KAUNAS UNIVERSITY OF TECHNOLOGY

ELECTRICAL AND ELECTRONICS FACULTY

PRASANTH RANGARAJAN JAYASANKAR

Comparative study of ensemble methods for multi-class data classification

Final project for Master degree

Supervisor
Prof. Dr. A.Lipnickas

Reviewer
Prof. Dr. R.Simutis

KAUNAS, 2016

KAUNAS UNIVERSITY OF TECHNOLOGY
FACULTY OF ELECTRICAL AND ELECTRONICS ENGINEERING

Comparative study of ensemble methods for multi-class data classification

Control Technologies (621H66001)

Final project for Master degree

Supervisor
Prof. Dr. A.Lipnickas

Project made by
Prasanth Rangarajan Jayasankar

Reviewer
Prof. Dr. R.Simutis

KAUNAS, 2016

KAUNAS UNIVERSITY OF TECHNOLOGY

FACULTY OF ELECTRICAL AND ELECTRONICS ENGINEERING

PRASANTH RANGARAJAN JAYASANKAR

FINAL DEGREE PROJECT T000M041

"**Comparative study of ensemble methods for multi-class data classification**"

DECLARATION OF ACADEMIC HONESTY

10 June 2016
Kaunas

I confirm that a final project by me, on the subject " **Comparative study of ensemble methods for multi-class data classification** " is written completely by myself; all provided data and research results are correct and obtained honestly. None of the parts of this thesis have been plagiarized from any printed or Internet sources, all direct and indirect quotations from other resources are indicated in literature references. No monetary amounts not provided for by law have been paid to anyone for this thesis.

I understand that in case of a resurfaced fact of dishonesty penalties will be applied to me according to the procedure effective at Kaunas University of Technology.

(PRASANTH RANGARAJAN JAYASANKAR) *(signature)*

SUMMARY

Author – Prasanth Rangarajan Jayasankar

Title - Comparative study of ensemble methods for multi-class data classification

Language- English

This work contains: 50 Pages

Keywords: Ensemble methods, Classifier evaluation, OVO-AVA, Single class, Multiclass, KNN, SVM, Bayes.

The main Goal is to Estimate the performance (error rate) of a classifier. The lower the error, better the classifier. In this paper some sets of data's selected are set for evaluation. The data's are trained and validated by a couple of classifiers with different classifier methods, then the results are compared by performing ensemble combining approach were investigated and compared with the well-known one-vs.-one and one-vs.-all decomposition strategies for multiclass data classification were the best classifiers with smallest misclassification rate are found and also the combination method helps to improve the single classifiers result.

The main motive of this research work was to discover methods for building a generalized ensemble of classifiers. As the performance on an empirical comparison of several multi classifier systems using several data sets those with different problems. Our experimental results shows that our ensemble methods on classifier will show the outperform of best state-of-art standalone ensemble methods.

Table of Contents

Comparative study of ensemble methods for multi-class data classification

Prasanth Rangarajan Jayasankar,

Department of Automation Engineering, Kaunas University of technology, Lithuania
rj_a380@outlook.com,

1.Abstract

Extensive study of different methods for building ensemble of classifiers are analyzed on this paper. Were various ensemble methods that are based on performing feature are examined and illustrates the power of using this variant by applying them to the number of problems of data's. So from the analyzed extensive study the best performing ensemble are found. The inputs are made by combining an approach are done on randomly with a cluster based input on ensemble method. Compared to other state of art standalone classifiers and ensemble, this method consistency performed well across other twelve diverse benchmark datasets. Another useful finding in this approach does not require parameters to be carefully tuned for each dataset, making our ensemble method well suited for practioners since there is less risk of over-training. Another interesting finding is that random subspace can be occupied with several other ensemble methods to improve performance. Then with these ensemble methods of classifiers the set of data's are compared with classifiers. And also to investigate and compare the well-known one-vs.-one and one-vs.-all decomposition strategies for multiclass data classification, Here ensemble method is the main base with the binary inputs were the special attention to the final step of the ensembles are paid, i.e. the combination of the outputs of the binary classifiers. The majority voting technique are compared by performance of the ensembles. The experimental study is carried out with several well-known data classification algorithms such as Known Nearest Neighbor (KNN), Support Vector Machines (SVM), Bayes, … on very well-known multi class data such as Fisher Iris, Breast Cancer, Wine, Diabetics, Forest Fire, ... So the best classifiers with smallest misclassification rate are found and also the combination method helps to improve the single classifiers result.

KEYWORDS: Ensemble methods, Classifier evaluation, OVO-AVA, Single class, Multiclass KNN, SVM, Bayes.

1.1 Introduction

Until recently, it was difficult to collect measurements, also time consuming and expensive. More information that can be processed With cheaper and more powerful forms of computing and data storage. To truly assist practitioners in other fields, researchers in machine intelligence need to develop general purpose classification methods that are capable of handling a broad Varity of problems and data types. These classification methods also need to be easy to use and they need to compete with less flexible state-of-art methods that have been crafted for very specific problems.

It is necessary to improve flexibility and accuracy to build systems when multiple classifiers are combined. The main idea behind multi classifier system is to average the hypotheses of a diverse group, or ensemble, of classifiers to produce a better approximation to a true hypothesis. The main aim of this paper is to compare several approaches of classifiers by building ensembles methods and to find a best method that works well across a serious of datasets without careful parameters tuning for each dataset. We find that this method compares very well with several state-of-art standalone and ensemble methods.

Another interesting finding is that random subspace can be occupied with several other ensemble methods to improve performance. Then with these ensemble methods of classifiers the set of data's are compared with classifiers. And also to investigate and compare the well-known one-vs.-one and one-vs.-all decomposition strategies for multiclass data classification. Here ensemble method is the main base with the binary inputs were the special attention to the final step of the ensembles are paid, i.e. the combination of the outputs of the binary classifiers. The majority voting technique are compared by performance of the ensembles. The experimental study is carried out with several well-known data classification algorithms such as Known Nearest Neighbor (KNN), Support Vector Machines (SVM), Bayes, … on very well-known multi class data such as Fisher Iris, Breast Cancer, Wine, Diabetics, Forest Fire, ... So the best classifiers with smallest misclassification rate are obtained as the result and also the combination method helps to improve the single classifiers result. More Datasets is processed in this method to gain more accuracy.

