International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue XI, Nov 2018- Available at www.ijraset.com

Predicting Cancer using Machine Learning Algorithms

Sangeetha D[1], N. Siva Balan [2]
[1, 2]*Department of Computer Science, New Horizon College of Engineering, Bangalore-560048, Karnataka, India*

Abstract: ***Cancer has been characterized as a heterogeneous disease consisting of many different subtypes. The early diagnosis and prognosis of a cancer type (Basal, Melanoma, Lung, Leukemia, and Breast Cancer) have become a necessity in cancer research institutes and hospitals, as it can facilitate the subsequent clinical management of patients. In this work it has been proposed to use a Random Forest, Naïve Bayes algorithm and KNN algorithm to predict the cancer occurrence and Logistic Regression algorithms for validating the model results. Different Variable Prioritization Techniques is also used to perform a preliminary screening of variables and to receive important ranks. Then, the new data set is extracted from initial dataset according to top-k important predictors and is input into the ensemble machine learning algorithms. The model will be validated against the different datasets using k fold validation methods requires a base offer said amid split. For whatever length of time that those base.***

Keywords: Machine Learning, Naive Bayes, Random forest, k-nearest neighbour, K-Fold cross validation.

I. INTRODUCTION

The therapeutic network has been built up to deal with human wellbeing with educated and capable specialists. In the everyday life the medicinal determination for any ailments is the real undertaking. The treatment of concern infections utilizing medicinal pathology is fundamental work. The measurements of blood based dieses is expanding step by step. In the period of medicinal science the a large portion of the dieses is enlightenment is finished utilizing the PC based mechanized framework. Malignant growth has been portrayed as a heterogeneous malady comprising of a wide range of subtypes. The early determination and forecast of a malignant growth type (Basal, Melanoma, Lung, Leukemia, Breast Cancer) have turned into a need in disease investigate organizations and doctor's facilities, as it can encourage the ensuing clinical administration of patients. Which malignant growth will happens in human with symptom's. Skin disease is the most well-known type of malignant growth, with over 5.4 million new cases for each year in the United States. 80% of skin disease events are Basal Cell Carcinoma(BCC), making it the most well-known kind of malignancy on the planet Despite the high pervasiveness of BCC, metastatic development is very exceptional, as is demise . Last announced, the passing rate was evaluated to be 2000 passings for every year in the U.S. In spite of being the most widely recognized disease around the globe, multi year survival rates of skin malignancy are above 95%, as long as they are identified and treated before they have spread. The way to delayed survival is early location, making a mechanized examination apparatus basic. Current machine learning investigations of Basal Cell Carcinoma (BCC) dermoscopy pictures have neglected to make a model reasonable for use in clinical applications. Among a wide range of skin malignancy, a melanoma is the deadliest. Melanoma is commonly a little, typically dark or darker shaded mole, which can grow anyplace on the skin. Distinguishing melanoma in its soonest organizes is among the most critical factors in enhancing the result of a melanoma analysis. Despite the fact that Melanoma covers just 1% of all skin malignant growths, it is in charge of the most elevated rate of death among all skin disease related setbacks. A measurement from The Skin Cancer Foundation demonstrates that melanoma represents roughly 10130 passings in the US every year. For 2017, the American Cancer Society gauges 87,110 new rates of melanoma bringing about 9,730 passings. To keep these passings, a practical answer for fast and early location of melanoma has turned into an exploration theme. Like different malignant cells, melanoma can quickly spread to different parts of the body causing serious harm. Nonetheless, the likelihood of melanoma cells getting to be metastasized stays low in beginning times. Doctors can find a way to monitor this malignancy, whenever analyzed and treated in these early periods. Lung malignant growth is the real reason for disease passing on the planet. The manifestations of lung malignant growth come into light at the Final stage. So it is exceptionally difficult to recognize in its starting stage. Hence, the demise rate is high for Lung malignant growth in examination with every other sort of disease. The two sort of lung sickness which create and spread in a startling way, are little cell lung malignancies (SCLC) and non-little cell lung tumors (NSCLC). The period of lung infection insinuates how much the development has spread in the lung. As indicated by a measurements led by world wellbeing association that consistently more than 7.6 million individuals kicked the bucket of lung

