DARSGAH-E-AHLEBAIT

darsgah-e-ahlebait
Center for Advanced Islamic & Modern Scientific Research

Pathological Response Predictor Method (PReP Method): to Predict the Pathological Condition of Cancer Patients Rejected During Predictive Analysis

Muhammad Rehan Abbas[1], Malik Sajjad Ahmed Nadeem[2], Azhar Abbas Khan[3], Syed Rizwan Abbas[4], Attiya Batool[5]

1. Department of Information Technology, University of Layyah
2. Department of Computer Science and IT, The University of Azad Jammu & Kashmir, Muzaffarabad
3. Department of Entomology and Food Science, University of Layyah
4. Department of Biological Sciences; Karakorum International University; Gilgit
5. Lincoln University College, Malaysia

Corresponding Author: Muhammad Rehan Abbas; Email: dr.mrabbas@gmail.com

Abstract

Reject option suggests that, during machine learning predictive analysis, samples for which a posteriori probabilities are insufficiently high should not be classified. A rejection region is defined in the feature space and all examples within this region are rejected by the classifier. The classifier rejects an example if the prediction is not sufficiently reliable and falls into the rejection region. In case of medical diagnosis, the rejected sample is actually the record of a patient which is forwarded to the physician(s) for further analysis and necessary action. A team of physicians is involved to investigate the patient's record by using multiple resources and medical tests. It requires more time and effort to accomplish the task. Consequently, a cost has to be beard in this entire scenario. In this work, we introduce Pathological Response Predictor Method (PReP Method) to help the physician(s) for further investigation on rejected samples during predictive analysis. PReP Method compares the distribution of a feature from the set of rejected samples to the distribution of the same feature in the test set. Differences between the distributions of a feature indicate why the sample(s) was rejected. Experiments were conducted on six publicly available clinical cancer datasets using five state of the art machine learning classification algorithms. Varying distributions reveals different aspects of analysis leading the physician to analyze and predict the pathological condition of the patient and indicate the factors for which rejection has encountered.

Keywords: PReP Method, cancer diagnosis, classification, reject option, predictive analysis, Machine Learning

1. Introduction

1.1 Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed (Sharma & Rani, 2021). The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide (Arif & Abbas, 2024; Urooj & Abbas, 2024). The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

1.2 Use of Machine Learning in Medical Diagnosis

Machine learning can positively impact patient care delivery strategies. For example, it can help clinicians identify, diagnose and treat disease. Applications of machine learning in healthcare can also streamline healthcare tasks and optimize surgery planning, preparation and execution (Ibrahim & Abdulazeez, 2021).

1.2.1 Use of Machine Learning for Cancer Diagnosis and Prediction

Recent developments in machine learning can help increase healthcare access in developing countries and innovate cancer diagnosis and treatment. Over the past two decades, machine learning techniques have been used extensively for cancer diagnosis (Abbas et al., 2019; Masud, Sikder, Nahid, Bairagi, & AlZain, 2021; Sharma & Rani, 2021; Torkey, Atlam, El-Fishawy, & Salem, 2021; Vaka, Soni, & Reddy, 2020). Scientists applied different methods, such as screening in an early stage, in order to find types of cancer before they cause symptoms (Abbas, Nadeem, & Abbas, 2020; Abbas, Nadeem, Aziz, Shaheen, & Saeed, 2018; Cruz & Wishart, 2006; Esteva et al., 2017; Kourou, Exarchos, Exarchos, Karamouzis, & Fotiadis, 2015; Rejani & Selvi, 2009; Roy & Tiirikainen, 2020; Syed, Jabeen, & Manimala, 2018). Moreover, they have developed new strategies for the early prediction of cancer treatment outcome. With the advent of new technologies in the field of medicine, a large amount of cancer data have been collected and available to the medical research community. However, the accurate prediction of a disease outcome is one of the most interesting and challenging tasks for physicians. As a result, machine learning methods have become a popular tool for medical researchers (Rudzińska et al., 2021). These techniques can discover and identify patterns and relationships between them, from complex datasets, while they are able to effectively predict future outcomes of a cancer type.

1.3 Reject Option in Classification

As medical diagnosis is a sensitive issue due to involvement of human being's life, so for practical applications in medicine, it is better for predictions to be made only when they are sufficiently reliable, with no prediction made by the classifier in other cases. In practice, this results in the addition of a "reject" option to the classifiers (Hanczar & Dougherty, 2008; Nadeem, Zucker, & Hanczar, 2010). If the category to which an example should be classified cannot be predicted reliably as shown in
Figure 1, the classifier rejects the observation by not assigning it any of the class labels. The reject option introduced by Chow (1970) suggests that, results in decisions not being taken for samples for which confidence is low, to reduce the likelihood of error. This phenomenon is illustrated in Figure 1.

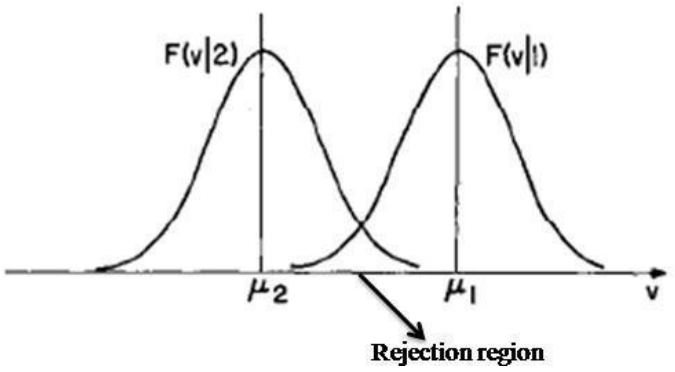

Figure 1: Two class example in a normal distribution. Figure source: (Chow, 1970).

Several people used reject option technique in machine learning predictive analysis. Dubuisson & Masson (1993) extended reject option proposed by Chow (1970) by defining two different types of reject option: an ambiguity reject option and a distance reject option. Fumera & Roli (2002) addressed the problem of implementing the reject option in SVMs. They suggested that, under the framework of the structural risk minimization principle, the rejection region must be determined during the training phase of a classifier. Landgrebe et al (2006) performed an analytic study on a controlled experiment followed by some experiments on real-world data sets with the distance-based reject-option classifier. Hanczar & Dougherty (2008) proposed an algorithm that aimed to produce a classifier with the smallest error rate possible. Yuan and Wegkamp (2010) stated the problem of binary classification with a reject option in which one can withhold the decision of classifying an observation at a cost lower than that of misclassification. Pillai et al., (2013) considered multi-label classification problems in application scenarios where classifier accuracy did not satisfactory, but manual annotation was too costly. In particular, multi-label accuracy measures were extended to take into account rejections and defined manual annotation cost as a cost function. Abbas et al., (2019) introduced Accuracy Rejection Normalized Cost Curves (ARNCCs) a three dimensional visualization technique to illustrate the performances of reject option based classifiers over different rejection regions and normalized cost to select the robust classifier.