1.2 Methodology of the Work

Data collecting measurements was difficult, time consuming and expensive. With increasingly cheaper and more powerful forms of computing and data storage.

To investigate and compare the well-known one-vs.-one and one-vs.-all decomposition strategies for multiclass data classification, Here ensemble method is the main base with the binary inputs were the special attention to the final step of the ensembles are paid, i.e. the combination of the outputs of the binary classifiers.

The majority voting technique are compared by performance of the ensembles. The experimental study is carried out with several well-known data classification algorithms such as Known Nearest Neighbor (KNN), Support Vector Machines (SVM), Bayes, … on very well-known multi class data such as Fisher Iris, Breast Cancer, Wine, Diabetics, Forest Fire, ... We use the following methods and experimental study to reduce the misclassification rate.

Used methods:

- Single classifiers methods
- Ensemble methods
- Decomposition strategies in multi-classification (OvO and AvA)
- Majority Voting Technique

Experimental study Methods:

- Single classifiers performance for multiclass data classification
- Ensemble classification methods for multiclass data classification
- Decomposition methods for multiclass data classification

A First Idea

Multiclass classifying is the problem of classifying instances into more than two classes. The multiclass classification problem can be solved by naturally extending the binary classification technique for some algorithms. These include neural networks, decision trees, k-Nearest Neighbor.[2]

2.Multi-class data classification

Each training point belongs to one of N different classes. The goal is to construct a function which, given a new data point, will correctly predict the class to which the new point belongs. Suppose we knew the density, pi(x), for each of the N classes. Then, we would predict the class.

Of course we don't know the densities, but we could estimate those using classical techniques. The Problem with Densities and Motivation Estimating densities is hard, especially in high dimensions with limited data. For binary classification tasks, we have seen that directly estimating a smooth separating function gives better results than density estimation.

Supervised classification algorithms aim at producing a learning model from a labeled training set. Various successful techniques have been proposed to solve the problem in the binary classification case. The multiclass classification case is more delicate, as many of the algorithms were introduced basically to solve binary classification problems. In this short survey we investigate the various techniques for solving the multiclass classification problem.[1]

Binary classification is a well studied special case of the classification problem. Statistical properties of binary classifiers, such as consistency, have been investigated in a variety of settings. Binary classification methods can be generalized in many ways to handle multiple classes. It turns out that one can lose consistency in generalizing a binary classification method to deal with multiple classes. We study a rich family of multiclass methods and provide a necessary and sufficient condition for their consistency. We illustrate our approach by applying it to some multiclass methods proposed in the literature.

2.1 TYPES OF CLASSIFIERS

I. **Kn-Nearest neighbor –** A robust non-parametric classifier. Classification has high computational complexity when. Must select metric and value of k. k must be set using validation. Can have excellent performance for arbitrary class conditional pdfs. [4]

II. **Parzen window –** Robust non-parametric. Must select form of kernel and size parameter h. Complexity and performance is similar to k-NN method.

III. **Neural network –** The multi-layer perception (a non-parametric classifier) is the standard network to use for supervised learning. Other types of neural networks are useful for unsupervised learning. Training can be very slow, but classification is fast. The number of hidden nodes must be set using validation (see below). Can have excellent performance. Impossible for a human to “understand” the classifier. Performance is vulnerable to unforeseen input data Gives a set of rules that can be understood.

IV. **Support Vector Machine:** Advantages are High accuracy, nice theoretical guarantees regarding overfitting, and with an appropriate kernel they can work well even if you’re data isn’t linearly separable in the base feature space. Especially popular in text classification problems where very high-dimensional spaces are the norm. Memory-intensive, hard to interpret, and kind of annoying to run and tune, though, so I think random forests are starting to steal the crown.[5]

V. **Bayes –** Always the optimal (minimum error rate or minimum risk) but requires exact knowledge of class prior probabilities and class conditional probabilities of features. Seldom possible because exact knowledge rarely exists. [6]

VI. **Bayes linear –** Assumes Gaussian distribution of features with equal covariance matrices for each class. A modest number of parameters to estimate. Fast training and classifying. In general, performance is limited. [6]

VII. **Bayes quadratic –** Assumes Gaussian distribution of features with a separate covariance matrix for each class. Requires many parameters (feature covariance's) to be estimated. Fast training and classifying. Performance may be poor when data is significantly non-Gaussian. Nearest neighbor (1-nearest neighbor) – A simple nonparametric method that uses all the training data for classification. Has high computational complexity for classification, though some acceleration methods exist. Must select a metric. Upper bound on error rate approaches twice that of ideal Bayes classifier.[6]

2.2 Single Classifier Method

This type of systems is to achieve the best possible classification performance. In general, a number of classifiers could be tested and evaluated in these systems, and the most appropriate one chosen for the problem at hand. However, it has recently become common practice to use more than one classifier rather than a single one for pattern recognition tasks. This is because different classifiers usually make different errors on different samples, which means that by combining classifiers we can create an ensemble that makes more accurate decisions.