disease. In addition, the demise rates of lung malignancy are relied upon to continue ascending, to end up around 17 million worldwide in 2030. We discovered that lung malignant growths passings in Bangladesh achieved 9,660 or 1.33% of aggregate passings, as indicated by the most recent WHO information distributed. In year of 2005, around 1,362,825 new malignant growth cases are normal and around 571,590 passings are relied upon to occur because of disease in the United States. It was assessed that there will be 162,921 passings from lung disease, which happens 30% of all malignant growth passings. The leukemia is the most risky malignant growth through blood. The versatility of leukemia dieses changes according to age gathering. Thought about the a wide range of malignancy, the leukemia is the normal sort. Leukemia is most as often as possible analyzed among individuals matured 65-74. The graphical portrayal of level of new instances of leukemia and level of death of leukemia is depicted in the accompanying. The demise and year insights of leukemia understanding. By and large, leukemia is separated into four kinds, for example, Acute Lymphocytic Leukemia (ALL), Acute Myeloid Leukemia (AML), Chronic Lymphocytic Leukemia (CLL) and Chronic Myeloid Leukemia (CML). The kinds of Leukemia and events. The most perilous illness on the planet is malignancy in which bosom disease is the risky for ladies. Numerous ladies kick the bucket each year due to bosom malignant growth. Recognizing the bosom disease physically takes a considerable measure of time and it is troublesome for the doctor to grouping. So the distinguishing the malignant growth through different programmed symptomatic methods is extremely important. According to clinical master distinguishing this malignancy in its first stage helps in sparing lives. According to cancer.net offers individualized aides for in excess of 120 sorts of disease and related innate disorders.

In this paper, I proposed to utilize a Random Forest, Naïve Bayes calculation and KNN calculation to anticipate the malignant growth event and Logistic Regression calculations for approving the model outcomes. Distinctive Variable Prioritization Techniques is additionally used to play out a primer screening of factors and to get imperative positions. At that point, the new informational collection is removed from beginning dataset as indicated by best k imperative indicators and is contribution to the gathering machine learning calculations. The model will be approved against the distinctive datasets utilizing k overlap approval strategies .

II. LITERATURE SURVEY

Kittler, et al., has appeared in [1], dermoscopy has enhanced the indicative precision of melanoma by 10-27% over basic exposed eye examinations. In any case, the precision of examining dermoscopy pictures still relies upon the experience of a doctor. A dermatologist not prepared in perusing dermoscopy pictures can be less exact than bare eye examination [2]. When the time required to acquire dermoscopy pictures is thought of it as, is certain that mechanized acknowledgment frameworks are more productive. Straightforward neural systems that accomplish an unassuming dimension of precision are not hard to make; in any case, a framework that can accomplish an exactness suitable for clinical utilize is very testing and relies upon a few highlights, for example, varieties among size and shape, differentiate between where sores start and end, and curios, for example, hair shading and veins. Moreover, BCC itself is characterized into 2 noteworthy subtypes: nodular and shallow, each with their own characterizing qualities [3]. There is little research on BCC and neural systems, and the works gave have been inadequate up to this point. In 2011, Cheng, et al., distributed two papers identified with BCC. The main paper analyze BCC by means of telangiectasia examination. 96.7% exactness was guaranteed, be that as it may, in genuine clinical practice, genuine precision will be altogether less because of absence of telangiectasia's in early injuries, and telangiectasia's with comparative highlights being in kindhearted skin sores [4]. The second distributed paper utilized a novel method of joining highlight extraction from sores with highlights of the patient's close to home profile and physical exam attributes of injuries [5]. There was, in any case, no precision, specificity, or affectability esteems gave. novel strategy for dissecting BCC dermoscopy pictures. By utilizing profound ResNets our trial results demonstrate that, on paper, the model could be utilized by and by as a screening instrument. To the best of our insight this is the first run through such a technique has been connected to the recognizable proof of BCC. We incorporated our model with two phases: division and order. Our division display utilizes a FCRN equipped for distinguishing a sore in a picture and disposing of superfluous data. Our characterization display at that point takes this section and dissects it utilizing a profound lingering system 152 layers profound. This model works consistently from a solitary information picture to a last yield with no necessity of manual work. To enhance this model later on, specialists can utilize more information to change from paired characterization to a downright model, retrain the model on optical pictures, and potentially utilize the utilization of exchange picking up, drawing upon models pre-prepared on Distinctive surfaces [6].Though, dermoscopy is the best strategy for identifying melanoma, the unwavering quality of the discovery additionally relies upon the working expertise of the dermatologists. As the identification relies upon human vision and past experience, making it programmed is a urged research point to seek after. One work demonstrates that, a sack of-highlights grouping technique can be utilized on dermoscopic pictures for naturally identify melanoma [7]. Two strategies were displayed as

