Journal of Theoretical and Applied Information Technology
30th November 2023. Vol.101. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A DOCTOR RECOMMENDATION AND BREAST CANCER PREDICTION USING MODIFIED K-MEANS AND SVM

[1]RETHINAKUMAR, [2]GOPINATH GANAPATHY, [3]JEONG-JIN KANG
[1]Research Scholar, Bharathidasan University, Assistant Professor,
Saranathan College of Engineering,India.
[2]Professor, Dept of Computer Science, Bharathidasan University, India.
[3]Professor, Dept of Information and Communication, Dong Seoul University,South Korea.

ABSTRACT

Today, people are exposed to various diseases due to their living habits and environmental conditions. Early diagnosis is very important, but it can be hard to predict accurately due to the symptoms. The correct prediction is therefore the most challenging part of the job. Data mining is a process that can help predict a disease. Through the use of medical data, it can analyze the patterns of the disease. In this research the K-means algorithm is used in this study along with a support vector machine. In previous work, we utilized the CNN algorithm to identify breast cancer, but this was unsuitable for large-scale datasets. We have now modified this algorithm to handle training breast cancer information.The modified K-means algorithm can reduce the training number and produce a new dataset that is completely original. It also provides the necessary information about the predicted disease and its recommended doctors. The Proposed algorithm takes into account various factors such as the distance from the predicted location, the experience of the doctors, and the feedback of the users to provide a personalized recommendation. Through the proposed algorithm, the user can get the most effective and personalized treatment possible. It also reviews the recommended doctors and provides its own recommendations to improve the accuracy of the diagnosis.

Keywords: *K-means, Support Vector Machine, Data mining and Machine Learning*

1.INTRODUCTION

One of the most challenging issues researchers face nowadays is the diagnosis of cancer. Around 10% of women worldwide will be diagnosed with breast cancer in their lifetime.[1].By utilizing computer-aided diagnosis methods, researchers can now spot breast tumors at an early stage. This method is commonly used in combination with other biomedical technologies such as X-ray radiography [2]. Due to the varying dimensions of breast tumors, classification in machine learning is regarded as a challenging task.Unlike lab tests that only consider the imaging of a magnetic resonance image or a PET scan [3].

Machine learning can be used to detect breast cancer. Different methods have been utilized. These include multi-level perceptron's, logistic regression, and support vector machines [4]. The performance and stability of these techniques are dependent on various factors, such as the model's structure, algorithms' parameters, and the features of the model [5]. The development and implementation of effective models for predicting breast cancer directly affects the safety and treatment of patients. This is why it is important for those working in the data mining industry. The performance of statistical models in predicting breast cancer was good. They were able to distinguish between benign and malignant tumors [6]. K-means is a widely used algorithm for determining and treating breast cancer. It clusters the collected data into groups to find patterns related to unlabeled information [7]. Different clusters of breast cancer data have varying characteristics. Through K-means, researchers can easily group the collected information into clusters that are composed of cases of malignant and benign breast cancer [8]. SVMs have been widely used in the past few years due to their impressive achievements. This method splits the data into multiple classes and recognizes the highest margins in each class to ensure that the results are accurate [9].

When choosing a doctor, patients often have a hard time choosing the right one based on the criteria that they used to make their decision. Although there are various online health portals and recommender systems that can help solve this issue, they are not able to provide a comprehensive solution. Although there are numerous health portals that contain extensive information about healthcare, it can be a daunting task to find the right information quickly [10]. The recommendations made by a physician must capture the patient's desires in order to be effective.This paper aims to identify the various features of a doctor and their importance in making informed decisions.

2. RELATED WORKS

Kumar et al (2021) [11] -This project used a system that can predict a patient's disease based on its symptoms. The patient can then input their symptoms. Various classification algorithms, such as the KNN, Random Forest, and Navie Bayes, had been utilized to forecast the disease. For instance, if the system detects a disease such as diabetes and heart disease, it can predicted its symptoms based on its True or False classification. Such systems may be useful in predicted the future medical treatment of diseases. They can inform patients about their condition based on their symptoms. After the system had predicted the disease, it can then recommend the appropriate doctor for the patient. The paper discussed the various applications of Machine Learning in predicting diseases. It showed how to improve the model by taking advantage of massive datasets collected by hospitals.

Ragab et al (2019) [12] -The objective of this project was to developed a new CAD system that would allow the classification and detection of various types of tumors in mammograms. There were two segmentation techniques that were suggested. The first method used circular contours to reduce the ROI. The red contour in the DDSM dataset was used to label the tumors.The second technique utilized a region-based approach to determine the largest area within the range. The threshold was then set at 76.During the extraction phase, DCNN was utilized. To differentiate between surgical and medical images, the AlexNet underwent a re-calibration.The first segmentation technique performed better than the second one by 1.8% in the dataset with the help of the DCNN. To improve its accuracy, the last layer was substituted with a SVM. When comparing the two methods, it was revealed that the second technique, which utilizes a linear kernel function, performed well. The suggested CAD system could be utilized to detect breast abnormalities that are not related to the tumor itself., it could be further developed with the deep convolutional network and the residual framework for improve the accuracy.

Botlagunta et al (2023) [13] -The development of various programmed modules for the use of data processing and text mining techniques was carried out. The classification framework performance was evaluated using a validated criterion. The ROC, accuracy, and AUC of the predictions were then compared. The data's statistical significance was then determined by performing the Welch Unpaired test. The use of text mining techniques in EHR helped identify patients with MBC. The mean difference between the healthy and MBC individuals was revealed by monocytes. The accuracy of ML models improved significantly with the removal of outliers. A Decision Tree (DT) classifier was able to achieve an accuracy of 83% and an AUC of 0.87, while a Flask-based web application was created to help physicians identify MBC patients. We concluded that these models could help improve the survival outcome of those who are in intensive care but it lagging to identify disease category

Shafique et al (2023)[14] -The objective of this study was to find a way to detect breast cancer using a fine-needle aspiration technique. The suggested KNN method was able to provide a 100% accurate prediction of breast cancer. It performed better than current techniques. The KNN method was also able to improve its accuracy by using the 15 essential features of the PCA algorithm. It was also able to perform better than other models when it used the 20 most crucial parameters. The suggested method was also 100% accurate when used with an SVD. Unfortunately, it was limited by the experiments that were being conducted with the WBCD, which is a standard dataset used in detecting breast cancer.

Yan et al (2020) [15] -This paper discussed a hybrid algorithm that used deep learning to recommend doctors. It learns the initial value of patient reviews and doctors' professional knowledge using an automatic depth algorithm. The PMF-CNN model was able to learn the various context features of reviews by using a convolutional neural network. This allows it to perform more accurate modelling of the data. The paper suggests a method that learns the initial value of the interactions between doctors' professional knowledge and patient reviews in the matrix decomposition framework. It utilized a noise reduction technique to avoid getting stuck in the algorithm's recommended solution. However, the main issue with the PMF-CNN model was that it can only recommend certain historical doctors. The

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

paper planned a method that learns the initial value of the interactions between doctors' professional knowledge and patient reviews in the matrix decomposition framework. It then used this to provide a personalized recommendation based on the collected data. In the future, it would also consider incorporating multiple context features to improve the accuracy of its recommendations.