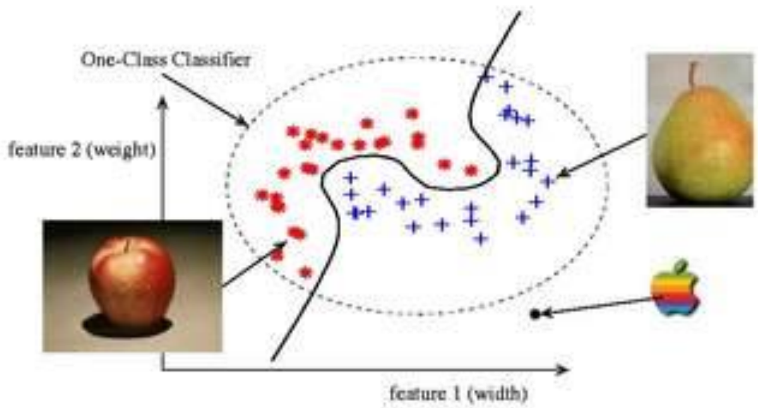

Fig.1 Single Class Classifier separation

Multiclass classification is one of the fundamental tasks in bioinformatics and typically arises in cancer diagnosis studies by gene expression profiling. There have been many studies of aggregating binary classifiers to construct a multiclass classifier based on one-versus-the-rest or other coding strategies, as well as some comparison studies between them. Correlation and single variable classifier methods are very simple algorithms to select a subset of variables in a dimension reduction problem, which utilize some measures to detect relevancy of a single variable to the target classes without considering the predictor properties to be used[7]. We apply this method to various classification problems including a synthesized data set and some cancer diagnosis data sets from gene expression profiling. The results demonstrate that, in most situations, our method can improve classification accuracy over simple voting heuristics and is better than or comparable to state-of-the-art multiclass predictors.

2.3 Ensemble Methods

Ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms Unlike a statistical ensemble in statistical mechanics, which is usually infinite, a machine learning ensemble refers only to a concrete finite set of alternative models, but typically allows for much more flexible structure to exist between those alternatives.

Goals are to Estimate the performance (error rate) of a classifier. The lower the error, the better. Often used to compare two or more types of classifiers. Compare the performance of two classifiers.

Classifiers are trained using real data, not simulated data. There are a limited number of samples to work with (for both training and testing).

The main strategies are,

- Bagging – use different samples of observation (Bootstrap Aggregation)
- Boosting – Make examples currently misclassified more Important.

To improve weak classifiers bagging and boosting could be used. These techniques are based on combining classifiers. Usually, a simple majority vote or a weighted majority vote are used as combining rules in bagging and boosting. However, other combining rules such as mean, product and average are possible[8].

Training, validation, and testing: Some classifier methods need certain “super”-parameters, In this case one should divide the training data into two groups, one for actually training the classifier and one for validation purposes[4]. The results of the validation tests are used for selecting the needed super-parameters; this is still part of the overall training process. The test data should be separate and not used for selecting the super-parameters. It should be use only after all the parameters have been decided. After all super-parameters of classifier are selected, then a classifier is designed using the all available training data[7]. The resulting classifier can be tested on the test data to give an unbiased estimate of its error rate.

3.Methods For Classifier Evaluation

3.1 Resubstitution:

First uses all available data to design a classifier. Then uses the same data again to test the classifier. Produces an “optimal” classifier in the sense that it uses all available data for design. Easy and fast. Using the same data for training and testing gives an optimistically biased estimate of the error rate, but variance is relatively smaller since all data is used for training. It is useful for determining a lower bound on classifier error rate. Since it suffers from “testing on the training data,” resubstitution is not recommended unless the training set is very large.

3.2 Data Partition (Holdout):

First separate data into two groups for training and testing. Often the training set will be selected to be twice the size of the test set. The classifier is sub-optimal in the sense that it uses only part of available data for training. Testing also suffers from a small test sample. Easy and fast. Result estimate of error rate is unbiased, but has a large variance (i.e., uncertainty).

3.3 Cross-validation :

A generalization of the holdout method. N total samples are divided into m groups of equal size. m different classifiers are trained each using m –1 groups, holding out each of the groups. For each of the m classifiers, the group left out is tested. The m test results are averaged. All samples get used for both training and testing. The result is unbiased and with minimum variance. Good method when a large number of samples are available. 5x2 cross-validation: Randomly divide data set into two parts equal sized parts: D1 and D2. First train on D1 and validate (test) on D2, then reverse the roles. Repeat this process 4 more times for a total of 10 train and test runs[5]. Five random divisions is chosen as a compromise between getting a large enough sample of results (the more, the better) and diminishing returns (since there is much redundancy in the train/validate sets).

A good method to use for selecting the appropriate classifier type to use and for determining certain classifier[9] “super”-parameters, e.g., k for nearest neighbor, h for Parzen, number of hidden nodes in neural net.

3.4 Jackknife (Leave-one-out) :

A limiting case of cross-validation. Where m = N. N different classifiers are trained each using N–1 samples. For each of the N classifiers, the one left out sample is tested. The N test results are averaged. Classifiers are very close to optimal. All samples are used for testing. Result is unbiased and with minimum variance[9]. If a fast leave-one-out algorithm is available (to estimate necessary parameters using an update scheme: e.g., mean, covariance and its inverse and determinant), jackknife is fairly fast. But, if no fast algorithm is available, then it is very slow, such as neural networks. Fast algorithms exist for estimating mean, covariance matrix, as well as inverse and determinant of covariance matrix.

So, useful for: 3 Bayes quadratic, k-nearest neighbor (using Euclidean or Mahalanobis distance), Parzen with Gaussian window.[6]. The best method for estimating performance

3.5 Bootstrap:

Bootstrap Various forms exist. A common one is: A training set is generated by randomly selecting N samples using replacement (i.e., samples can be selected more than once). The samples not selected for training are used for testing. The process is repeated many times, e.g., 200. The results are averaged to give final estimate of the error rate. The classifiers produced are never optimal. The resulting estimate is unbiased. This method is very computationally intensive (slow). The method very good for use when only a few samples.