1.4 Proposed Technique

As the rejected sample is referred for further analysis and investigation, it needs more time and effort to reach the final decision. Manual diagnosis requires more resources, medical tests, time and cost. There should be a system which may help the physicians to conduct further investigation on rejected samples. To the best of our knowledge no one proposed such a technique which may be able to take into account the further analysis on rejected samples. In this work, we proposed a methodology; Pathological Response Predictor Method (PReP Method) to automate further analysis of rejected samples. PReP Method works by comparing the distributions of features from rejected samples to their respective distributions in the test set. As test set consists of accurate class labels, so it is a realistic approach to compare the distributions of features from rejected samples to their distributions in the test set.

2. Materials and Methods

2.1 Cancer Datasets

Six publicly available clinical cancer datasets were selected for this study. Four of them are from UCI machine learning repository (Bache & Lichman, 2013), and others are Lung Cancer from North Central Cancer Treatment Group (NCCTG) (Loprinzi et al., 1994) and Veteran's Administration Lung Cancer Trial (Kalbfleisch & Prentice, 1980). All datasets have repeatedly appeared in the machine learning literature. Wisconsin Breast Cancer Database (WBCD) was obtained from the University of Wisconsin Hospitals, Madison (Wolberg, Street, & Mangasarian., 1995) by UCI machine learning repository. It was created using Fine Needle Aspirates (FNA) of human breast tissue. Sixteen instances contain missing values which were removed and as a result, we have 683 samples. Wisconsin Diagnostic Breast Cancer (WDBC) was also obtained from the UCI Machine learning repository. It was created by Wolberg & Mangasarian (1995) during the diagnosis of different patients. Wisconsin Prognostic Breast Cancer (WPBC) dataset is actually the follow-up of original breast cancer dataset by Wolberg & Mangasarian (1995). The Breast-cancer dataset is obtained from UCI Machine learning repository. It was donated by Zwitter & Soklic (1988) from University Medical Center Institute of Oncology in Ljubljana, Yugoslavia. This dataset is one of the three domains provided by the Oncology Institute.

NCCTG Lung Cancer dataset is obtained from North Central Cancer Treatment Group (Loprinzi et al., 1994). The dataset describes survival in patients with advanced lung cancer. It was collected from the patients completed questionnaires in 1994 by the North Central Cancer Treatment Group. Size of the dataset after removing missing value attribute samples is 167 where each sample contains 10 variables. We used *status* variable as a response, where 1 for censored and 2 for dead. Veteran's Administration Lung Cancer Trial Dataset contains the records of patients which is actually the survival analysis from

cancer. The source of the dataset is Veteran's Administration Lung Cancer Trial, which was first published by Kalbfleisch & Prentice (1980). The baseline characteristics of the data used in the study are presented in Table 1.

Table 1. Baseline characteristics of datasets used in the study.

Dataset	#Instances	#Attributes	+ive class count	-ive class count
WPBC	194	33	46	148
WDBC	569	30	212	357
WBCD	683	9	239	444
Breast-cancer	286	9	85	201
NCCTG lung cancer	167	9	47	120
VA lung cancer	137	7	09	128

2.2 Methodology

Working of proposed technique can be well understood by algorithm as follows:

Algorithm of Proposed PReP Method

Input: $(\chi, BP, Rs_j, F_iRs_j, ts_i^{neg}, ts_i^{pos}, clAlgo, rejL_p)$ where

- $\chi = \{x^r, y^r\}_{r=1}^{N}$ is a matrix of original data where x^r represents the feature space, y^r the target, while r denotes an example in the dataset χ.
- F_n *is the total number of features*
- BP; Box Plot
- $clAlgo$; a list of traditional learning algorithms without RO.
- Rs_j is the set of rejected samples at j^{th} rejection level
- ts_i^{neg} is the distribution of feature *i* in negative class member subset of test set
- ts_i^{pos} is the distribution of feature *i* in positive class member subset of test set
- F_iRs_j is the distribution of feature *i* in rejection set *j*
- $rejL_j$ is the list of specific rejection levels

Output: a set of box plot graphs showing distributions i.e., ts_i^{neg}, ts_i^{pos}, and F_iRs_j.

1. Split data χ into k folds i.e. $D_k = \chi_{k=1}^{n}$
2. **Repeat** steps 3 to 13 for each fold *k*, where $k := 1$ to n
3. $tr_k := D_{n-1}$ ▷ train set
4. $ts_k := D_n$ ▷ test set
5. $learnedModel := model(tr_k)$
6. $predictions := predict(learnedModel, ts_k)$
7. **Repeat** steps 8 to 13 for each *t*, where $t := 0$ to 1 with 0.002 increment
8. **Repeat** steps 9 to 13 for each *j*, where $j := 1$ *to* d, d is the size of original dataset *x*
9. $probDiff := probC1 - probC2$, where *probDiff* is the difference of probabilities of Class1 and Class2.
10. $if(probDiff \leq probC1\ and\ probDiff \geq probC2)$
11. $rejDf := x[j]$, *rejDf* is rejected set of samples
12. *else*
13. Assign class label
14. **Repeat** steps 12 to 16 for each $rejL_p$ rejection level
15. **Repeat** steps 13 to from 1 to f_n for each feature
16. Draw box plot *BP* using distributions $F_iRs_j, ts_i^{neg}, ts_i^{pos}$

The proposed technique works as first, a classifier Cl is induced by some machine learning algorithm to make predictions for a given test data. Resulting model M is capable to classify and reject the samples in given circumstances. Rejection statistics are calculated on the basis of rejection threshold t ranging from 0 to 1 consists of 500 rejection regions. In this way, a classifier rejects the samples from 0% to 100%. p arbitrarily rejection levels 5%, 10%, 20%, 30% and 40% were selected for analysis purposes. There is no restriction to select only these rejection levels. At certain rejection levels, there might be some rejected samples which are thrown into a pre-defined rejection space (rejection level). Finally, the distributions of feature i in test set as negative class member tsi neg and positive class member tsi pos were compared with the distribution of same feature FiRsj in rejection set. This process was repeated for all features of rejected samples. Five commonly used state of the art machine learning classifiers were used in this study. Selected classifiers were, Support Vector Machines with linear kernel (SVML), Support Vector Machine with radial kernel (SVMR) (Cortes & Vapnik, 1995), Linear Discriminant Analysis (LDA) (Fisher, 1936), Random Forest (RF) (Ho, 1995, 1998) and k-Nearest Neighbors (kNN) (Altman, 1992).

