International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 12 ǁ December 2022 ǁ PP. 556-563

Cancer Prognoses Classification using Ensemble on A combined Rough sets, chi2 and Correlation Models

Mohamed Taybe Elhadi[1] and Fathi Abdalla[2]
[1]*Department of Software Engineering, Zawia University, Zawia, Libya*
[2] *Bahrain Royal Medical Services (BDF) , Department of Pathology*

ABSTRACT
This paper repots on experiments performed combining techniques from rough sets theory and machine learning known classifiers with ensemble to help select and evaluate the important features from of a set of factors. One dataset was download and the rest were collected from local hospitals in Libya. Rough Sets reduction techniques were used as a bases in generating equivalent subsets of features (Referred to as Reducts) from among the available factors. The selected subsets of attributes are then compared and used as a bases for further augmentation by a number of other feature selection techniques, namely that of Chi2 and Correlation. The suggested approach showed very encouraging results. An improvement is obtained through the augmentation of rough sets with Chi2 and Correlation. The used approach was quite successful with variable training and evaluation accuracies. The experiments conducted have shown that rough set is capable of suggesting a reasonable initial subset (Reducts) that serve as bases for more improved subsets using the complementary techniques of Chi2 and Correlation. The set of alternative factors when used for classification have shown very good training and evaluation accuracies. A number of machine learning classifiers have been adopted and applied to the selected and prepared data. Further improvement is obtained through the application of ensemble process on the set of classifiers results. Random Forest have shown the best performance and combined datasets were the best when accuracies and ensample were collected.

KEYWORDS
Feature Selection, Rough Set Reducts, Chi Square, Correlation, Machine Learning Classifiers, Random Frost, Nearest Neighbour, Support Vector Machines, Logistic Regression

---

Date of Submission: 15-12-2022 Date of acceptance: 31-12-2022

I. INTRODUCTION

With the prevalence of huge datasets and the use of both statistical and machine learning classification techniques, features (attributes, variables, factors) selection has become an important part of any data modelling. The ability to optimize and reduce large feature set through the selection of most appropriate ones is of vital importance for the analysis and classification applications. Not only it increases efficiency of algorithms by the elimination of the redundant and noisy data, it also saves on important resources when feature measurements are created by providing alternative feature sets. Applications such as cancer evaluation and prognostication require good standardized systems to determine the grade of cellular abnormalities, the stage of spread, and the prediction of survival, all judged on the basis of a set of data attributes. Making a judgment on medical cases is a critical and very demanding work that needs human expertise.

Cancer grading involves describing abnormal and cancerous cells and tissue. It is done through the use of microscope to compare cells and tissue to original, healthy cells, while cancer staging is the process by which a a health expert like a doctor decides the person's level of cancer progression using diagnostic tests, imaging scans, and samples taken from surgery. Cancer survival is refer to the portion percentage of people who survive a certain type of cancer for a specified period of time [1].

Statistical and machine learning techniques can help the experts in deciding and in selecting important or relevant factors from among the available ones. With the ever-increasing size and availability of data both on the web and otherwise, there is more demand for useful statistical classification methods and machine learning techniques for use with such large data sources. The selection of the right features, factors, attributes or variables has become an important matter in data mining, knowledge discovery and machine leaning application, especially predictive approaches.

The motivation behind feature selection varies depending on work objectives, data sizes and usage constraints. In general, several reasons are behind the use of feature selection techniques including model simplification to make it easier to do interpretation [2], to best use time and resources, and [3] to minimize and reduce dimensionality to allow the use of models [4,5]. Of interest to us in this work is feature selection in

which a set of features are selected to provide equivalent discrimination power as that of the whole data set. Alternative feature selections, which are not of interest to us include dimensionality reduction.

The rest of the paper is made of section 2 on feature selection and classification techniques; section 3 on datasets; section 4 on Procedure followed; section 5 on Experiments conducted, results obtained and result discussions; final section 6 on conclusions and feature work.