3.6 Data Bases Used for evaluation

3.6.1 Wine Data Set: [12]

Data Set Information:

These data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines. The units are not important for the purposes of this problem as I recommend using the scaled data in your analysis anyway. [16]

Attribute Information:

All attributes are continuous

NOTE: attribute information [17]

3 classes Data size: 178 entries

Data distribution: 59, 71, and 48 entries for each class

12 features corresponding to the values from chemical analysis, no missing data:

- Alcohol
- Malic acid
- Ash
- Alkalinity of ash
- Magnesium
- Total phenols
- Flavonoids
- Nonflavanoid phenols
- Proanthocyanins
- Color intensity
- Hue
- Proline

3.6.2 Diabetes Data Set:

Data Set Information:[13]

The dataset represents 10 years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. It includes over 50 features representing patient and hospital outcomes. Information was extracted from the database for encounters that satisfied the following criteria. [18]

(1) It is an inpatient encounter (a hospital admission).
(2) It is a diabetic encounter, that is, one during which any kind of diabetes was entered to the system as a diagnosis.
(3) The length of stay was at least 1 day and at most 14 days.
(4) Laboratory tests were performed during the encounter.
(5) Medications were administered during the encounter.

Attribute Information:

The data contains such attributes as patient number, race, gender, age, admission type, and time in hospital, medical specialty of admitting physician, and number of lab test performed, HbA1c test result, diagnosis, number of medication, diabetic medications, and number of outpatient, inpatient, and emergency visits in the year before the hospitalization, etc. [19]

3.6.3 Breast Cancer Data Set:

Data Set Information:[14]

Samples arrive periodically. The database therefore reflects this chronological grouping of the data. This grouping information appears immediately below, having been removed from the data itself: [20]

Attribute Information:

- Sample code number: id number
- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape
- Marginal Adhesion
- Single Epithelial Cell Size
- Bland Chromatin
- Normal Nucleoli
- Mitoses

Class: (2 for benign, 4 for malignant)

3.6.4 Iris Data Set:

Data Set Information:[15]

This is perhaps the best known database to be found in the pattern recognition literature. Fisher's paper is a classic in the field and is referenced frequently to this day. The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are NOT linearly separable from each other. Predicted attribute: class of iris plant. This is an exceedingly simple domain. This data differs from the data presented in Fishers article.

Attribute Information:

- sepal length in cm
- sepal width in cm
- petal length in cm
- petal width in cm
- class: 3
  -- Iris Setosa
  -- Iris Versicolour
  -- Iris Virginica

3.6.5 Forest Fire Dataset

Data Set Information:[16]

Several Data Mining methods were applied. After fitting the models, the outputs were post-processed. Four different input setups were used. The experiments were conducted using a 10-fold (cross-validation) x 30 runs. Two regression metrics were measured: MAD and RMSE. A Gaussian support vector machine (SVM) fed with only 4 direct weather conditions (temp, RH, wind and rain) obtained the best MAD value: 12.71 +- 0.01 (mean and confidence interval within 95% using a t-student distribution). The best RMSE was attained by the naive mean predictor. An analysis to the regression error curve (REC) shows that the SVM model predicts more examples within a lower admitted error. In effect, the SVM model predicts better small fires, which are the majorities

4. Decomposition strategies in multi-classification

Several motivations for the use of binary decomposition strategies in multi-class classification problems can be found in the literature. For example, in the reduction of the complexity involved in the classes' separation when using a decomposition approach was shown. Noise is a common problem that produces negative consequences in classification problems. When a problem has more than two classes, that is, a multi-class problem, an interesting approach to deal with noise is to decompose the problem into several binary subproblems, reducing the complexity and consequently dividing the effects caused by noise into each of these subproblems. This contribution analyzes the use of decomposition strategies, and more specifically the One-vs-One scheme, to deal with multi-class datasets with class noise.[26]. This way, the binary predictors generated may impose preferences for some of the classes. Decomposition also opens up new possibilities for the use of parallel processing, since the binary sub problems are independent and can be solved with different processors. Dividing a problem into several new problems which are then independently solved implies the need for a second phase where the outputs of each problem have to be aggregated. Therefore, decomposition includes two steps:

- Problem division
- Combination of the outputs.

4.1 A Simple Idea — One-vs.-One (OVO)Classification

The simplest approach is to reduce the problem of classifying among K classes into K binary problems, where each problem discriminates a given class from the other K −1 classes [18]. For this approach, we require N = K binary classifiers, where the k th classifier is trained with positive examples belonging to class k and negative examples belonging to the other K − 1 classes. When testing an unknown example, the classifier producing the maximum output is considered the winner, and this class label is assigned to that example. state that this approach, although simple, provides performance that is comparable to other more complicated approaches when the binary classifier is tuned well.
Pick a good technique for building binary classifiers. Build N different binary classifiers.

4.2 Another Simple Idea — All-vs.-All (AVA) Classification

In this approach, each class is compared to each other class. A binary classifier is built to discriminate between each pair of classes, while discarding the rest of the classes. This requires building $\frac{k(k-1)}{2}$ binary classifiers. When testing a new example, a voting is performed among the classifiers and the class with the maximum number of votes wins. Results show that this approach is in general better than the one-versus-all approach.

Also called all-pairs or one-vs.-one classification.

4.3 OVO vs. AVA

OVO and AVA are so simple that many people invented them independently. It’s hard to write papers about them. So there’s a whole cottage industry in fancy, sophisticated methods for multiclass classification. To the best of my knowledge, choosing properly tuned regularization classifiers (RLSC, SVM) as your underlying binary classifiers and using one-vs.-one (OVO) or all-vs.-all (AVA) works as well as anything else you can do. We propose a feature selection method for multiclass classification. The proposed method selects features in backward elimination and computes feature ranking scores at each step from analysis of weight vectors of multiple two-class linear Support Vector Machine classifiers from one-versus-one or one-versus-all decomposition of a multi-class classification problem[27].

Viewed naively, AVA seems faster and more memory efficient. It requires classifiers instead, but each classifier is (on average) much smaller. If the time to build a classifier is super linear in the number of data points, AVA is a better choice. With SVMs, AVA’s probably best. However, if you can solve one RLS problem over your entire data set using a matrix factorization, so with Regularized least square RLS, OVA are a great choice.

