13. S. Pramanik and S. K. Bandyopadhyay, "Hiding Secret Message in an Image", *International Journal of*

- Innovative Science, Engineering and Technology, vol. 1 pp. 553-559, (2014).
14. S. Pramanik and S. K. Bandyopadhyay, "Image Steganography Using Wavelet Transform and Genetic Algorithm", *International Journal of Innovative Research in Advanced Engineering*, vol. 1 pp. 1-4, (2014).
15. Howlader N, Noone AM, Krapcho M, Garshell J, Neyman N, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). *SEER Cancer Statistics Review, 1975-2010*, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/archive/csr/1975_2010/, based on November 2012 SEER data submission, posted to the SEER web site, April (2013).
16. Chen YC, Ke WC, Chiu HW. Risk classification of cancer survival using ANN with gene expression data from multiple laboratories. *Comput Biol Med.* 2014 May;48:1-7.doi: 10.1016/j.combiomed.2014.02.006. Epub (2014) Feb 22. PMID: 24631783.
17. Park K, Ali A, Kim D, An Y, Kim M, Shin H. Robust predictive model for evaluating breast cancer survivability. *Engl Appl Artif Intell* (2013);26:2194–205.