II. FEATURE SELECTION AND CLASSIFICATION TECHNIQUES

The growth of data along with advanced mining and knowledge discovery tools along with the explosive increase of data sources makes for huge opportunities for businesses and alike. Machine learning algorithms have become an integral part of many data analyses especially classification and prediction. Approaches for machine learning are normally divided into three approaches [6]: (1) Supervised learning which is used if the available training data has a labelled attribute and other (new) data does not contain a label; (2) Unsupervised learning which has no labelled information, but the algorithms strive to discover any existing patterns or relationships in the data; (3) Deep learning which learns and improves using artificial neural networks with larger and complex networks that aid in classification problems.

2.1 Feature Selection Methods

The main aim in feature selection methods is to reduce the number of provided variables to those that are believed to be most useful to a model in order to predict the target variable. It works by removing redundancy and allowing informative predictors for the model. Large numbers of features can slow the development and training of models, put a demand on system resources both space and speed; they can also add uncertainty to the predictions and increase the overall effectiveness and cost of the model.

The type of data described by the features can also contribute to some categorization. There are those used to handle continuous data values and those that concentrated on discrete valued data or categorical data. Many techniques apply to both. Here, it is only concentrated on feature selection methods applicable to supervised methods on categorical data both binary and multivariate [7].

Feature selection have been categorized in many ways. One very common way of grouping of feature selection techniques is thorough adopted features selection which is normally divided into (1) filter-based, (2) wrapper based and (3) embedded methods [8].

2.1.1 Filter methods:

These are feature selection methods that tend to be less computationally intensive but can produce a feature set which is not tuned to a specific type of predictive model [9]. They provide lower prediction performance and so are useful for exposing the relationships between the features. They use certain measures to score a feature subset that is fast to compute but capable of capturing the utility of the subset. Examples of filter-based methods include the pointwise mutual information [10], mutual information [11], Pearson correlation coefficient [12], inter/intra class distance, relief-based algorithms [13] or the scores of significance tests for each class/feature combinations [11,14]. Filters are not tied to any particular induction algorithm. That is why they are applied in many domains.

2.1.2 Wrapper methods:

This method searches through the feature subset space using the estimated accuracy from an induction algorithm as a measure of subset suitability. Although wrappers may produce better results, they are expensive to run and can break down with large numbers of features. These algorithms are resource intensive and try to make available the best performing feature set for a particular type of model. They rank feature subsets where each new subset is used to train a model. Testing while tracking the error rate of the model makes the score for that subset. Recursive feature elimination RFE is an example of a wrapper feature selection methods [15].

2.1.2.1 Types of wrapper methods: SelectKBest which selects features based on k heist scores; Forward Selection which an empty set of features is setup. Then the best feature among the original set is determined and selected and we keep iterating until all best are selected; Backward Elimination which works in reverse to the previous technique. It starts with the full set of features and keeps removing the worst features remaining in the set; Recursive Feature Elimination RFE works by searching for a subset of features by starting with all features in the training dataset and successfully removing features until the desired number remains. Fitting of the given machine learning algorithm to be used in the core of the model, ranking features by importance, discarding the least important features, and re-fitting the model. This process is repeated until a specified number of features remains. The procedure is almost the same as in the case of backward elimination.

2.1.3 Embedded methods:
These are a catch-all subset of techniques. the linear model LASSO and its improvements in Bolasso [16] Elastic net regularization, and FeaLect are examples of this approach. They scored all the features based on combinatorial analysis of regression coefficients.[17].

2.2. Rough Sets and Reduction

Rough Set Theory (RST) introduced by Z. Pawlak in the 1980s is the base for many data analysis techniques for the discovery of any structural relationships within imprecise data. RST makes use of the idea of equivalence classes within the given training data where data tuples forming an equivalence class are indiscernible. Rough sets can define classes such as those class that cannot be distinguished in terms of the available attributes.

Other uses of rough set include attribute reduction and feature suggestion or selection. It works by removing any attributes that do not contribute to the classification of the given training data. RST has also become a tool for handling uncertainty expressed as indiscernibility between objects in a set. Full presentations of rough set and its many concepts is beyond the scope of this paper. Literature is full of references. One paper that presents the many concepts specially from perspective of features reduction or selection can be found in [18]. It contains basic concepts Approximation Space, Information Tables, Dependency Analysis and Data Reduction, Computation of Rules which for the bases of the theory and are implemented by a number of systems are discussed in literature.