Figure 2 is the demonstration of experimental design of the proposed methodology.

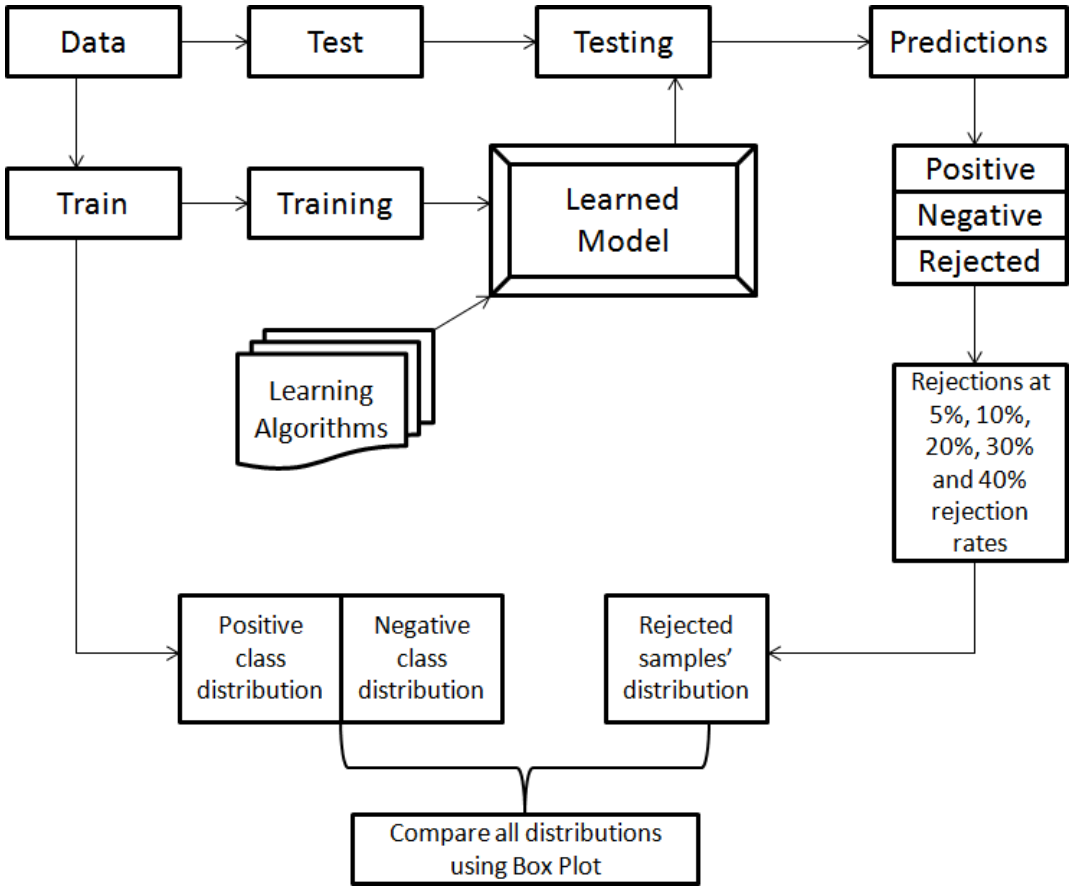

Figure 2: Experimental design of the proposed methodology.

3. Results and Discussions

This section presents the results of the proposed technique. Experiments were conducted at 5%, 10%, 20%, 30% and 40% rejection levels. Some typical results are presented here.

3.1 Results of Wisconsin Brest Cancer Database (WBCD)

Following results show the comparisons of distributions of feature '*cell_thickness*' for WBCD dataset using SVML classifier. Figure 3 (A) is a comparison of distributions without rejection and rejected samples. Without rejection, majority of the values are between 2 to 6, whereas 50% of the values of rejected samples are between 4 to 7. Figure 3 (B) shows the distributions of feature '*cell_thickness*' from rejection set and test set (as negative and positive class membership). Labels 'no_rej (N)' and 'no_rej (P)' represent the distributions of feature '*cell_thickness*' without rejection from test set and 'Rejected' is the distribution of same feature in the set of rejected samples at 5% rejection level.