3 CANCER PREDICTION AND DOCTOR RECOMMENDATION SYSTEM USING P-Model

We presented a K-means algorithm-based model with a modified SVM technique in this study. In previous work, we used CNN for the identification of breast cancer. This method is only suitable for high-dimensional data. To address this issue, we extracted various texture features. The K-means training algorithm is utilized to create a breast cancer analysis training dataset. By modifying its algorithm, the sample size can be reduced, and the dataset can be created entirely new.During the pre-processing phase, the training data is broken down into two categories: malignant and benign. K-means is utilized to reduce the number of entries in each classification category. The resulting set is then trained to improve the classifier's performance.Compared to the training data with full data, the high-quality training dataset provides the shortest training time.

Since finding the right doctor can be very challenging, recommender systems are being used by many people after they have been diagnosed with breast cancer. These systems use a combination of machine learning and statistical techniques to analyze and forecast the available resources. Recently, the utility of recommender systems has been acknowledged in the medical field. More research is being conducted to improve their capabilities in healthcare. Prior to doctor recommendation, patients had two options when choosing a healthcare provider. The scope and application of these options are limited. One of the options is to ask for recommendations from friends or relatives. Although these sources are usually reliable, the chances of having a similar medical history are low. Also, it's possible that the recommendations from a patient's social circle might not cover all of their healthcare options.

The P-Model proposes a hybrid algorithm that combines the SVM and mod K-means clustering to improve the prediction of cancer. Frame Work Figure 1 shows the model's methodology.

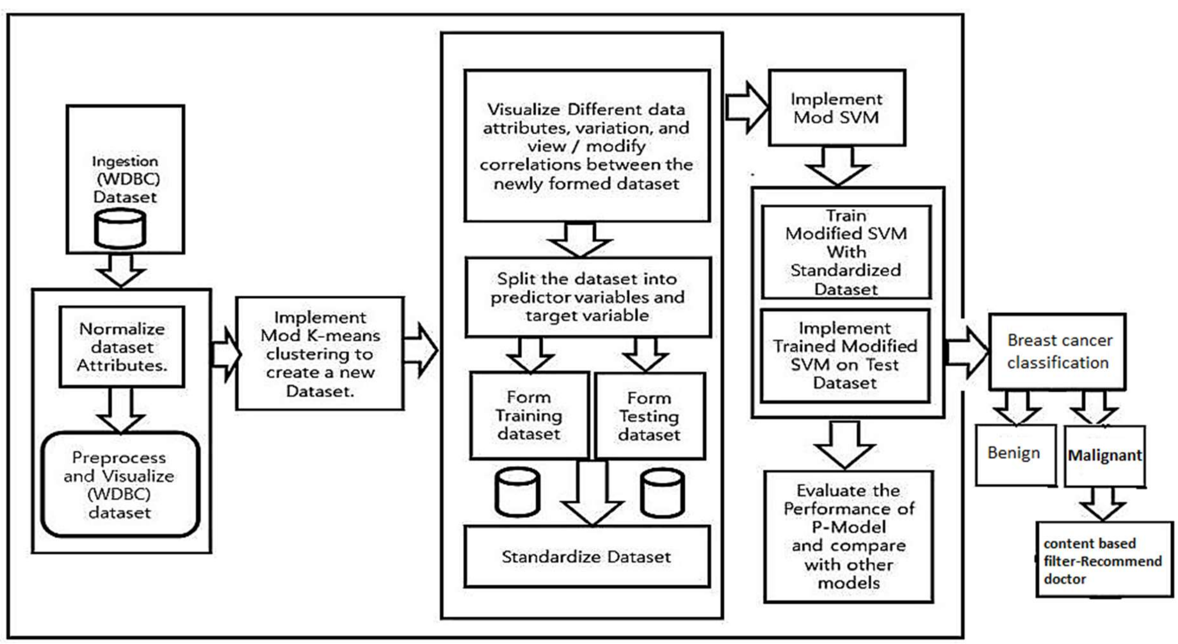


Figure 1. Proposed Architecture

The P-Model takes the WDBC datasets and visualizes them. After normalizing the data, it then pre-processes it. In order to create new clusters, we use a modified clustering algorithm that's designed to create groups with the shortest distances between them. The clustering process will automatically create new clusters if the distance between the centroids of the clusters exceeds the threshold value. All WDB samples are then tested for this procedure. The initial centroid of the data is chosen as the starting point of the clustering process. The K-means clustering algorithm then identifies the clusters that can be utilized in the training data. The K-means method is more flexible compared to other clustering techniques. It doesn't require the computation of a cluster number. The goal of this method is to minimize the number of entries in the training dataset through clustering. In the case of the Benign category, the clustering algorithm will automatically create clusters with similar characteristics. The clustering mean for the new samples will determine their classification.

3.1 Dataset description

The quality of the samples that were produced should match the original training data. After pre-processing and visualizing the data, we then check the correlations and attribute relationships between the various elements. To create a frame for the analysis, we'll divide the data into target and predictor variables. For the training set, 80% of the dataset was used, while the remaining 20% was used for testing. After identifying breast cancer, the doctor recommendation was performed using a content-based collaborative filtering method.

3.2 Feature Extraction And Visualize The (Wdbc) Dataset

Breast Cancer Wisconsin's data set contains various features that can be computed by using a digitized image. These features are used to record the cancer's prognosis. Breast thermograms are categorized into normal and abnormal. The former generally have symmetric vascularization or temperature distribution across the right and left breasts. On the other hand, abnormal breasts have a large hyperthermic area that is associated with the tumor. The degree of abnormality detected during thermography can increase the likelihood of cancer and worsen the prognosis. The appearance of vascularities and hyperthermic areas can cause subtle changes in the texture of the breast, which can be imperceptible to the naked eye but can affect its asymmetry. A texture analysis technique can be utilized to detect breast abnormalities. It can be obtained by extracting features from the thermogram's spectral or spatial domain representations. A statistical method can then be utilized to analyze the texture levels in the breast using the spatial domain's statistical features. The method for obtaining the spatial distribution of the gray values in a thermogram is known as the feature extraction process. It involves computing the various features at each point of the image and then coming up with a set of statistics.