2.3. Chi Square

Chi Square is statistical measure or test used to decide on the validity of a certain hypothesis by either accepting it or rejecting it depending on its weight and p-value. The Chi^2 measure is good at discovering of any relationship between any two attributes, the strength or significance of such relationship. In particular, it is useful on evaluation of correlation between input factors and an output factor [19]. Relationship with high significance and with a specific p-value (normally less than 0.05) are considered good and can confirm existence of relationship between any two factors.

2.4. Correlation

Correlation is statistical measure used to decide on the existence of any relationship (correlation) between a single or multiple factors and the decision or output factor. The range from positive one to negative one is the result of the correlation coefficient. Values in the zero level are considered insignificant and indicates no relationship between attributes. A positive value close to 1 is an indication of strong relationship while a value close to -1 is considered as a negative or a reversed correlation [20]. Presence of correlation between say input factors and output factor is an indication of significance and relationship between involved features or attributes.

2.5 Classification and Learning Models

The following is a brief mention of the machine learning classifiers used in our work to validate the proposed set of features and to suggest a number of prediction models.

- **Logistic Regression (LR):** Logistic regression, takes advantage of prior observations of a data set to predicts a binary outcome. LR is capable of producing probabilities and of classification of new data based on continuous and discrete datasets using the well-known sigmoid [21].
- **Naive Bayes (NB)**: is based on Bayes' theorem. The classifiers are fast and easy to implement, however, they are based on the simplistic assumption of independence of predictors. But it has its own uses applications such recommendation systems, spam filtering, sentiment analysis [22].
- **Artificial Neural Networks (ANN):** a Multi-layer Perceptron classifier that make use of base Neural Network to perform classification [23].
- **K-Nearest Neighbour (KNN):** The k-nearest neighbour is a classifier that makes use of proximity (similar points can be found near each other) where a class label is assigned on the basis of the label that is mostly found around a given data point [24].
- **Random Forest (RF):** is based on decision trees model which allow an orderly way to draw logical conclusions in various problem-solving context [25]. RF is grouping of decision trees with aggregated results into one representative result for the purpose of reducing overfitting while maintaining the error to lowest possible levels. This normally done by training on different samples of the data.
- **Support Victor Machines (SVM):** this is regression and classification model with the objective of finding a hyperplane in an N-dimensional space representing the number of features that distinctly classifies the data points. Support vectors are data points that are used to maximize the margin of the classifier closer to the hyperplane [26].

III. DATASETS AND COLLECTIONS

Many machine learning techniques are data-driven and data-cantered. They are trained to learn from exiting data and to analyse and classify unknown data. In this work, empirical data is used for the training and evaluation. Data used was collected from local medical hospitals and oncology institutions [27] except for one dataset, however, was downloaded form web. It was used for experimentation, pre-development and evaluation of the procedure, namely, Cancer-Large-Wisconsin [28]. Once the procedure is tried and evaluated using the downloaded dataset, it was applied to the local data made of four sets for three important decision attributes that are normally handled by a human expert. As in Table 1, the datasets are:

1. Breast cancer which is 569 rows by 32 columns (**BCW-0**)
2. Colon dataset with grade as decision attribute. It is made of 183-rows by 5 columns (short named **Colon-1**)
3. Colon dataset with Survival Period as decision attribute. It is made of 230-rows by 14 columns (short named **Colon-2**)
4. ICU cancer patients 273 rows by 8 columns (**ICU-3**).
5. Clinicopathology 153 rows by 17 columns (**CPath-4**)

IV. SUGGESTED PROCEDURE

The suggested procedure, as is shown in Figure 1 above, is made of a number of major phases which of which is made of some steps or tasks. The steps are taken to prepare data and select subsets of important features. Each major step is explained next:

4.1 Role of Rough Set Reducts: Base Subsets Generation

This a base stage where Rough Sets is used to generate a set of subsets that is supposed to provide an equivalent classification as the original data. The following tasks are performed in this phase:

(1) Using the rough set tool Rosetta [29] and its reduct generation algorithm [30] to create a number of reducts each of which is used as bases for the next phases. In particular, the algorithm is applied on the dataset to produce the set of reducts.