4.4 Majority Voting Technique:

The idea behind the voting classifier implementation is to combine conceptually different machine learning classifiers and use a majority vote or the average predicted probabilities (soft vote) to predict the class labels. An approach based on combining classifiers has shown a significant potential gain in comparison to the performance of an individual best classifier. This improvement turned out to be subject to a sufficient level of diversity exhibited among classifiers, which in general can be assumed as a selective property of classifier subsets. Given a large number of classifiers, an intelligent classifier selection process becomes a crucial issue of multiple classifier system design[28].

Program Output Results

Algorithm 1. Known Nearest Neighbor (KNN)

Dataset 1.KNN Breast Cancer:

Output Results:

E26

	A	B	C	D	E	F	G	H	I	J	K	L
1	**KNN(BreastCan**		**One Vsone**		**ALL VS ALL**		**Maximum Voting**			**Mis classification**		
2												
3	Radius		100%		25.80%		174(100%)			0(0%)		
4	Texture		87.36%				152(87.36%)			22(12.64%)		
5	Perimeter		10.34%				18(10.34%)			156(89.66%)		
6	Area		83.33%				145(83.33%)			29(16.67%)		
7	Smoothness		5.75%				10(5.75%)			164(94.25%)		
8	Compactness		12.64%				22(12.64%)			152(87.36%)		
9	Concavity		87.36%				152(87.36%)			22(12.64%)		
10	ConcavePoints		74.71%				130(74.71%)			44(25.29%)		
11	Symmetry		2.30%				4(2.30%)			170(97.70%)		
12												
13												
14												

Tab.1 Output on Tabular Representation

Specific representation of output that are compared with data's within its attributes on classes which is shown on One Vs. One for Breast Cancer Dataset with KNN classifier;

On All Vs. All, all the data's from all the attributes are taken in account were the average of the values are taken which gives the accuracy of all the attributes.

The maximum voting represents the data's compared in which the data's that are corresponding to the class on its attributes to its maximum data's on its sets.

The misclassification represents the data's compared in which the data's that are not corresponding to the class on its attributes.

	Radius	Textur	Perim	Area	Smoot	Comp	Conca	Conca	Symm
Radius	100.00% (174)	0	0	0	0	0	0	0	0
Texture	0	100.00% (174)	0	0	0	0	0	0	0
Perimeter	0	82.18% (143)	17.82% (31)	0	0	0	0	0	0
Area	0	83.33% (145)	14.37% (25)	2.30% (4)	0	0	0	0	0
Smoothness	0	85.06% (148)	12.64% (22)	0	2.30% (4)	0	0	0	0
Compactness	0	94.25% (164)	1.15% (2)	1.15% (2)	0	3.45% (6)	0	0	0
Concavity	0	75.86% (132)	20.11% (35)	1.15% (2)	1.15% (2)	0	1.72% (3)	0	0
ConcavePoints	0	77.01% (134)	8.05% (14)	1.72% (3)	1.72% (3)	6.90% (12)	0.57% (1)	4.02% (7)	0
Symmetry	0	82.76% (144)	15.52% (27)	0.57% (1)	0	0	0	0.57% (1)	0.57% (1)