worldwide and neighborhood for the characterization. The worldwide technique was performed via programmed division, trailed by an extraction of shading and surface highlights for preparing the classifier. The nearby strategy was slanted towards picture investigation and acknowledgment. As of late, to make this discovery procedure more congenial for the all-inclusive community, computerized pictures are being utilized. One work with computerized picture utilized a blend of Otsu and k-implies grouping division strategies for recognizing the influenced region, and removing a few straight and non-direct highlights from the sore part [8]. we propose a strategy which can consequently distinguish the nearness of melanoma qualities in a mole from dermoscopic picture and gives an ongoing forecast rate about conceivable state of that mole, utilizing a constant protest identification procedure (YOLOv2) [9]. For encouraging the PC supported conclusion related work on melanoma numerous datasets can be discovered on the web. In this work, PH2 dataset is utilized, which is a dermoscopic picture database [10]. Each picture from that dataset was explained by the clinical order of the mole. For making the framework powerful, the information was expanded by working turn and obscuring techniques. Widening and disintegration were additionally connected for growth yet dismissed in light of the poor execution of the framework. The framework was prepared utilizing a best in class question recognition framework YOLOv2. The prepared framework was tried utilizing a five-overlay cross approval method and the outcomes gave an exactness of 86% while indicating invariance to location of any real hair. The proposed framework was kept running on a PC using a TITAN-X GPU. Future work could incorporate the advancement of a portable application where the client can take pictures of a suspicious mole, and the versatile application may remotely run on the web and present the anticipated outcome progressively [11]. Shown a CAD calculation using thresholding, morphological taking care of and Fisher Linear Discriminant to section, perceive patients knobs and remove from false positives. The structure got an exactness of 82.66% with 3 FP per case being approved with 143 handles. Gomathi and Thangaraj [12] used picture handling calculation, Fuzzy CMean figuring and neural classifier in the periods of preprocessing, discontinuity recognize patients knobs and separately. This calculation had a precision of 76.9%. Kumar et al. [13] proposed a CAD calculation that used Biorthogonal Wavelet Transform, locale becoming and fluffy based system in preprocessing, discontinuity and recognizable proof of knobs. The calculation can distinguish whether the info picture contains tumor cell or not and prepared to foresee if there is any probability of being development. The proposed calculation gives an exactness of 97% for malignancy distinguishing proof and 87% for disease forecast. The proposed framework would be suitable in helping the specialist in perceiving the lung as destructive or non-cancer-causing. The exactness of the framework can be improvised by means of setting it up on an enormous picture set and course of action in light of inherited figuring of hereditary calculation and profound neural network[14]. Built up on the result, the incomplete complexity extending strategy is viewed as the best procedure among all differentiation extending strategies that enhances nature of picture [15]. By and large, leukemia is isolated into four sorts, for example, Acute Lymphocytic Leukemia (ALL), Acute Myeloid Leukemia (AML), Chronic Lymphocytic Leukemia (CLL) and Chronic Myeloid Leukemia (CML) [16]. tried trial on 200 examples of Leukemia minuscule smear pictures. In this exploration paper we are featured the consequences of 10 test pictures. For the pre-handling of the database binarization, histogram and edge location procedures are utilized. The pre-handled pictures are portioned utilizing the Otsu picture division. The picture coordinating is finished utilizing the Maximally Stable Extremely Regions (MSER). Utilizing MSER we are separated the matrix and worldwide highlights of each picture. A similar arrangement of typical and distinctive set, for example, Leukemia with ordinary are coordinated with the match highlights focuses. The execution of the framework is computed by utilizing the FAR and FRR. The proposed framework announced exactness of 95.12% where FAR is 5.0% and FRR is 4.75% [17]. the intermingling time altogether increments and it gets harder to upgrade the system. In the event of convolutional neural system, we have discovered the best outcome with three hundred component maps. A similar outcome may be acquired with various design of the system. Our consequences of CNN classifier (98.06% precision) demonstrate nearly better execution in examination crafted by Karabatak and Cevdet-Ince where the exactness was 97.4% utilizing Association Rules(AR) and Neural Network(NN). Such near examination on bosom disease characterization would give further consolation and bits of knowledge on the productive methodologies for location of malignant growth issues [18]. Joining pre-prepared CNN (VGG-16) with a one-FC NN-classifier can accomplish normal precision about 0.905 for arranging strange versus ordinary cases in the DDSM database. In this investigation, the grouping exactness of the adjusting model is simply 0.008 higher than that of the component extraction demonstrate however the time cost of the element extraction show is just about 5% of that of calibrating model. Along these lines, this investigation demonstrates that applying move learning in CNN can identify bosom malignant growth from mammogram, and preparing a NN-classifier by highlight extraction is a quicker technique in exchange learning [19]. Univariate Feature Selection calculation utilized chi2 technique for choice Best 16 Features from UCI dataset. After gather last 16 highlights from univariate Feature Selection calculation we execute strategic and neural system calculation on these 16 highlights and last connected casting a ballot calculation on result and accomplished 98.50% exactness. Wisconsin Breast Cancer Dataset have contain 699 columns with highlights classes 30