(2) A set of k-attributes are created using best attributes from attributes that were collected through Chi^2, where best is represented by p-values of less than .05.

(3) A third set was also created using correlation between the set of all features and the labeled feature. Those with absolute correlation of .20 or more were selected.

(4) Further combining of the created subsets are done to come up with various combinations of sets.

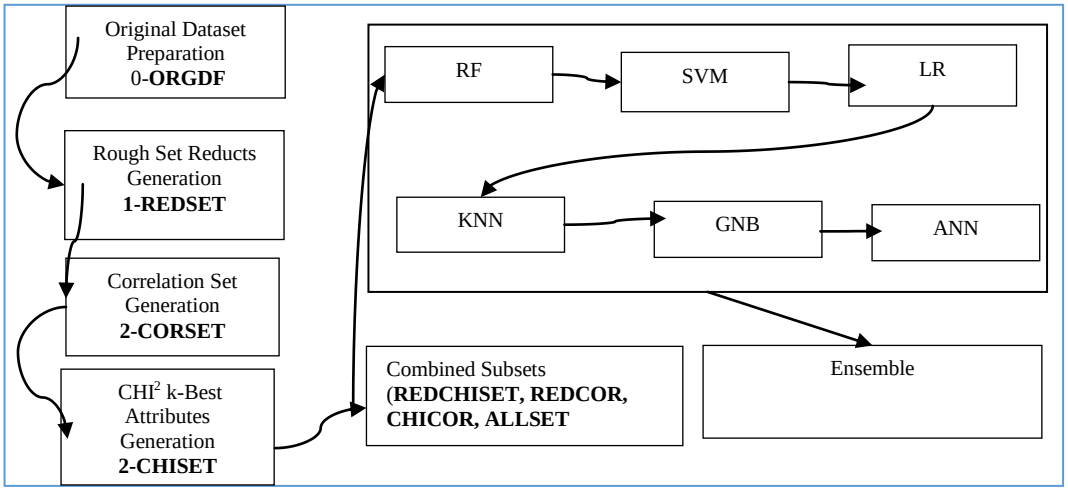

Figure 1: Suggested Procedure

The combined set of the previously created sets including **Reducts and Correlated**, **Reducts and Chi2**, **Chi2 and Correlated** and a final superset made of the combination from the **all** of the previous sets (**reducts, Chi'ed and correlated**).

4.2 Machine Learning Classification: Model Building

Ppreviously mentioned machine learning classifier (subsection 2.5) have been applied to the groups of data created including Random Frost (RF), Nearest Neighbor (KNN), Support Vector Machines (SVM), Logistic

Regression (LR), Neural Nets (ANN) and Bayes' Classifier (NB). All in all, we have the following sets of data each of which is complete and without any missing data:

1. **Original Dataset (0-ORGSET)**: this is the complete set of attributes as is.
2. Reducts Datasets (**1-REDSET**): these are a subset based on Rough Set reduction.
3. **Correlated Dataset (2-CORSET)**: produced using correlation methods with best k-attributes selected. The k is the size of the Reducts used. Only those attributes whose p-value is less than 0.05 are used.
4. **Chi2 Dataset (3-CHISET)**: This a Chi^2 best k-attributes selected, where the k is the size of the Reducts used.
5. The other four sets are made by the combining of the previous sets and are: (4-**REDCHISET** which is the combination of REDSET, and CHISET; 5-**REDCOR** which is a combination of REDSET and CORSET, 6-**CHICOR** which is a combination of **CHISET** and **CORSET**; finally, the set of all sets (7-**ALLSET**).

In total we have five datasets and seven subsets for each of the five datasets totaling 35 datasets. Each subset is subjected to six classification algorithms.