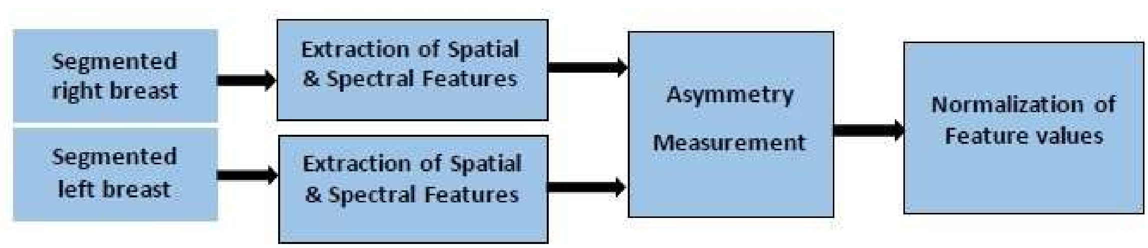


Figure 2: Feature Extraction Steps

3.2.1. Extraction of Texture Features from Spatial Domain

The spatial features of a given area are extracted using RLM, NGTDM, and the computing histogram, as well as the second and first order statistical features with the help of the ROI and the GLM. The computed mean, variance, skewness, entropy, and kurtosis are presented by the histogram of the image. The four-direction GLCM matrixes are also taken into account. The various features of the GLCMs are extracted from the given area. These include energy, variance, contrast, and entropy. The average features of the four GLCMs are then compared and compared with the chaotic breast cancer development.

The basic information about an image is presented in a histogram, which is composed of various statistical features. Some of these include the mean, variance, skewness, and entropy.

$$\mu = \sum_{i=1}^{G-1} iP(i) \quad (1)$$

The number of gray levels in an image is the probability of each level being represented by a different color is known as P(i). The mean is the average intensity of the image.

$$\sigma^2 = \sum_{i=1}^{G-1} (i-\mu)^2 P(i) \quad (2)$$

Variance indicates variability of intensity values present in the region.

$$Skewness = \frac{1}{\sigma^3} \sum_{i=1}^{G-1} (i-\mu)^3 P(i) \quad (3)$$

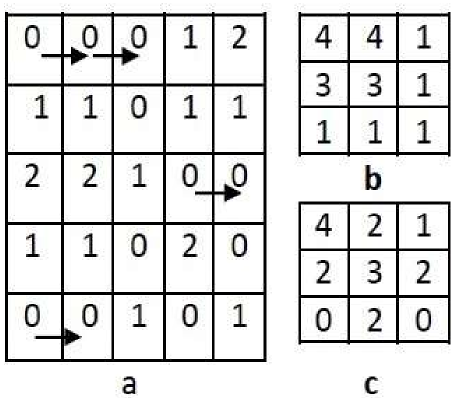


Figure 3: Instance for GLCM

Skewness known as third moment, is a measure of asymmetry. Its value remains nearly the same in case of normal breast.

$$Kurtosis = \frac{1}{\sigma^4} \sum_{i=1}^{G-1} (i-\mu)^4 P(i) \quad (4)$$

Kurtosis known as the fourth moment is a measure of the asymmetry between the breasts. It helps identify areas with a higher degree of symmetry deviation.

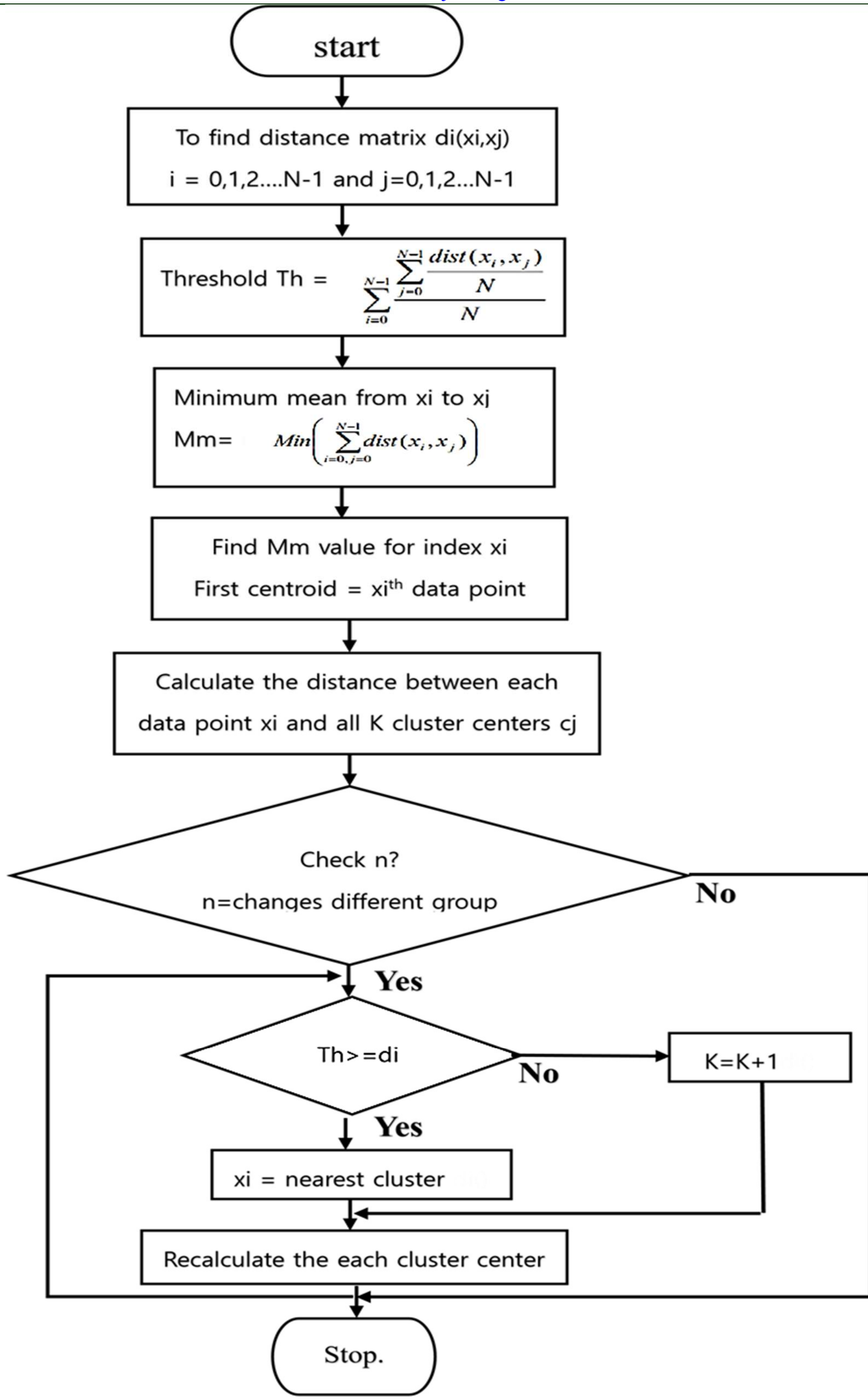
$$Entropy = \sum_{i=0}^{G-1} P(i) \log_2[P(i)] \quad (5)$$

The quantity of uncertainty that exists in the breast is measured by entropy. This is a useful tool for detecting asymmetry.