highlights. After connected Univariate Feature Selection technique top 16 highlights are chosen from conclusive model execution. Since extensive highlights are impact on expense of model usage. Accomplished exactness is great from individual accomplished precision from both machine learning calculation [20]. Wisconsin Breast Cancer datasets, we utilized our two primary calculations, which are: NB and KNN, since our objective and test from bosom malignancy arrangement is to fabricate classifiers that are exact and solid. After a precise correlation between our calculations, we seen that KNN accomplished a higher productivity of 97.51%, be that as it may, even NB has a decent exactness at 96.19 %, if the dataset is bigger, the KNN's the ideal opportunity for running will increment [21]. the bosom malignant growth dependent on the way that whether their disease will be backslid or not. To do this, various types of classifiers are connected to various structures. The reproduction results communicated unpleasant neural system with two yields in the concealed layer prompted the most astounding exactness. Likewise, by and large, harsh neural system brought about the least difference contrasted with different techniques. This profits to the structure of unpleasant neural system which considers interim weight matrixes[22].

III.EXISTING SYSTEM

Strategic Regression regulated ML calculations have been utilized to order the Cancer dependent on chronicled information. The machine learning calculation process demonstrated that Machine learning calculations can be utilized viably with high exactness rate. Display execution measurements was estimated as far as TPR , FPR , Precision and Recall . Models developed utilizing singular classifiers are will in general have less exactness. Models are not approved crosswise over various examples. Thus, If irregular examples are taken, there is high possibility of getting less precision Enough needs was not given for underneath methods which can enhance the exactness of the models to the incredible degree to Variable choice (Prioritization of Factors), Feature specialist and Variable changes, Variable decreases. Strategic relapse needs to indicate a regularization coefficient each time the more noteworthy the esteem. the more grounded the capacity to fit the nonlinearity. there is no great basis for the underlying estimation of the customary term coefficient is set to 0 to decide a superior learning rate. At that point learning rate is settled. Furthermore, the esteem, (for example, 1) is given to the lambda is expanded or lessened by multiple times . two sets are intended for preparing set with 75% examples and test set containing 25% examples. its gives less precision. Calculated Regression is additionally not a standout amongst the most amazing calculations out there and can be effectively beated by more perplexing ones. Another burden is its high dependence on an appropriate introduction of your information. This implies calculated relapse is anything but a valuable device except if you have effectively distinguished all the critical autonomous factors. Since its result is discrete, Logistic Regression can just anticipate a clear cut result. It is likewise an Algorithm that is known for its weakness to over fitting. A burden of it is that we can't take care of non-direct issues with strategic relapse since its choice surface is straight. All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