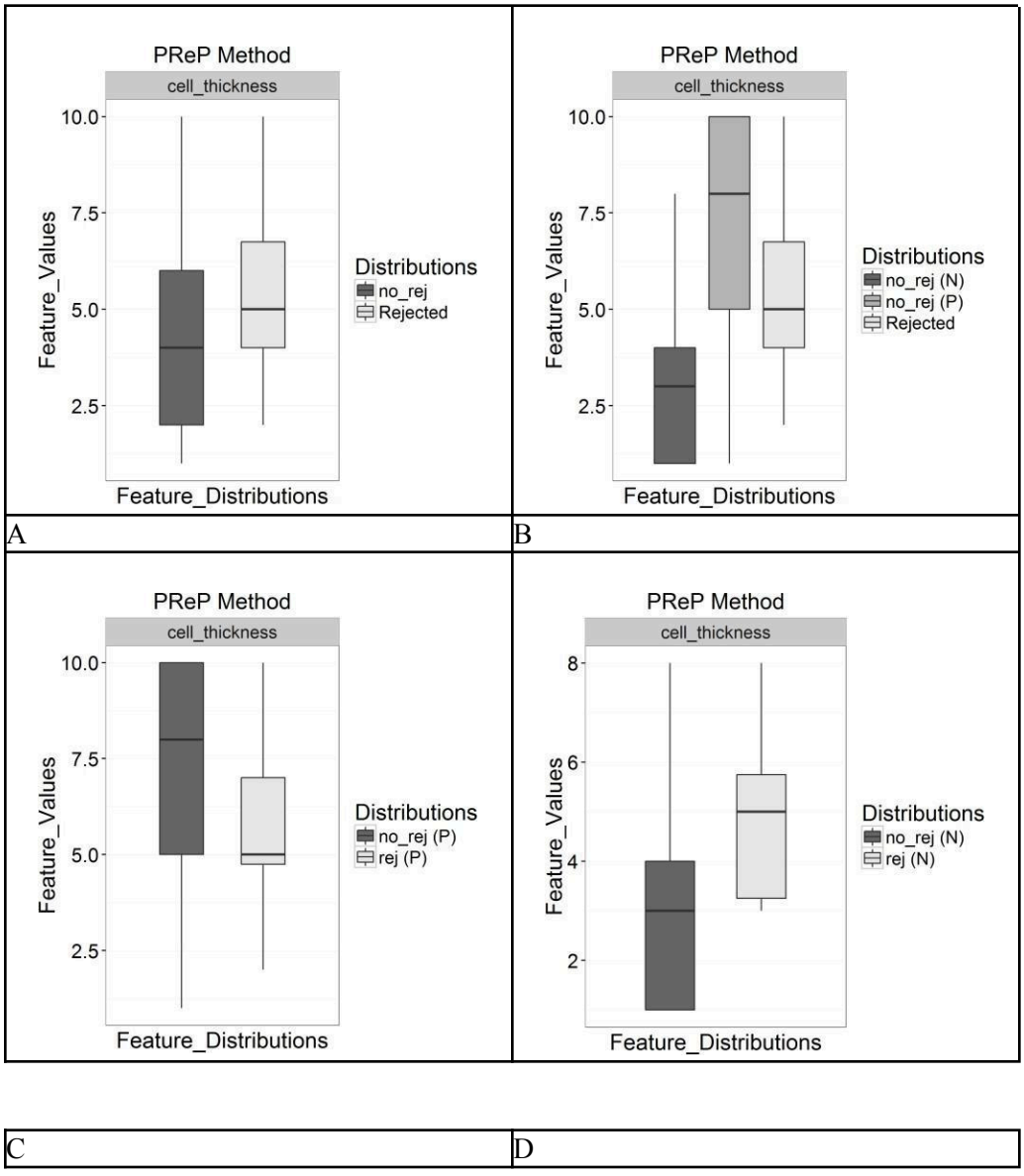

Figure 3: Distributions of feature '*cell_thickness*' at 5% rejection level based on the result of SVML for WBCD dataset.

Negative class membership distribution of feature '*cell_thickness*' shows 75% values are between 1 to 4 and median is 3. Similarly, for Positive class distribution of feature '*cell_thickness*', 75% of the distribution is between 5 to 10, and the median is 8. In rejected samples, half of the distribution of '*cell_thickness*' is between 4 to 6.5 and median is at 5. Upper whisker shows 25% of the distribution is above 6.5. The distribution of feature '*cell_thickness*' in rejected samples is close to positive class. The

analysis done for all features at 5 percent rejection and shown in Figure 5. All the remaining features including '*cell_thickness*' have been shown there. Results reveal that majority of the rejected samples at 5% rejection belongs to positive class.

Figure 3 (C) is the comparison of distributions as positive class membership of feature '*cell_thickness*' without rejection and rejected samples. Similarly, Figure 3 (D) is the comparison of distributions as negative class membership of feature '*cell_thickness*' without rejection and rejected samples. Figure 3 (C) and Figure 3 (D) validate the results by showing relevancy between actual and rejected samples' distributions. At 5% rejection majority of values of distributions from rejected samples are close to positive class. To validate the results, we observed that, there are 34 samples rejected at 5% rejection rate, from which 24 rejected samples belong to positive class and 10 samples belong to negative class.

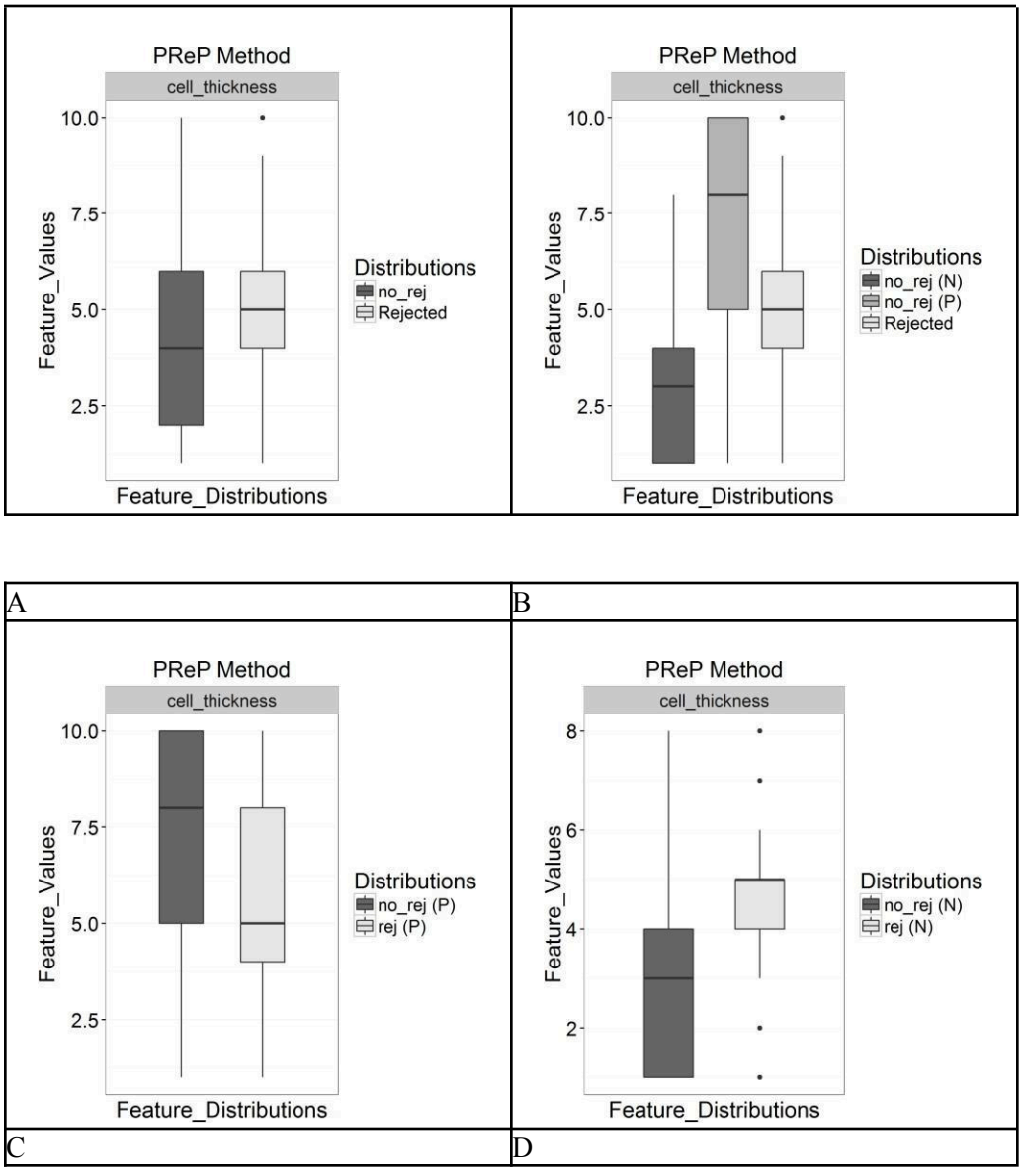

Figure 4: Distributions of feature *'cell_thickness'* at 20% rejection level based on the result of SVML for WBCD dataset.