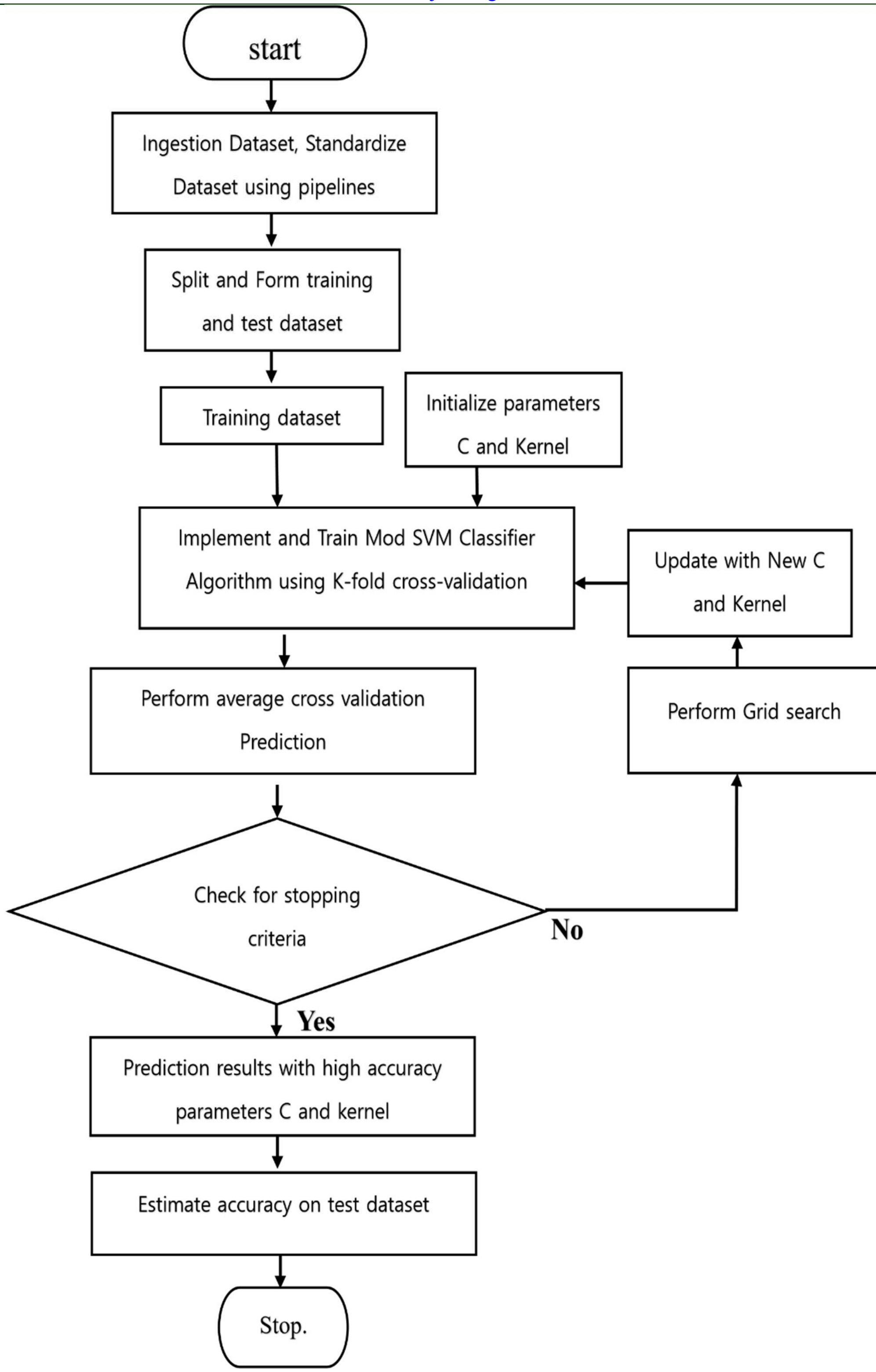
Histogram moments do not determine the position of the pixels. The neighborhood relations of the images are computed by taking into account the NGTDM and GLCM values. The number of gray levels in an image is computed by comparing the two pixels' distance and orientation.

3.3 Modified K Means With Svm For Breast Cancer Identification

The proposed model combines the SVM and K-means methods. The steps in the process can be found in Figure 4. The system's performance is then improved by using a new dataset. Classifiers for breast cancer are based on theories and statistical methods. The classification process can then split the collected data into groups according to the N-dimensional model's hyperplane. Numbers and training samples are the representation of the classes in the collection. Support vector machines can also be used to find the closest point in the model's space. The distance between the model's hyperplane and the closest point in the space is then calculated by the SVMs.


start
To find distance matrix di(xi,xj)
i = 0,1,2....N-1 and j=0,1,2...N-1
Threshold Th = $\sum_{i=0}^{N-1} \frac{\sum_{j=0}^{N-1} \frac{dist(x_i, x_j)}{N}}{N}$
Minimum mean from xi to xj
Mm= $Min\left(\sum_{i=0,j=0}^{N-1} dist(x_i, x_j)\right)$
Find Mm value for index xi
First centroid = xith data point
Calculate the distance between each
data point xi and all K cluster centers cj
Check n?
n=changes different group
No
Yes
Th>=di
No
K=K+1
Yes
xi = nearest cluster
Recalculate the each cluster center
Stop.

(A)



(B)

Figure 4: Flow chart of (a) Modified K-Means Algorithm (b)Modified SVM Algorithm

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Minimise

$$W(\alpha)\frac{1}{2}\sum_{i=1}^{N}\sum_{N_{j=1}} y_i y_j \alpha_i(x_i, x_j) - \sum^{N}{}_{i=1}$$

Subject to:

$$\forall i: 0 \leq \alpha_i \leq C and \sum^{N}{}_{i=1} \alpha_i y_i = 0$$

A vector of N variables is defined as α, which indicates that the soft margin is identified by C. The kernel functions of special geometry representations (SVMs) are C>0, C=0,0, and K(xi, xj).

Classification models are designed to minimize the bias and variance issues that can occur in their classification. This is done through the use of learning methods that are designed to minimize these errors.It can be caused by the algorithm or method itself. Variances are random errors that can happen due to the uncertainty in the training data or the learning method procedure. In general, a prediction model with low bias can have overfitting problems, The accuracy of a model can be affected by various factors, such as under fitting. On the other hand, if a model has low variance, it can cause issues that can lead to inaccurate results.Breast cancer classification is carried out through the use of a single SVM, which has various settings that can impact its accuracy. Besides these, other factors such as the model's structure and kernel functions can also affect its accuracy. A suitable SVM is proposed using a radial basis function followed by a grid-based strategy. It is then trained to classify the training materials into two categories: benign and malignant. The classification model's performance is evaluated by means of the confusion matrix.