IV.PROPOSED SYSTEM

A group is itself an administered learning calculation, since it very well may be prepared and afterward used to make forecasts. The prepared gathering, in this way, speaks to a solitary speculation. This speculation, in any case, isn't really contained inside the theory space of the models from which it is assembled. Subsequently, groups can be appeared to have greater adaptability in the capacities they can speak to. This adaptability can, in principle, empower them to over-fit the preparation information in excess of a solitary model would, yet practically speaking, some troupe strategies will in general decrease issues identified with over-fitting of the preparation information. Experimentally, troupes will in general yield better outcomes when there is a huge decent variety among the models. Numerous troupe techniques, subsequently, look to advance assorted variety among the models they consolidate. Albeit maybe non-instinctive, more arbitrary calculations (like irregular choice trees) can be utilized to deliver a more grounded outfit than exceptionally conscious calculations (like entropy-decreasing choice trees). Utilizing an assortment of solid learning calculations, in any case, has been appeared to be more viable than utilizing systems that endeavor to impair the models with the end goal to advance decent variety. As the troupe classifiers are utilized, the model improves as a decision for forecasts. Approving the models with K-Fold cross approval procedure results in the model adequately working for any sort of information.

K-Fold cross approval method: K-Fold cross approval system in some cases called turn estimation, or out-of-test testing is any of different comparative model approval strategies for evaluating how the aftereffects of a factual examination will sum up to a free informational collection. It is predominantly utilized in settings where the objective is forecast, and one needs to gauge how precisely a prescient model will perform practically speaking. In an expectation issue, a model is generally given a dataset of known information on which preparing is run (preparing dataset), and a dataset of obscure information (or first observed information) against which the model is tried (called the approval dataset or testing set). The objective of cross-approval is to test the model's

capacity to foresee new information that was not utilized in evaluating it, with the end goal to hail issues like over fitting or choice predisposition and to give an understanding on how the model will sum up to a free dataset (i.e., an obscure dataset, for example from a genuine issue). Cross Validationinvolves apportioning an example of information into correlative subsets, playing out the investigation on one subset (called the preparation set), and approving the examination on the other subset (called the approval set or testing set). To lessen inconstancy, in many strategies numerous rounds of cross-approval are performed utilizing distinctive parcels, and the approval results are joined (e.g. found the middle value of) over the rounds to give a gauge of the model's prescient execution. In outline, cross-approval consolidates (midpoints) proportions of wellness in expectation to determine a more precise gauge of model forecast execution. The proposed framework will help the accompanying advantages contrasted with the current models, for example, Less Variance and Biasness, Better selection of factors through element designing, prioritization lattice, High Accuracy.