Figure 4 shows the results at 20% rejection level for WBCD dataset using SVML classifier. Figure 5 shows the comparison of the distributions of all features from rejected samples at 5% rejection, with their distributions without rejection as positive and negative class membership.

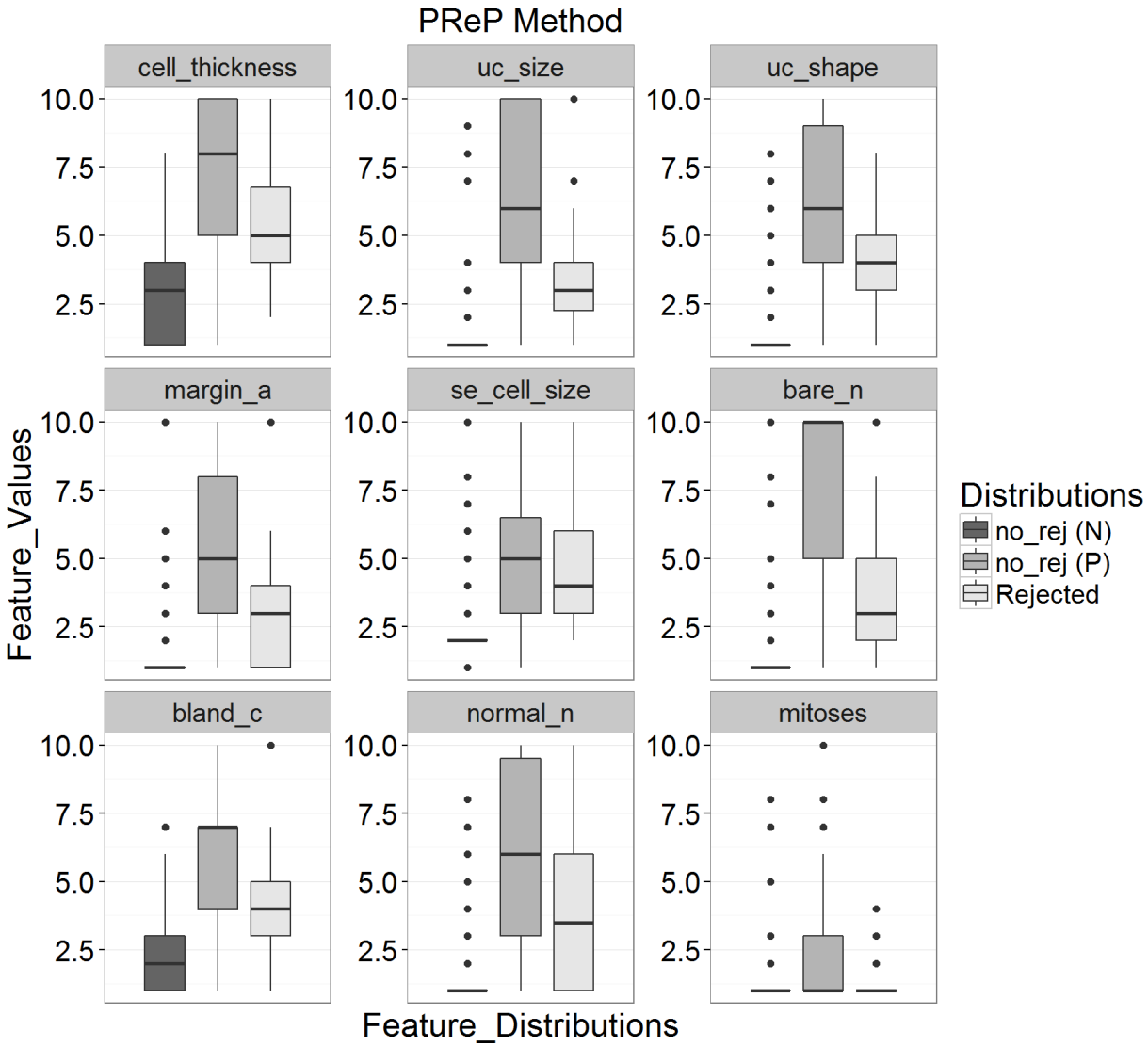

Figure 5: Comparing the distributions of all features from rejected samples at 5% rejection, with their distributions in test set without rejection as positive and negative class membership. The classifier used SVM Linear for the dataset WBCD.

3.2 Results of Wisconsin Prognostic Breast Cancer (WPBC)

Results of WPBC obtained on the bases of repeating all previous steps that were performed earlier for WBCD dataset. There are 198 observations after removing the samples containing 'NA' values. There are 148 negative and 46 positive observations.

Figure 6 is the representation of results based on computations at 5% rejection level for feature '*tumor_size*'. Figure 6 (A) is the comparison between non-rejected and rejected samples' distributions. Without rejection, 75% values exist between 1 to 3.5. On the other hand 75% values of feature '*tumor_size*' from rejected samples are ranged in 3 to 7.5. Figure 6 (A) is an initial investigation telling the overall picture. Figure 6 (B) exhibits the more specific details by comparing the distribution of feature '*tumor_size*' with its distributions as negative and positive class membership in the test set. Box plot representing negative class distribution shows, 75% values are below 3.5 but on the other hand positive distribution shows, 75% values lie between
2.5 to 6.0. Distribution of feature '*tumor_size*' in the rejected samples is much closer to the positive class membership distribution. This is not a final decision. After analyzing all features, physician(s) may be able to set a direction for classifying the rejected samples. Anyway, by analyzing the individual features, physician(s) may be able to predict the pathological condition of the patient for specific or all disease factors.