4.3 Evaluations of Selected Features

For each group of features including the complete data set and the set of classifiers adopted, a cross validation was performed collecting the (1) Mean Training Accuracy (**MTA**), and (2) Mean Validation Accuracy (**MVA**) in evaluation phase for every dataset. The conducted experiments are highlighted and further discussed in the next section.

4.4 Ensemble

Finally, Ensemble voting on the resulting classifications was performed in order to select the classifiers with the best classification result.

V. EXPERIMENTS, RESULTS AND DISCUSSIONS

As is shown in sample results Table 1 and 2. A set of experiments involving six different machine learning techniques were performed for each of the 5 datasets. The used datasets of BCW, Colon-1, Colon-2, ICU, CPath produced 20, 1, 14, 5 and 5 reducts respectively with variable sizes.

Table 1: Sample of 2 **Training** resulting reducts for Cancer Dataset

		GNB	**LR**	**SVC**	**KN**	**RF**	**NN**				
RNo	**TYPE**	**TA**	**TA**	**TA**	**TA**	**TA**	**TA**	**ENS**	Average	**Min**	**Max**
0	ORG	96.27	97.33	65.85	97.77	100	97.8	0.96	92.5	65.85	100
0	COR	96.3	97.26	97.77	98.06	100	100	0.97	98.23	96.3	100
1	RED	95.72	96.45	97.04	97.26	100	100	0.96	97.75	95.72	100
1	CHI	100	99.93	97.77	97.88	100	100	1	99.26	97.77	100
1	CHIRED	100	100	97.47	97.66	100	100	1	99.19	97.47	100
1	CHICOR	100	99.93	98.1	98.35	100	100	1	99.4	98.1	100
1	REDCOR	100	99.93	98.1	98.39	100	100	1	99.4	98.1	100
1	ALL	100	99.93	98.1	98.35	100	100	1	99.4	98.1	100
2	RED	95.57	96.05	96.93	97.8	100	100	0.96	97.73	95.57	100
2	CHI	100	99.93	97.77	97.88	100	100	1	99.26	97.77	100
2	CHIRED	100	100	97.47	97.66	100	100	1	99.19	97.47	100
2	CHICOR	100	99.93	98.1	98.35	100	100	1	99.4	98.1	100
2	REDCOR	100	99.93	98.1	98.39	100	100	1	99.4	98.1	100
2	ALL	100	99.93	98.1	98.35	100	100	0.99	99.4	98.1	100

The following tasks were performed:

- Using Reducts: All of the dataset produced a number of reducts except for one dataset that had only 1 set.

- Using Chi^2: It was limited to those features with p-value of .05 or less. The choice of the cute value is a matter of experimentations.
- Using Correlation: an absolute value of correlation of 0.20 was adopted as cutoff value. Correlation results are normally score on a range of -1 to +1 with positive is good representation of an association and negative as inverse association. Use of absolute value will guaranty the inclusion of those values that represent an association relationship.
- The classification results were then feed to ensemble model.

As an example of results, Table 1 & 2 show the Training Accuracy (TA) obtained for the first two reducts of a data set. Last three columns show the average accuracy cross the classifiers, minimum TA values and max TA values for each subset of features. Since ORGSET and CORSET are done once for the dataset, they are highlighted for easy of compression.