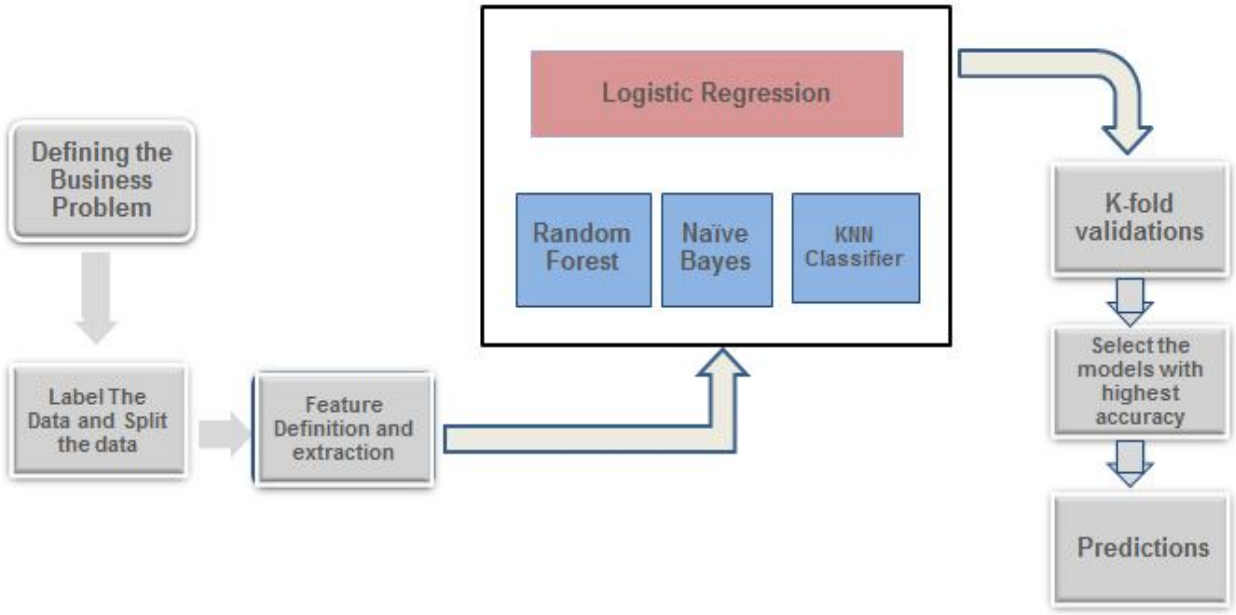

Figure: Predicting types of cancer model

Foreseeing disease utilizing models Naive Bayes classifier, k-closest neighbor, Random woods appeared in figure 1.Pre-handling method is gathering the informational collections by characterizing the business issue. Clean the informational collections and union the information. At that point name the information and split the information .Once part is finished. Extraction will process informational indexes utilizing the models Naive Bayes classifier, k-closest neighbor, Random timberland. Strategic relapse calculation utilizing in this model to demonstrate it will give less precision as I talked about in existing framework. When extraction is finished. Models will run the informational indexes utilizing K-overlay approvals strategy informational indexes will process via train and test method. Model will give the most elevated exactness as results. At that point we can anticipate the whether the patient as malignant growth or not. This model will be useful for healing centers and it's likewise useful for specialists effortlessly they can comprehend whether the patient has malignant growth or not utilizing the front end device. Front end device will actualize for specialists to utilize.

V. CONCLUSIONS

This paper centers the group classifiers are utilized, the models improves as a decision for expectations. Which it will gauge the patient as malignant growth or not. An investigation of machine learning will be connected with the end goal that an expectation of sorts of malignant growth can be made dependent on the information gathered from individuals. An examination can be produced using the past technique to guarantee that The proposed framework will help the accompanying advantages contrasted with the current models give Less Variance and Biasness, Better selection of factors through element designing, prioritization grid , High Accuracy.

VI. ACKNOWLEDGMENT

I am very much thankful to Dr. B. Rajalakshmi, professor and Head, Dept of computer science & Engineering, NHCE, Bangalore for her constant support extended towards me during the course of this project. Her help and advice instilled me to complete the project on time.