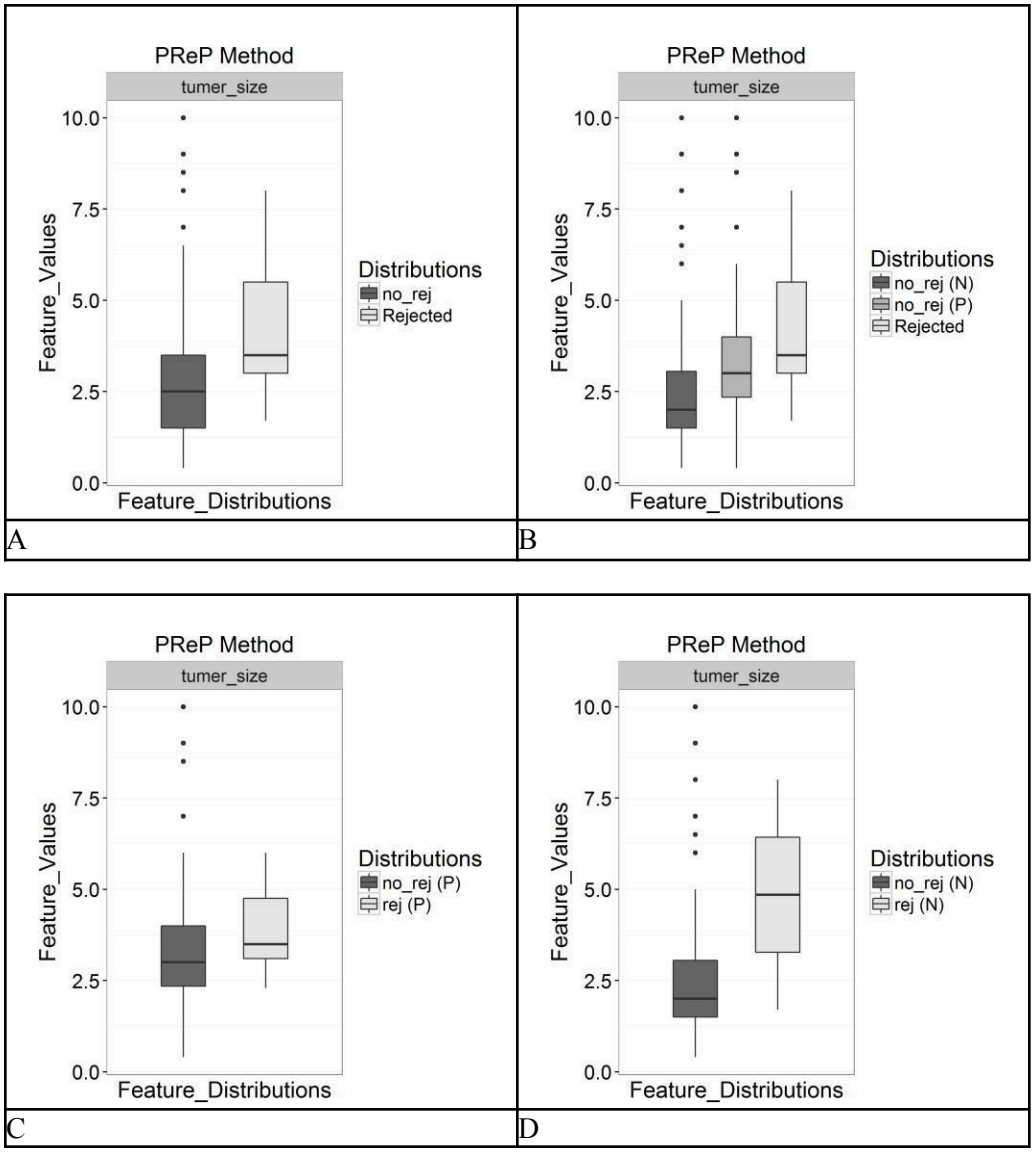

Figure 6: Distributions of feature *'tumor_size'* at 5% rejection level based on the result of SVML for WPBC dataset.

Figure 7 represents the results of WPBC dataset using SVML classifier on 20% rejection rate for the feature *'tumor_size'*.