Table 2: Sample of 2 **Valuation** Resulting Reducts for Cancer Dataset

		GNB	LR	SVC	KNN	RF	NN				
RNo	TYPE	VA	VA	VA	VA	VA	VA	ENS	Average	Min	Max
0	ORG	95.9	96.49	62.53	83.91	100	95.62	0.96	90.29	62.53	100
0	COR	95.9	96.64	96.35	96.64	100	94.15	97.7	95.9	100	100
1	RED	95.61	96.05	95.32	95.61	100	93.71	0.96	97.75	95.72	100
1	CHI	100	99.71	96.64	95.18	100	99.56	1	99.26	97.77	100
1	CHIRED	100	99.71	95.62	95.91	100	99.71	1	99.19	97.47	100
1	CHICOR	100	99.42	96.49	96.64	100	99.12	1	99.4	98.1	100
1	REDCOR	100	99.42	96.49	96.49	100	99.27	1	99.4	98.1	100
1	ALL	100	99.42	96.49	96.64	100	99.12	1	99.4	98.1	100
2	RED	95.47	95.32	95.62	95.18	100	94.59	0.96	97.73	95.57	100
2	CHI	100	99.71	96.64	95.18	100	99.56	1	99.26	97.77	100
2	CHIRED	100	99.71	95.62	95.91	100	99.71	1	99.19	97.47	100
2	CHICOR	100	99.42	96.49	96.64	100	99.12	1	99.4	98.1	100
2	REDCOR	100	99.42	96.49	96.34	100	98.68	1	99.4	98.1	100
2	ALL	100	99.42	96.49	96.64	100	98.98	0.99	99.4	98.1	100

An average training and validation accuracy is calculated for each subset and Reducts in nutshell, under different classifiers, the Reducts and the other subsets were able to match the accuracy of the original full set of features. In particular, the combined subsets that complemented the Reducts have improved the overall accuracy in very notable way. The best results came from the set of all subsets combined.

5.1 Experiment 1: Applying the Procedure to BCW Dataset

BCW was selected and applied as a pre-test to the application of the local datasets. In the BCW case, there were 19 reducts produced by rough set with sizes ranging from 4 to 5 features. Results obtained showed an improvement on the results wherever combined subsets are present.

5.2 Experiment 2: Applying the Procedure to Local Datasets

In this second experiment, we used four locally collected data sets to create small subset of features form the available data with the objective of being able to come up with a model that can predicted the labels for each dataset. The experiments done were:

- **Colon-1 Dataset**: The original table had only features one first and second are Age and Age Group so only the latter was used. The decision feature was Grade. As can be seen from the table, there was only a single reduct made of one attribute were as the original table contained 3 condition attributes of the 4 attributes that make up the condition attributes.

Table 3: Results of the different subsets of factors

		GNB		LR		SVC		KN		RF		NN		ENS
SZ	**TYPE**	**TA**	**VA**	**TA**	**VA**	**TA**	**VA**	**TA**	**VA**	**TA**	**VA**	**TA**	**VA**	
3	ORG	58.39	50.59	59.07	56.7	60.72	56.7	50.28	42.24	60.72	60.72	60.72	56.16	**0.54**
1	COR	55.49	55.45	55.49	55.45	56.04	55.45	54.39	53.83	56.04	56.04	56.04	55.45	**0.56**
1	RED	55.49	55.45	55.49	55.45	56.04	55.45	54.39	53.83	56.04	56.04	56.04	55.45	**0.56**
1	CHI	52.2	45.14	52.2	44.02	52.2	44.02	51.79	48.96	52.2	52.2	52.2	44.02	**0.40**
2	CHIRED	100	100	100	100	100	100	100	100	100	100	100	100	**1**
2	CHICOR	100	100	100	100	100	100	100	100	100	100	100	100	**1**
1	REDCOR	55.49	55.45	55.49	55.45	56.04	55.45	54.39	53.83	56.04	56.04	56.04	55.45	**0.56**
2	ALL	100	100	100	100	100	100	100	100	100	100	100	100	**1**

The reduct size was 33% less, but Ensemble was albeit higher than that of the original. The combined sets of CHIRED, CHICOR, and ALL had a perfect 100% with 33% less in size;

- **Colon-2 Dataset**: This is the same features as Colon-1 but with a different decision attribute, namely Survival Period. The original table had only 14 attributes inclusive of decision attribute, survival period. There were 16 reducts with size ranging from 1 to 6 attributes. Combined attribute sets reached 8 attributes in size, but still much less than original size. Results were in the high 90s for the combined and marginally higher than original in most cases from the reducts. Only 3 of the 16 reducts were less accurate than the original.

- **ICU Dataset**: This is another data set made of 237 objects and 8 features, The decision attribute was Survival. There were 5 reducts with size of 4. Combined attribute sets reached 6 attributes in size, but still much less than original size. Results were consistently in the high 90s for the combined and marginally higher than original in most cases from the reducts.