REFERENCES

[1] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of dermoscopy," The lancet oncology, vol. 3, no. 3, pp. 159– 165, 2002.
[2] M. Binder, M. Puespoeck-Schwarz, A. Steiner, H. Kittler, M. Muellner, K. Wolff, H. Pehamberger, et al., "Epiluminescence microscopy of small pigmented skin lesions: short-term formal training improves the diagnostic performance of dermatologists," Journal of the American Academy of Dermatology, vol. 36, no. 2, pp. 197–202, 1997.
[3] A. C. Society, "Key statistics for basal and squamous cell skincancers."
[4] B. Cheng, "Automatic vessel and telangiectases analysis in dermoscopy skin lesion images," 2009.
[5] B. Cheng, R. J. Stanley, W. V. Stoecker, S. M. Stricklin, K. A. Hinton, T. K. Nguyen, R. K. Rader, H. S. Rabinovitz, M. Oliviero, R. H. Moss, and et al., "Analysis of clinical and dermoscopic features for basal cell carcinoma neural network classification," Skin Research and Technology, vol. 19, no. 1, 2012.
[6] Eric Vander Putten, Ameer Kambod1, and Mobeen Kambod2"Deep Residual Neural Networks for Automated Basal Cell Carcinoma Detection" March 2018.
[7] C. Barata, M. Ruela, M. Francisco, T. Mendonca, and J. S. Marques, "Two Systems for the Detection of Melanomas in Dermoscopy Images Using Texture and Color Features," IEEE Syst. J., vol. 8, no. 3, pp. 965–979, Sep. 2014.
[8] T. T. K. Munia, M. N. Alam, J. Neubert, and R. Fazel-rezai, "Automatic Diagnosis of Melanoma Using Linear and Nonlinear Features from Digital Image," In
[9] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," Dec. 2016.
[10] "ADDI - Automatic computer-based Diagnosis system for Dermoscopy Images." [Online]. Available: https://www.fc.up.pt/addi/ph2 database.html. [Accessed: 01-Sep-2017].
[11] Shudipto Sekhar Roy, Akkas Uddin Haque, Jeremiah Neubert "Automatic Diagnosis of Melanoma from Dermoscopic Image Using Real-Time Object Detection" 2018.
[12] S. K. Kumar, J. Ramesh, P. T. Vanathi, K. Guunavathi, "Robust and automated lung nodule diagnosis from CT images based on fuzzy systems." in International Conference On Process Automation Control and Computing, Coimbatore, India, IEEE, 2011.
[13] M. Subrajeet, S. S. Sushanta, P. Dipti, S. Sanghamitra "Fuzzy based blood image segmentation for automated leukemia detection."in International Conference on Devices and Communications, Mesra, India, IEEE, 2011.
[14] Janee Alam1, Sabrina Alam2, Alamgir Hossan3 "Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classifier" 2018
[15] Raja Rajeshwari V, N.Ramesh, "Contrast Stretching Enhancement Techniques for Acute Leukemia Images", Publication of Problems and Application in Engineering Research Papers, vol.4, special Issue 01, pp 190-194, 2013
[16] Virmani J., Kumar V., Kalra N., Khandelwal N., "Characterization of primary and secondary malignant liver lesions from B-mode ultrasound", Journal of Digital Imaging, Vol. 26, No. 6, pp. 1058- 1070, February 2013.
[17] Rege M.V. Mohammed Basil Abdulkareem Santosh Gaikwad B.W. Gawli "Automatic Leukemia Identification System Using Otsu Image segmentation and MSER Approach for Microscopic Smear Image Database" 2018
[18] Shajib Ghosh , Jubaer Hossain , Dr. Shaikh Anowarul Fattah , Dr. Celia Shahnaz , Asir Intisar Khan ": Efficient Approaches for Accuracy Improvement of Breast Cancer Classification Using Wiscons in Database" (2017).
[19] Shuyue Guan , Murray Loew "Breast Cancer Detection Using Transfer Learning in Convolutional Neural Networks" (2017).
[20] Naresh Khuriwal , Nidhi Mishra "Breast Cancer Diagnosis Using Adaptive Voting Ensemble Machine Learning Algorithm Authors : Shuyue Guan , Murray Loewo (2018)".
[21] Meriem AMRANE , Saliha OUKID , Ikram GAGAOUA, Tolga ENSAR "Breast Cancer Classification Using Machine Learning " (2018).
[22] Noushin Jafarpisheh , Nahid Nafisi , Mohammad Teshnehlab "Breast Cancer Relapse Prognosis by Classic and Modern Structures of Machine Learning Algorithms "(2018).