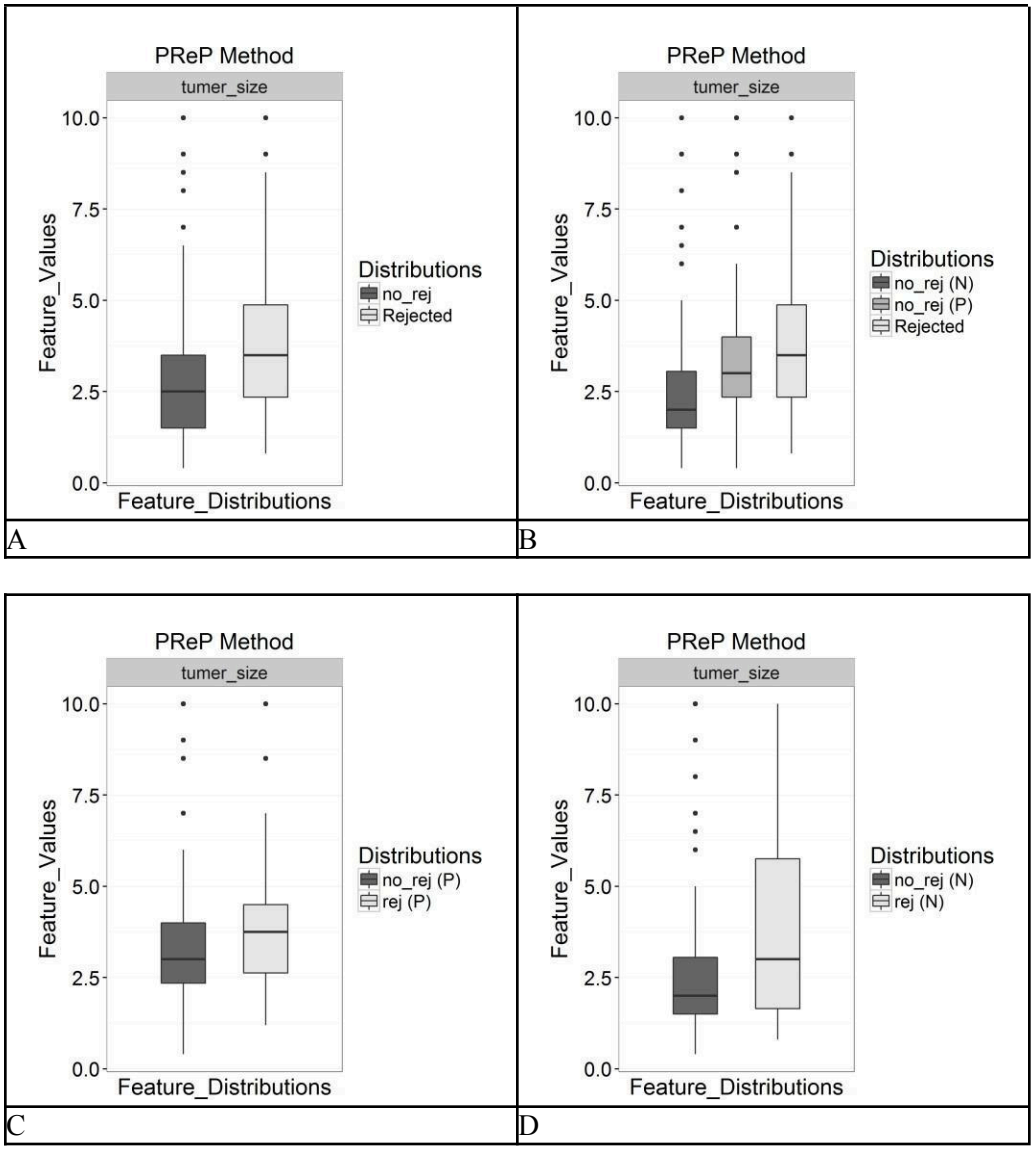

Figure 7: Distributions of feature 'tumor_size' at 20% rejection level based on the result of SVML for WPBC dataset.

Figure 8 is a complete picture of distributions of all features at 5% rejection level. It is very clear that majority of the distributions are close to the distribution of positive class. At 5% rejection rate, there are 9 rejected samples in which 7 samples belong to *'Positive'* class and 2 from *'Negative'* class.

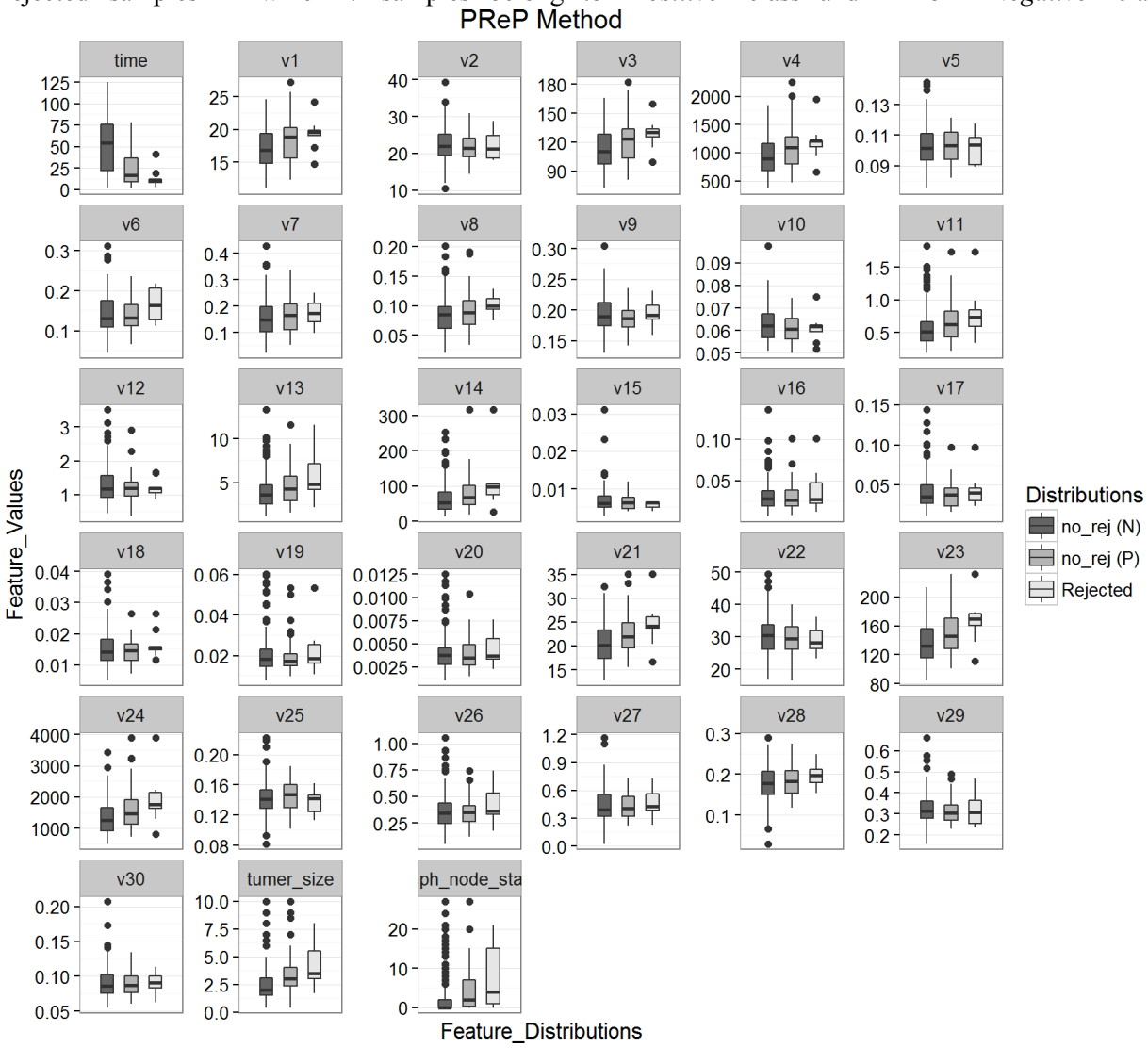

Figure 8: Comparing the distributions of all features from rejected samples at 5% rejection, with their distributions in test set without rejection as positive and negative class membership. The classifier used SVM Linear for WPBC dataset.

3.3 Discussions

Classification with reject option enhances the predictive ability of a classifier by rejecting the doubtful samples on a certain rejection threshold. Particularly in the domain of healthcare, it is recommended that the doubtful prediction should not assign a class label and it must be referred to the physician for further investigation. Rejected samples need more time and effort for additional tests and examinations. PReP Method helps the physicians for further investigations on rejected samples. Different distributions of rejected samples show that, the reason why they were rejected. The physician may suggest the patient wait for some time till the next medical checkup. The test set contains the precise and accurate feature values to assure correct class labels. As PReP Method compares the distribution of a feature from rejected samples to its distribution in the test set, it may help the physician to predict the current pathological condition of the patient. It also works as a helping tool to classify the rejected samples.

4. Conclusion

In this work, we presented Pathological Response Predictor Method (PReP Method) to predict the pathological condition of patient's who were rejected during predicted analysis. Reject option enabled classifiers reject the sample(s) not attaining the required confidence level to make a correct decision. In the domain of healthcare, rejected samples are referred to the physicians or surgeons for further necessary action. Physicians have to investigate the rejected samples (patients) by using manual methods. The manual procedure takes much time and efforts to reach the final decision.