- **Clinic-o-Pathology**: This is another dataset made of 153 objects and 17 features and a Grade decision attribute. The original table had only 17 attributes inclusive of decision attribute. There were 49 reducts with size ranging from 1 to 4 attributes. Combined attribute sets reached 8 attributes in size, but still much less than original size. The ensample values for ORIG, COR were 0.59 and 0.65 respectively. Results were in the high 90s for the combined and higher than original in most cases from the reducts.

VI. CONCLUSION AND RECOMMENDATION

The ability to reduce a large feature set through the selection of most appropriate ones is of a vital importance for the analysis and classification applications. It can positively influence efficiency of algorithms by the elimination of the redundant and noisy data. Many applications fields in medical applications and particularly cancer evaluation and prognoses require estimation of their grading, staging and survival prediction based on a set of data selected attributes. Making a judgment on medical cases or any case for that matter is a critical role and a very demanding work that needs special care and qualified human expertise.

Cancer grading is the task of describing how atypical and aggressive the cancer cells and tissue look under a microscope when compared to healthy cells. Lower grade cancers are characteristically less aggressive and have a better prognosis; Staging is the process where a medical expert decides on a person's cancer existence based on results of diagnostic tests, scans, and samples taken during surgery. Staging is very critical in deciding treatments and future prognoses; Finally cancer survival indicates the portion of people who survive a certain type of cancer for a specific amount of time;

Statistical and machine learning techniques are used to help the experts in deciding and in selecting important factors. A number of experiments combining techniques from statistics, rough sets theory and machine learning classifiers were applied to help select and evaluate important features from a set of factors in a sample dataset collected from local hospitals in Libya. Rough sets techniques were used in generating equivalent subsets of features from among the available factors. The selected subsets of attributes are then compared and used as a bases that can be augmented by a number of other feature selection techniques, namely that of Chi^2 and attribute correlation. The resulting subsets are used as a bases for classification models to help in deciding new cases. Then an ensemble is used to select the best results.

The suggested approach showed a very encouraging result. An improvement is obtained through the augmentation of rough sets with Chi^2 and correlation when applied to some data available from the web and other locally collected data. The used approach was quite successful with variable training and evaluation accuracies. The experiments conducted have shown that rough set is capable of suggesting a reasonable initial subset (Reducts) that serve as bases for more improved subsets using the other complementary techniques of chi^2 and correlation. The set of alternative factors when used for classification have shown very good accuracies. Machine learning classifier algorithms have been applied to selected and prepared data using the mentioned feature selection techniques. The set of classifiers included Nearest Neighbour, Support Vector Machines, Random Frost, Neural Nets Logistic Regression, and Bayes with variable degrees of accuracy. The adopted classifiers are further subjected to ensemble with very encouraging results.

REFERENCES

[1]. Schober, P., & Vetter, T. R. (2018). Survival analysis and interpretation of time-to-event data: the tortoise and the hare. Anesthesia and analgesia, 127(3), 792.

[2]. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). Moving beyond linearity. In an Introduction to Statistical Learning (pp. 265-301). Springer, New York, NY.

[3]. Sammut, C., & Webb, G. I. (2017). Encyclopaedia of machine learning and data mining. Springer Publishing Company, Incorporated.

[4]. Kramer, M. A. (1991). Nonlinear principal component analysis using auto-associative neural networks. AIChE journal, 37(2), 233-243.

[5]. Kratsios, A., & Hyndman, C. (2021). Neu: A meta-algorithm for universal uap-invariant feature representation. Journal of Machine Learning Research, 22, 92.

[6]. The 1-Applied Predictive Modelling 1st ed.(2013), Corr. 2nd printing 2018 Edition by Max Kuhn (Author), Kjell Johnson (Author)

[7]. Mirtaheri, S. L., & Shahbazian, R. (2022). Machine Learning: Theory to Applications. CRC Press.

[8]. Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. Knowledge and information systems, 34(3), 483-519.

[9]. Hira ZM, Gillies DF. (2015) A review of feature selection and feature extraction methods applied on microarray data. Advances in bioinformatics. 2015;.