3.4 K-Means And Content Based Doctor Recommendation For Cancer Care

After you have been diagnosed with cancer, it's important that you choose the right doctor. Although you may continue seeing the same doctor for your treatment, you might also go to another specialist. An oncologist is a type of doctor who specializes in treating cancer. Although you may eventually need to see a different specialist, your oncologist and other healthcare professionals will continue treating you for a long time. Through the content-based recommendation system, which is powered by the Google algorithm, you can get a personalized list of doctors based on your needs and the information collected from various sources.

- Before you choose a doctor for cancer treatment, it's important that you thoroughly research their experience and their capabilities.
- Are you an oncology board certified individual? This means that you have successfully completed a rigorous examination that's focused on the treatment of cancer.
- The length of your medical career is a vital question.
- How many individuals with this type of cancer are you treating each year?
- How many individuals with this type of cancer do you typically treat annually?

Are your patients aware of the availability of clinical trials?

4. Experimental Evaluation

The suggested method for developing a more accurate cancer prediction is by combining the SVM and K-Mean clustering algorithms. It was tested using a Core i7 and a 3.8.2 shell.

4.2 Dataset

The Wisconsin dataset contains 569 cases of diagnostic breast cancer. The proposed model takes into account the data set from UCI's repository, Which 357 are classified as malignant tumors and 212 are benign cases. It also has 32 key characteristics. The 32 key characteristics of patients are grouped together into three categories: The distinguishing characteristics of tumors are defined by their class indicator, patients ID, and distinguishing attributesare based on various aspects, such as radius, texture, area, compactness, and concavity. They can be obtained using a digitized image of the breast mass.

4.3 Statistical Analysis of Spatial Features using t Test

An independent test is conducted to determine the statistical significance that various spatial features have in distinguishing normal and abnormal breast tissue. It can be used to determine if the difference between the two groups' averages really reflects the population's differences. A significance level of 5% is tested by taking into account the variance between the two groups' averages. The asymmetry in the thermograms of the right and left breast is more pronounced than that of the normal ones due to the presence of tumor and vascularization.These abnormalities can be differentiated in early breast cancer detection.

4.5 Evaluation of P-Model on Standardized Data

We compare the performance of various learning algorithms against each other in Figure 5 such as KNN, NB, KART, and SVM, in a 20% testing dataset without tuning. We found that modified SVM with C=1.3 and RBF kernel performed well with an accuracy of 0.96 percent. The graph shows the performance of different SVM variants against popular machine learning frameworks such as NB and KNN in a 50%WDBC testing dataset with no tuning. Compared to the unmodified version, the modified version with C=0.1 and RBF kernel was able to achieve an accuracy of 0.97%. During the testing phase, it was used to classify the data into two categories: Benign and Malignant.

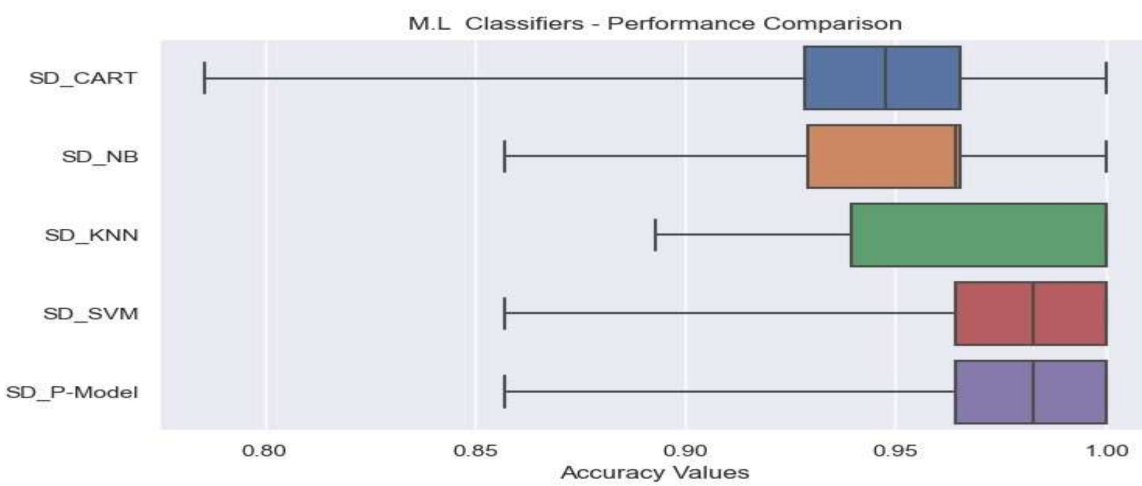


Figure 5: Comparison On Standardized Data (80%+20%)

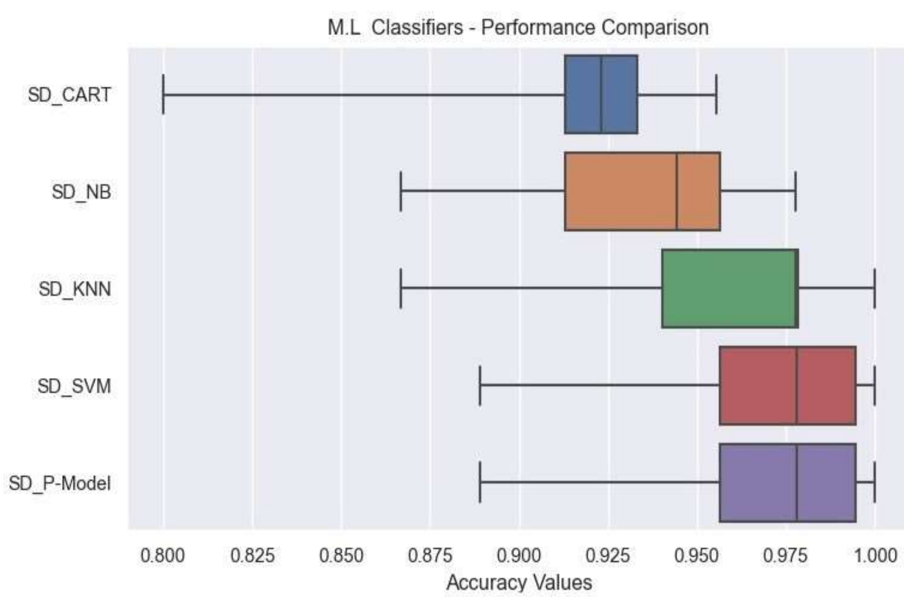


Figure 6: Comparison On Standardized Data (50%+50%)

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.6 Measure for Performance Evaluation / Experimental Results

A previous discussion noted that an algorithm for analyzing the data collected by the WDBC was evaluated using a ten-fold cross-validation approach. The diagnosis accuracy was maintained at 97.38%.The generated performance can be attributed to the confusion matrix. Figure 7 shows the accuracy of our proposed model and machine learning classifiers. In the 50 percent of testing dataset, our proposed model performed 96.49% accurate. In the 20% of testing data, our proposed model performed 99.12% accurate. The proposed model's accuracy is impressive.