Our proposed PReP Method helps the physicians to further in-depth analysis of the rejected samples. Five commonly used classification algorithms along with six publicly available clinical cancer datasets were used for experimentation. First, a classifier was induced by using 10-fold cross-validation to build a predictive model. Predictions were obtained after applying test data. Rejection statistics were computed on 5%, 10%, 20%, 30% and 40% rejection rates. Distribution of the feature from rejected samples was compared to the distributions (distribution of feature as positive class membership and distribution of feature as negative class membership) of the same feature in the test set. Results reveal that proposed PReP Method is applied nature and it is very useful in cancer diagnosis procedure.

The PReP Method is not limited to cancer diagnosis, it is also applicable to other diagnostic and prognostic problems. Beyond healthcare domain, PReP Method can also be used for other real-life problems. Currently, we designed the methodology for binary problems, it may be extended to multi-class problems for different domains.

References

Abbas, M. R., Nadeem, M. S. A., & Abbas, S. R. (2020). Effect of data normalization on the performance of classification algorithms. *Journal of Natural Sciences, 8*(2), 104–110.

Abbas, M. R., Nadeem, M. S. A., Aziz, W., Shaheen, A., & Saeed, S. (2018). Performance comparison of classifiers using clinical observations in reject option scenarios to detect cancerous subjects. *Indian Journal of Science and Technology, 11*(39), 1–9.

Abbas, M. R., Nadeem, M. S. A., Shaheen, A., Alshdadi, A. A., Alharbey, R., Shim, S.-O., & Aziz, W. (2019). Accuracy rejection normalized-cost curves (ARNCCs): A novel 3-dimensional framework for robust classification. *IEEE Access, 7*, 160125–160143.

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician, 46*(3), 175–185.

Arif, M., & Abbas, S. R. (2024). Chemoinformatics algorithms: Unlocking nature's potential. *JNS, 12*(2), 39–46. https://doi.org/10.17605/OSF.IO/J4GBC

Bache, K., & Lichman, M. (2013). *UCI machine learning repository.*

Chow, C. K. (1970). On optimum recognition error and reject tradeoff. *IEEE Transactions on Information Theory, 16*(1), 41–46.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Cruz, J. A., & Wishart, D. S. (2006). Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics, 2*, 117693510600200030.

Dubuisson, B., & Masson, M. (1993). A statistical decision rule with incomplete knowledge about classes. *Pattern Recognition, 26*(1), 155–165.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017).

Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115–118.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7*(2), 179–188.

Fumera, G., & Roli, F. (2002). Support vector machines with embedded reject option. In *Pattern Recognition with Support Vector Machines*, 811–819.

Hanczar, B., & Dougherty, E. R. (2008). Classification with reject option in gene expression data. *Bioinformatics, 24*(17), 1889–1895.

Ho, T. K. (1995). Random decision forests. In *Proceedings of the Third International Conference on Document Analysis and Recognition*, 278–282.

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(8), 832–844.

Ibrahim, I., & Abdulazeez, A. (2021). The role of machine learning algorithms for diagnosing diseases. *Journal of Applied Science and Technology Trends, 2*(1), 10–19.

Kalbfleisch, J., & Prentice, R. (1980). Veteran's Administration Lung Cancer Trial.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal, 13*, 8–17.

Landgrebe, T. C., Tax, D. M., Paclík, P., & Duin, R. P. (2006). The interaction between classification and reject performance for distance-based reject-option classifiers. *Pattern Recognition Letters, 27*(8), 908–917.

Loprinzi, C. L., Laurie, J. A., Wieand, H. S., Krook, J. E., Novotny, P. J., Kugler, J. W., ... Klatt, N. E. (1994). NCCTG Lung Cancer.

Masud, M., Sikder, N., Nahid, A. A., Bairagi, A. K., & AlZain, M. A. (2021). A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework. *Sensors, 21*(3), 748.

Nadeem, M. S. A., Zucker, J.-D., & Hanczar, B. (2010). *Accuracy-Rejection Curves (ARCs) for Comparing Classification Methods with a Reject Option.* Paper presented at the MLSB.

Pillai, I., Fumera, G., & Roli, F. (2013). Multi-label classification with a reject option. *Pattern recognition, 46*(8), 2256-2266.

Rejani, Y., & Selvi, S. T. (2009). Early detection of breast cancer using SVM classifier technique. *arXiv preprint arXiv:0912.2314*.

Roy, D., & Tiirikainen, M. (2020). Diagnostic power of DNA methylation classifiers for early detection of cancer. *Trends in cancer, 6*(2), 78-81.

Rudzińska, M., Daglioglu, C., Savvateeva, L. V., Kaci, F. N., Antoine, R., & Zamyatnin Jr, A. A. (2021). Current Status and Perspectives of Protease Inhibitors and Their Combination with Nanosized Drug Delivery Systems for Targeted Cancer Therapy. *Drug Design, Development and Therapy, 15*, 9.

Sharma, A., & Rani, R. (2021). A Systematic Review of Applications of Machine Learning in Cancer Prediction and Diagnosis. *Archives of Computational Methods in Engineering*, 1-22.

Syed, L., Jabeen, S., & Manimala, S. (2018). Telemammography: a novel approach for early detection of breast cancer through wavelets based image processing and machine learning techniques *Advances in Soft Computing and Machine Learning in Image Processing* (pp. 149-183): Springer.

Torkey, H., Atlam, M., El-Fishawy, N., & Salem, H. (2021). Machine Learning Model for Cancer Diagnosis based on RNAseq Microarray. *Menoufia Journal of Electronic Engineering Research, 30*(1), 65-75.

Urooj, S. S., & Abbas, S. R. (2024). Chemo-informatics Challenges: Tackling Big Data for Drug Discovery. *JBS, 12*(1), 18-24. doi: https://doi.org/https://doi.org/10.17605/OSF.IO/BYVQ5

Vaka, A. R., Soni, B., & Reddy, S. (2020). Breast cancer detection by leveraging Machine Learning. *ICT Express, 6*(4), 320-324.

Wolberg, D. W. H., Street, W. N., & Mangasarian., O. L. (1995). *Wisconsin Breast Cancer Database*.

Yuan, M., & Wegkamp, M. (2010). Classification methods with reject option based on convex risk minimization. *Journal of Machine Learning Research, 11*(Jan), 111-130.

Zwitter, M., & Soklic, M. (1988). Institute of Oncology University Medical Center Ljubljana, Yugoslavia -- Donors: Ming Tan and Jeff Schlimmer *(Jeffrey.Schlimmer@a.gp.cs.cmu.edu)* --.