[10]. Yang, Y., & Pedersen, J. O. (1997). A comparative study on feature selection in text categorization. In Icml (Vol. 97, No. 412-420, p. 35).

[11]. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of machine learning research, 3(Mar), 1157-1182.

[12]. Prion S, Haerling KA. (2014) Making sense of methods and measurement: Pearson product-moment correlation coefficient. Clinical simulation in nursing. 2014 Nov 1;10(11):587-8.

[13]. Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Moore JH. (2018) Relief-based feature selection: Introduction and review. Journal of biomedical informatics. Sep 1;85:189-203.

[14]. Forman G. (2003) An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. Mar 3;3(Mar):1289-305.

[15]. Chen, X. W., & Jeong, J. C. (2007, December). Enhanced recursive feature elimination. In Sixth International Conference on Machine Learning and Applications (ICMLA 2007) (pp. 429-435). IEEE.

[16]. Bach, Francis R (2008). Bolasso: model consistent lasso estimation through the bootstrap. Proceedings of the 25th International Conference on Machine Learning. pp. 33–40. doi:10.1145/1390156.1390161. ISBN 9781605582054. S2CID 609778.

[17]. Zare, H., Haffari, G., Gupta, A., & Brinkman, R. R. (2013). Scoring relevancy of features based on combinatorial analysis of lasso with application to lymphoma diagnosis. In BMC genomics (Vol. 14, No. 1, pp. 1-9). BioMed Central.

[18]. Zhang, Q., Xie, Q., & Wang, G. (2016). A survey on rough set theory and its applications. CAAI Transactions on Intelligence Technology, 1(4), 323-333.

[19]. Franke TM, Ho T, Christie CA. The chi-square test: Often used and more often misinterpreted. American journal of evaluation. 2012 Sep;33(3):448-58.

[20]. Bewick, V., Cheek, L., & Ball, J. (2003). Statistics review 7: Correlation and regression. Critical care, 7(6), 1-9.

[21]. Wang QQ, Yu SC, Qi X, Hu YH, Zheng WJ, Shi JX, Yao HY. (2019) Overview of logistic regression model analysis and application. Zhonghua yu fang yi xue za zhi [Chinese journal of preventive medicine]. Sep 1;53(9):955-60.

[22]. Kaur G, Oberai EN. (2014) A review article on Naive Bayes classifier with various smoothing techniques. International Journal of Computer Science and Mobile Computing. Oct;3(10):864-8

[23]. Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H. (2018) State-of-the-art in artificial neural network applications: A survey. Heliyon. Nov 1;4(11): e00938.

[24]. Taunk K, De S, Verma S, Swetapadma A. (2019) A brief review of nearest neighbour algorithm for learning and classification. In2019 International Conference on Intelligent Computing and Control Systems (ICCS) May 15 (pp. 1255-1260). IEEE.

[25]. Biau G, Scornet E. (2016) A random forest guided tour. Test. Jun;25(2):197-227.

[26]. Brereton RG, Lloyd GR. Support vector machines for classification and regression. Analyst. 2010;135(2):230-67.

[27]. Mussa A, Rabia A, Wesam E, Jamela BM, Firas A, Fathi AB. (2015) CLINICOPATHOLOGICAL CHARACTERISTICS OF COLON CANCER IN LIBYA. Misurata Medical Sciences Journal.:107. MMSJ Vol.2 Issue.2 (Winter).f[

[28]. Street WN, Wolberg WH, Mangasarian OL. (1993) Nuclear feature extraction for breast tumor diagnosis. InBiomedical image processing and biomedical visualization Jul 29 (Vol. 1905, pp. 861-870). SPIE

[29]. Øhrn A, Komorowski J. (1997) Rosetta--a rough set toolkit for analysis of data. InProc. Third International Joint Conference on Information Sciences 1997.

[30]. Zhi, J., Liu, J. Y., & Wang, Z. (2009). Rough set attribute reduction algorithm based on immune genetic algorithm. In 2009 2nd IEEE International Conference on Computer Science and Information Technology (pp. 421-424). IEEE.