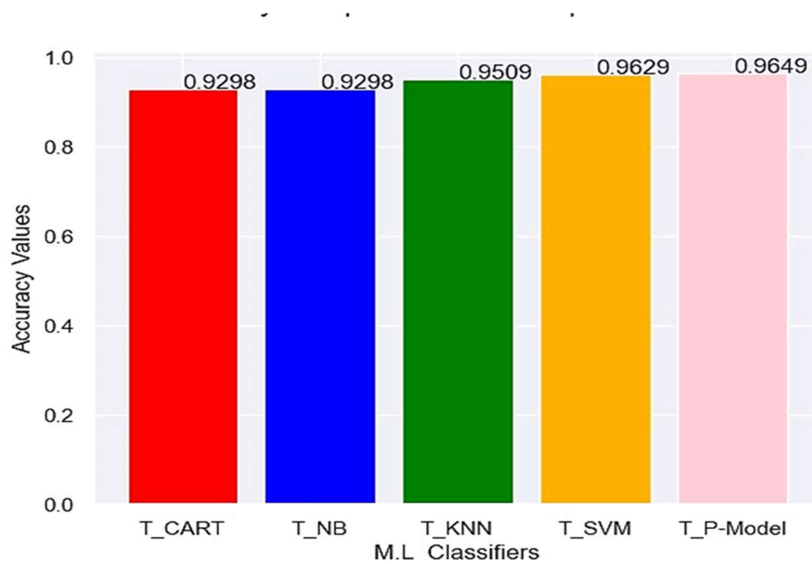


Figure 7: Accuracy (50%+50%)

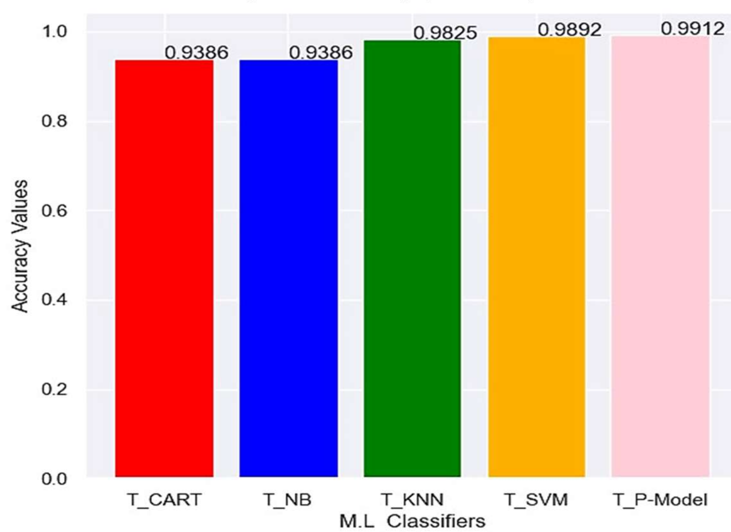


Figure 8: Accuracy (80%+20%)

The performance of classification models is affected by the matrix that shows the cases in two groups. one is positive, which is referred to as benign, and the other is negative, which is referred to as malignant. The proposed model performed better than the others in our tests after tuning. Table 2 summarizes the results of our analysis, while Table 6 summarizes the collected data from the Western Digital Business Center.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 5: Evaluation Comparison Based On 50% Of Testing WDBC Dataset

M.L Classifiers	Class	precision	recall	f1-score	support	Accuracy	Run Time
CART	Benign	0.94	0.94	0.94	181	0.929825	0.001999
	Malignant	0.90	0.90	0.90	104		
	Average	0.92	0.92	0.92	285		
NB	Benign	0.94	0.94	0.94	181	0.929825	0.000999
	Malignant	0.90	0.90	0.90	104		
	Average	0.92	0.92	0.92	285		
KNN	Benign	0.94	0.98	0.96	181	0.950877	0.001999
	Malignant	0.97	0.89	0.93	104		
	Average	0.96	0.94	0.95	285		
SVM	Benign	0.97	0.98	0.97	181	0.962912	0.001999
	Malignant	0.96	0.94	0.95	104		
	Average	0.96	0.96	0.96	285		
P-Model	Benign	0.97	0.98	0.97	181	0.964912	0.001999
	Malignant	0.96	0.94	0.95	104		
	Average	0.96	0.96	0.96	285		

Table 3: Evaluation Comparison Based On 20% Of Testing WDBC Dataset

M.L Classifiers	Class	precision	recall	f1-score	support	Accuracy	Run Time
CART	Benign	0.95	0.96	0.95	75	0.938596	0.000999
	Malignant	0.92	0.90	0.91	39		
	Average	0.93	0.93	0.93	114		
NB	Benign	0.95	0.96	0.95	75	0.938596	0.000999
	Malignant	0.92	0.90	0.91	39		
	Average	0.93	0.93	0.93	114		
KNN	Benign	0.97	1.00	0.99	75	0.982456	0.002999
	Malignant	1.00	0.95	0.97	39		
	Average	0.99	0.97	0.98	114		
SVM	Benign	0.98	0.99	0.99	75	0.989228	0.003994
	Malignant	0.98	0.97	0.97	39		
	Average	0.98	0.98	0.98	114		
P-Model	Benign	1.00	0.99	0.99	75	0.991228	0.002998
	Malignant	0.97	1.00	0.99	39		
	Average	0.99	0.99	0.99	114		

4.7 Doctor Recommendation results

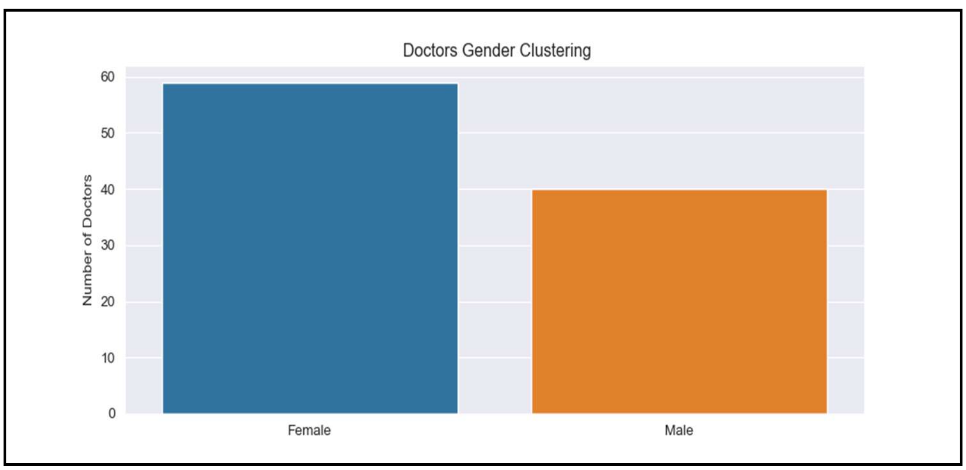
The suggested online doctor recommendation system would provide a more accurate and comprehensive analysis of patients' medical conditions. It takes into account their descriptions, medical images, and the expertise of their doctors. The suggested model takes into account the location of patients as well. This would enhance the convenience of online consultations and improve the value of prediagnosis data. The data presented in Figure 9 shows the various types of patient information collected by the suggested system. Figure 10 shows the plots of the doctor recommendations.