Fig.2 ALL Vs. ALL Representation

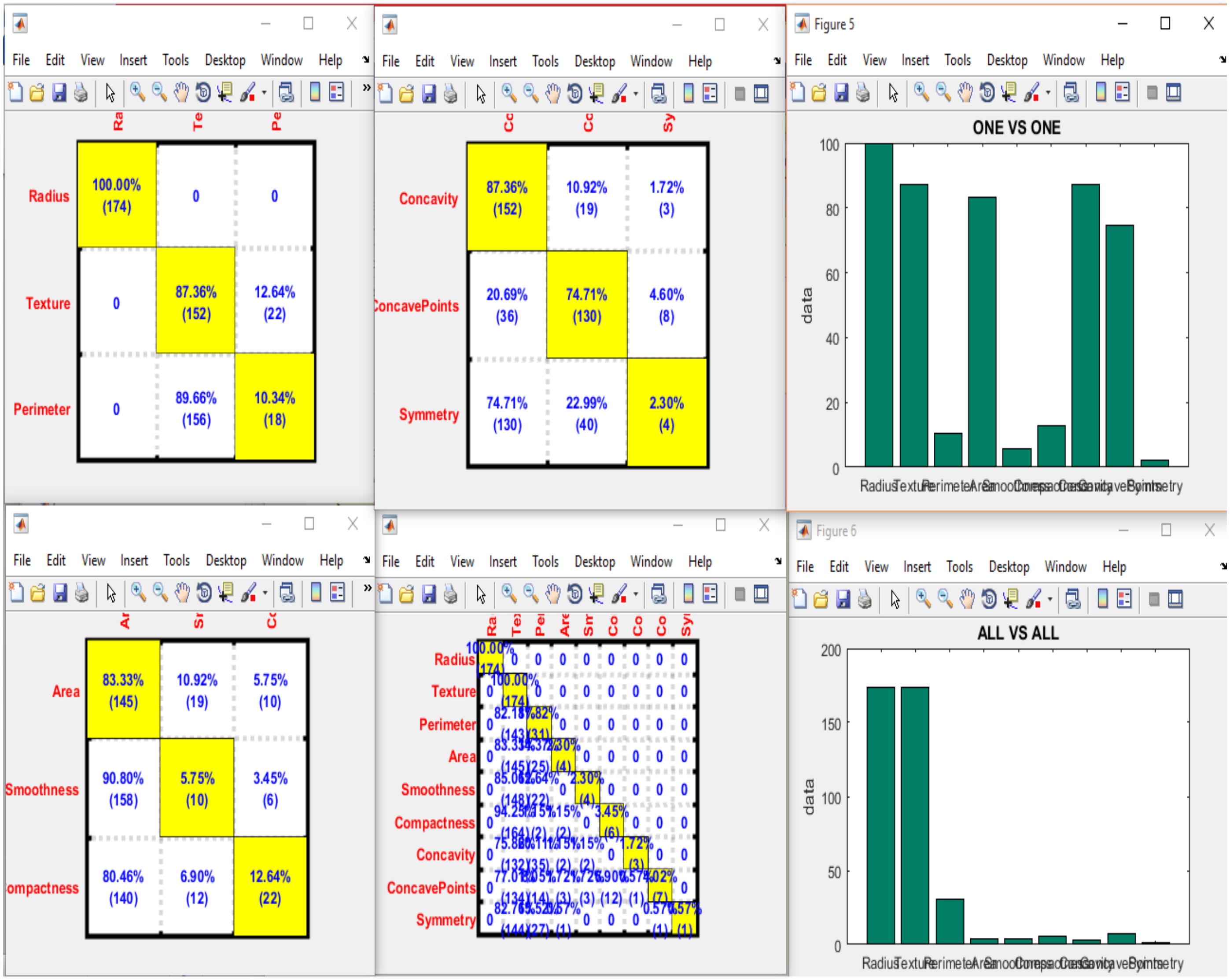

Fig.3 Overall Breast Cancer KNN Evaluation

The Outputs shows the Overall evaluation of the Breast Cancer Data set using KNN type of Classifier.

Thus give the accuracy of

OVO: 51.5288%

AVA: 25.80%

Majority: 100%

Dataset 2. KNN Diabetics:

Output Results:

A25

	A	B	C	D	E	F	G	H	I	J	K	L
1	KNN(Diabetics)		One Vsone		ALL VS ALL		Maximum Voting			Mis classification		
2												
3	EncounterID		100%		43.35%		2000(100%)			0(0%)		
4	Patientbr		99.95%				1999(99.95%)			1(0.05%)		
5	Admission typeid		100%				2000(100%)			0(0%)		
6	Discharge positionid		70.85				1417(70.85%)			583(29.15%)		
7	admissionsourceid		53.75%				1075(53.75%)			925(46.25%)		
8	Timein hospital		30.10%				602(30.10%)			1398(69.90%)		
9	numlab procedure		96.75%				1935(96.75%)			65(3.25%)		
10	numprocedure		96.60%				1932(96.60%)			68(3.40%)		
11	nummedications		85.70%				1714(85.70%)			286(14.30%)		
12	NumoutPatient		99.10%				1982(99.10%)			18(0.90%)		
13	numberemergency		0(no Data Value)				0(no DataValue)			2000(100%)		
14	numberinpatient		4.20%				84(4.20%)			1916(95.80%)		
15												
16												

Tab.2 Output on Tabular Representation

	encoun	patienti	admiss	dischar	admiss	timeinh	numlab	numprc	numme	numout	number	number
encounterid	2.30% (46)	97.70% (1954)	0	0	0	0	0	0	0	0	0	0
patientnbr	0.10% (2)	99.85% (1997)	0	0	0	0	0.05% (1)	0	0	0	0	0
admissiontypeid	0	0	99.95% (1999)	0.05% (1)	0	0	0	0	0	0	0	0
lischargepositionid	0	0	67.20% (1344)	3.35% (67)	0.10% (2)	0.80% (16)	22.75% (455)	0	5.80% (116)	0	0	0
admissionsourceid	0	0	10.95% (219)	29.00% (580)	60.00% (1200)	0.05% (1)	0	0	0	0	0	0
timeinhospital	0	0	42.90% (858)	17.15% (343)	10.45% (209)	28.80% (576)	0	0	0.70% (14)	0	0	0
numlabprocedure	0	0	0	0	0	0.10% (2)	96.50% (1930)	0	3.40% (68)	0	0	0
numprocedure	0	0	30.85% (617)	3.00% (60)	0.60% (12)	0.70% (14)	0	64.85% (1297)	0	0	0	0
nummedications	0	0	1.15% (23)	2.25% (45)	4.25% (85)	17.60% (352)	10.45% (209)	0	64.30% (1286)	0	0	0
numoutpatient	0	0	0.05% (1)	0	0	0	0	99.50% (1990)	0	0.25% (5)	0.05% (1)	0.15% (3)
numberemergency	0	0	0.10% (2)	0	0	0	0	99.90% (1998)	0	0	0	0
numberinpatient	0	0	6.05% (121)	0.25% (5)	0	0	0	93.50% (1870)	0	0	0.15% (3)	0.05% (1)