id	Name / Description	GENDER	AGE
Patient Id : 15 - Best Doctor for Cardiologist - Trichy.		MALE	AGE-51
Patient Id : 16- Best Doctor for Cardiologist - Bangalore.		FEMALE	AGE-52
Patient Id : 17 -Best Doctor for General Medicine, Cardiologist - Chennai		MALE	AGE-53
Patient Id :18 -Best Doctor for General Medicine - Bangalore		FEMALE	AGE-54
Patient Id :19 - Best Doctor for Radiation Oncologist-Chennai.		FEMALE	AGE-55
Patient Id :20 - Best Doctor for Radiation Oncologist-Trichy.		MALE	AGE-56
Patient Id :21 -Best Doctor for Breast cancer Oncology-Trichy.		FEMALE	AGE-57

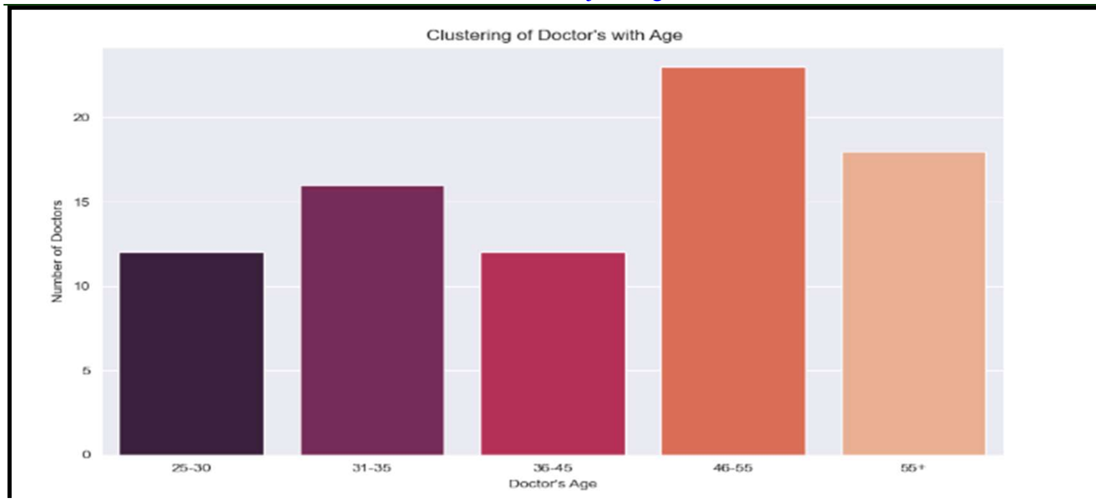
Figure 9: Sample Patient Description

id	Name / Description	GENDER	AGE	EXPERIENCE	RATING
1	Dr. Prabhakar Koregol -Cardiologist - Bangalore.	MALE	AGE-45	Experience : 15- Years	RATING-5*
2	Dr. Jayaranganath -General Medicine, Doctor for Cardiologist - Bangalore.	MALE	AGE-57	Experience : 26-Years	RATING-4*
3	Dr. Anand Kumar -Best Doctor for Cardiologist -Bangalore.	MALE	AGE-58	Experience : 20-Years	RATING-3*
4	Dr. Sridhara -General Medicine, Cardiologist -Bangalore.	MALE	AGE-50	Experience : 16-Years	RATING-4*
5	Dr. Nagesh -General Medicine, Cardiologist -Bangalore.	MALE	AGE-55	Experience : 25-Years	RATING-5*
6	Dr. Roopa -General Medicine, Cardiologist -Bangalore.	FEMALE	AGE-50	Experience : 22-Years	RATING-4*
7	Dr. Prahlad H Yathiraj-Radiation Oncologist - Chennai.	MALE	AGE-40	Experience : 9-Years	RATING-5*
8	Dr. Ranjan Kumar Mohapatra -General Medicine, Medical Oncologist- Chennai.	MALE	AGE-55	Experience : 25-Years	RATING-5*
9	Dr. Veda Padma Priya -General Surgery, Surgical Oncologist – Chennai.	FEMALE	AGE-35	Experience : 8-Years	RATING-4*
10	Dr. Ashwathy Susan Mathew -Radiation Oncologist - Chennai.	FEMALE	AGE-50	Experience : 25-Years	RATING-4*
11	Dr. Sumana-Radiation Oncologist-Chennai.	FEMALE	AGE-55	Experience : 18 -Years	RATING-5*
12	Dr. B. Anis-Oncology-Trichy.	MALE	AGE-40	Experience : 12-Years	RATING-5*
13	Dr. M. Mangala Devi-Radiation Oncologist-Trichy.	FEMALE	AGE-50	Experience : 20-Years	RATING-5*
14	Dr. B. Vijayasekaran-Cardiologist -Trichy.	FEMALE	AGE-55	Experience : 16-Years	RATING-4*

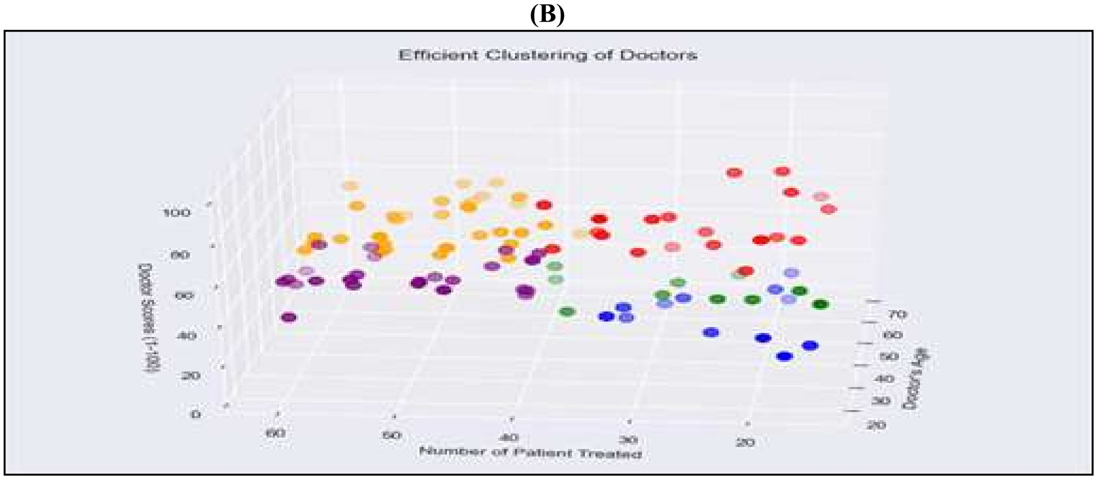
Figure 10: Doctor Recommendation List With Gender, Age, Experience And Score Rating