Fig. 4 All Vs. All Graphical output Representation

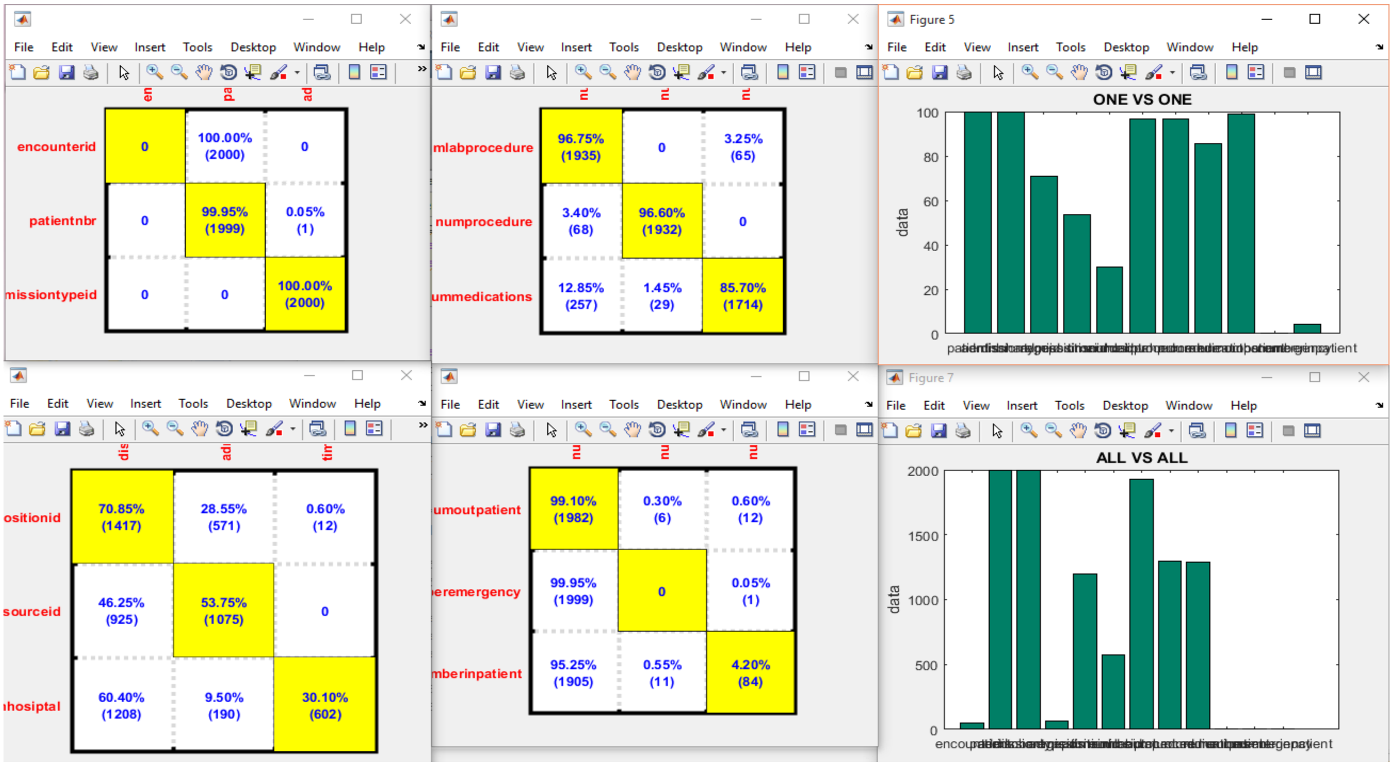

Fig. 5 Overall Evaluation of Diabetes Data for KNN Classifier

Specific representation of output that are compared with data's within its attributes on classes which is shown on One Vs. One for Diabetes Dataset with KNN classifier;

On All Vs. All, all the data's from all the attributes are taken in account were the average of the values are taken which gives the accuracy of all the attributes.

The maximum voting represents the data's compared in which the data's that are corresponding to the class on its attributes to its maximum data's on its sets.

The misclassification represents the data's compared in which the data's that are not corresponding to the class on its attributes.

The Outputs shows the Overall evaluation of the Diabetes Data set using KNN type of Classifier.

Thus give the accuracy of

OVO: 69.75%

AVA: 43.35%

Majority: 100%

Dataset 3. KNN Fisher Iris

Output Results:

KNN(Fisheriris)	One Vsone	ALL VS ALL	Maximum Voting	Mis classification
SETOSA	100%	92%	50(100%)	0%
VERSICOLOR	100%		50(100%)	0%
VIRGINICA	76%		38(76%)	24%

Tab.3 Output on Tabular Representation

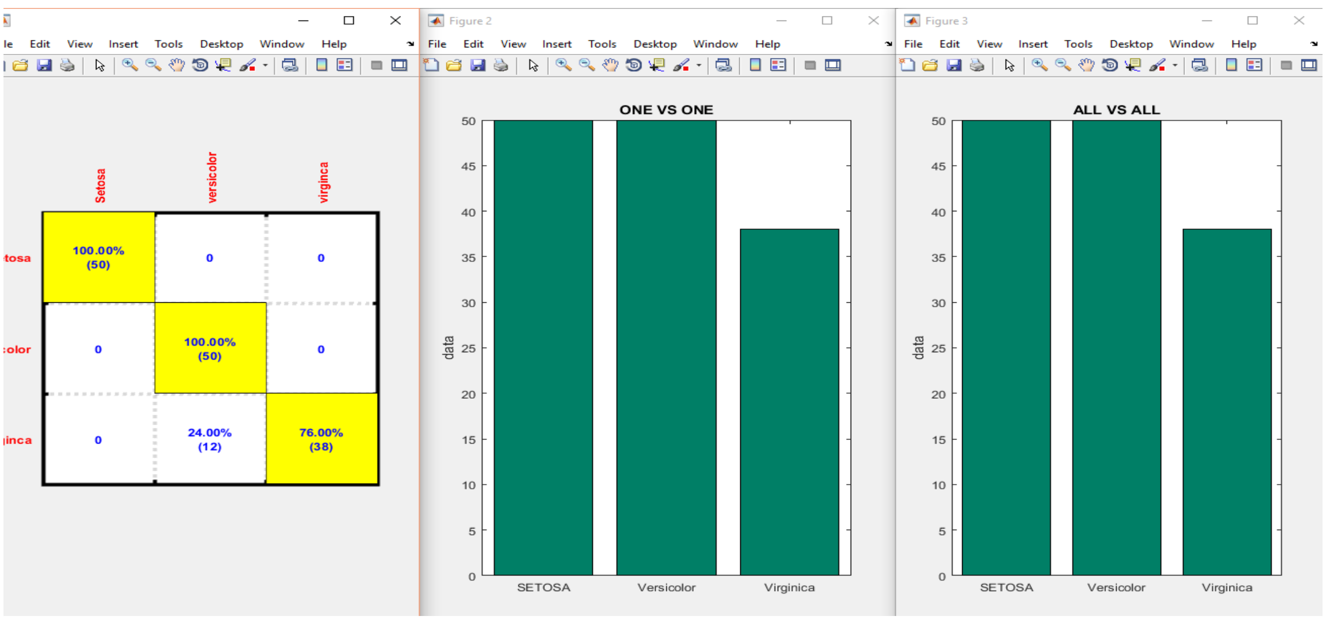

Fig.6 Overall Evaluation of Diabetes Data for KNN Classifier

Specific representation of output that are compared with data's within its attributes on classes which is shown on One Vs. One for Fisher Iris Dataset with KNN classifier; The Outputs shows the Overall evaluation of the Fisher Iris Data set using KNN type of Classifier.