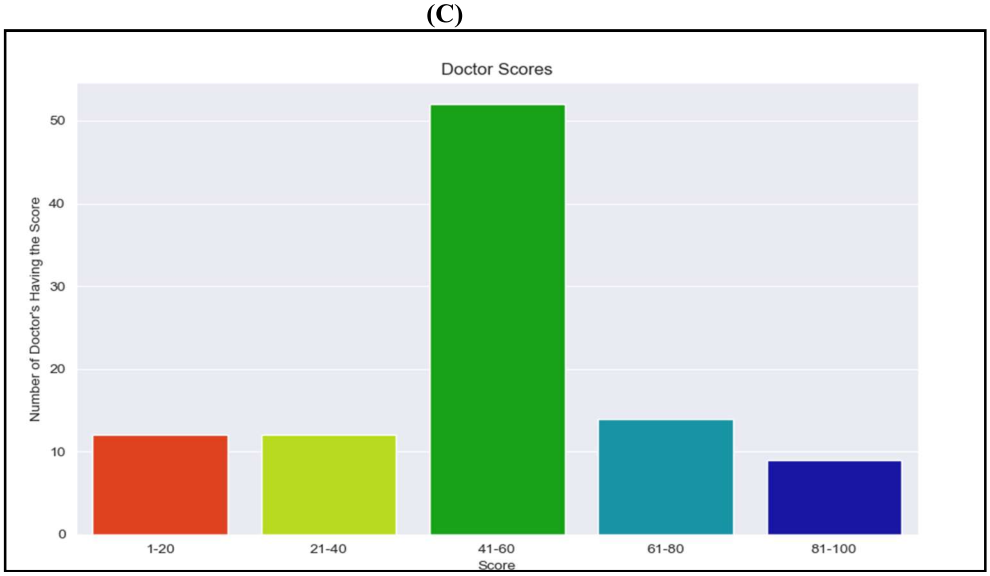
(A)



(B)



(C)



(D)

Figure 11: Doctor Recommendation Filtering With (A) Gender (B) Age (C) Experience (D) Score Rating

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The results of our study revealed that our proposed model was more accurate than the heuristic baseline. It also suggested that the information collected from other patients could complement a patient's consultation history to provide more accurate recommendations.

5. CONCLUSION

The goal of this study is to develop a method that can help predict the development of breast cancer and provide a doctor's recommendation, which are commonly used in healthcare. The advantage of this system is that it can use both unstructured and structured data, which are not always available in the current methods. The paper shows that the proposed model performed better than other algorithms. It has been suggested that a system that collects information about breast cancer patients could help decrease the disease's incidence and cost.The proposed model has one major issue, which is its inability to evaluate new doctors. This means that it can only recommend certain historical doctors. In order to improve its accuracy and solve the cold start issue, the researchers will add features to the system, such as patient reviews and medical consultation categories. The researchers additionally plan to integrate various contextual information sources using deep learning techniques.

REFERENCES

[1] Al-Yaseen, W.L., Jehad, A., Abed, Q.A. and Idrees, A.K., 2021. The Use of Modified K-Means Algorithm to Enhance the Performance of Support Vector Machine in Classifying Breast Cancer. *International Journal of Intelligent Engineering & Systems*, *14*(2).

[2] TsehayAdmassuAssegie, S.S., 2020. A Support Vector Machine and Decision Tree Based Breast Cancer Prediction. *International Journal of Engineering and Advanced Technology (IJEAT), ISSN*, pp.2249-8958.

[3] BERAHIM, M., SAMSUDIN, N.A. and MUSTAPHA, A., 2022. CORRELATION-BASED FEATURE SELECTION WITH BAG-BASED FUSION SCHEME FOR MULTI-INSTANCE LEARNING APPLICATION. *Journal of Engineering Science and Technology*, *17*(6), pp.3940-3955.

[4] Assiri, A.S., Nazir, S. and Velastin, S.A., 2020. Breast tumor classification using an ensemble machine learning method. *Journal of Imaging*, *6*(6), p.39.

[5] Jin, Z., Zhang, Z., Demir, K. and Gu, G.X., 2020. Machine learning for advanced additive manufacturing. *Matter*, *3*(5), pp.1541-1556.

[6] Mosayebi, A., Mojaradi, B., BonyadiNaeini, A. and Khodadad Hosseini, S.H., 2020. Modeling and comparing data mining algorithms for prediction of recurrence of breast cancer. *PloS one*, *15*(10), p.e0237658.

[7] Smiti, A., 2020. When machine learning meets medical world: Current status and future challenges. *Computer Science Review*, *37*, p.100280.

[8] Mohammadi, L., Einalou, Z., Hosseinzadeh, H. and Dadgostar, M., 2021. Cursor movement detection in brain-computer-interface systems using the K-means clustering method and LSVM. *Journal of Big Data*, *8*(1), pp.1-15.

[9] Pisner, D.A. and Schnyer, D.M., 2020. Support vector machine. In *Machine learning* (pp. 101-121). Academic Press.

[10] Pumplun, L., Fecho, M., Wahl, N., Peters, F. and Buxmann, P., 2021. Adoption of machine learning systems for medical diagnostics in clinics: qualitative interview study. *Journal of Medical Internet Research*, *23*(10), p.e29301.

[11] Kumar, A., Sharma, G.K. and Prakash, U.M., 2021. Disease prediction and doctor recommendation system using machine learning approaches. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, *9*, pp.34-44.

[12] Ragab, D.A., Sharkas, M., Marshall, S. and Ren, J., 2019. Breast cancer detection using deep convo-lutional neural networks and support vector machines. PeerJ 7: e6201.

[13] Botlagunta, M., Botlagunta, M.D., Myneni, M.B., Lakshmi, D., Nayyar, A., Gullapalli, J.S. and Shah, M.A., 2023. Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms. *Scientific Reports*, *13*(1), p.485.

[14] Shafique, R., Rustam, F., Choi, G.S., Díez, I.D.L.T., Mahmood, A., Lipari, V., Velasco, C.L.R. and Ashraf, I., 2023. Breast cancer prediction using fine needle aspiration features and upsampling with supervised machine learning. *Cancers*, *15*(3), p.681.

[15] Yan, Y., Yu, G. and Yan, X., 2020. Online doctor recommendation with convolutional neural network and sparse inputs. *Computational Intelligence and Neuroscience*, *2020*.