Thus give the accuracy of

OVO: 92%

AVA: 92%

Majority: 100%

Dataset 4. KNN Forest Fire Data:

Output Results:

A1 | KNN(ForestFire)

KNN(ForestFire)	One Vsone	ALL VS ALL	Maximum Voting	Mis classification
Xaxis	86%	80.04%	111(86.05%)	18(13.95%)
Yaxis	11.63%		15(11.63%)	114(88.38%)
Month	64%		82(63.57%)	47(36.43%)
Day	100%		129(100%)	0(0%)
FFMC	98.45%		127(98.45%)	2(1.55%)
DMC	88%		114(88.37%)	15(11.63%)
DC	93%		120(93.02%)	9(6.98%)
ISI	90.70%		117(90.70%)	12(9.30%)
TEMP	85.27%		110(85.27%)	19(14.73%)
RTI	0(no Value)		0(no Value)	129(100%)
WIND	99%		128(99.22%)	1(0.78%)
RAIN	0(no Value)		0(no Value)	129(100%)

Tab.4 Output on Tabular Representation

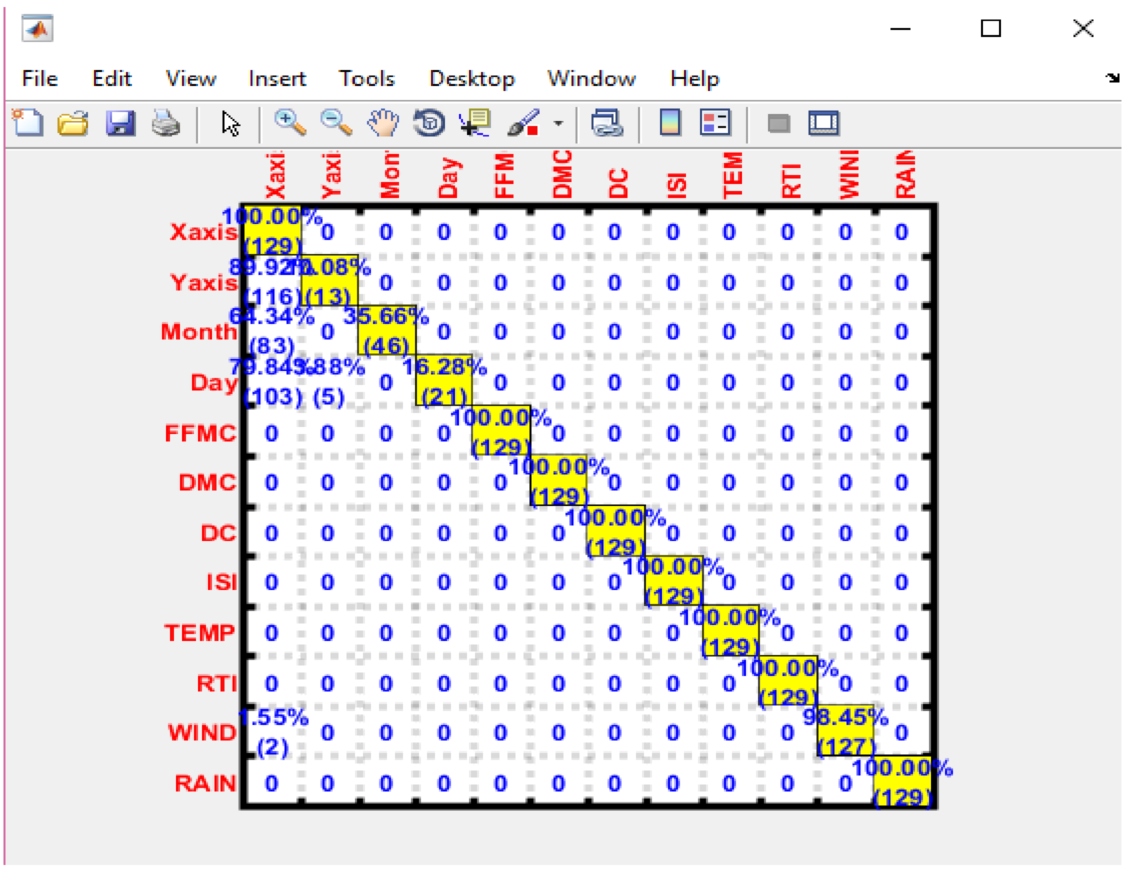

Fig.7 Evaluation of Forest Fire for KNN Classifier

Fig.8 Overall Evaluation of Forest Fire Data for KNN Classifier

Specific representation of output that are compared with data's within its attributes on classes which is shown on One Vs. One Forest Fire Dataset with KNN classifier; On All Vs. All, all the data's from all the attributes are taken in account were the average of the values are taken which gives the accuracy of all the attributes.

The Outputs shows the Overall evaluation of the Forest Fire Data set using KNN type of Classifier.

Thus give the accuracy of

OVO: 68.0044%

AVA: 80.04%

Majority: 100%

Dataset 5. KNN Wine:

Output Results:

	KNN(Wi	One Vsone	ALL VS ALL	Maximum Voting	Mis classification
1	KNN(Wi	One Vsone	ALL VS ALL	Maximum Voting	Mis classification
3	alcohol	100%	62%	44(100%)	0(0%)
4	Malic Acid	72.73%		32(72.73%)	12(27.27%)
5	Ash	95%		42(95.45%)	2(4.55%)
6	Alcalinityofash	100%		44(100%)	0(0%)
7	Magnesium	100%		44(100%)	0(%)
8	Total Phenols	100%		44(100%)	0(%)
9	Flavonids	75%		33(75%)	11(25%)
10	NonFlavonoidsPhen	100%		44(100%)	0(0%)
11	Proanthocyanis	70.45%		31(70.45%)	13(29.55%)
12	ColorIntensity	75%		33(75%)	11(25%)
13	Hue	100%		44(100%)	0(0%)
14	DilutedWine	70%		31(70.45%)	13(29.55%)

Tab.5 Output on Tabular Representation

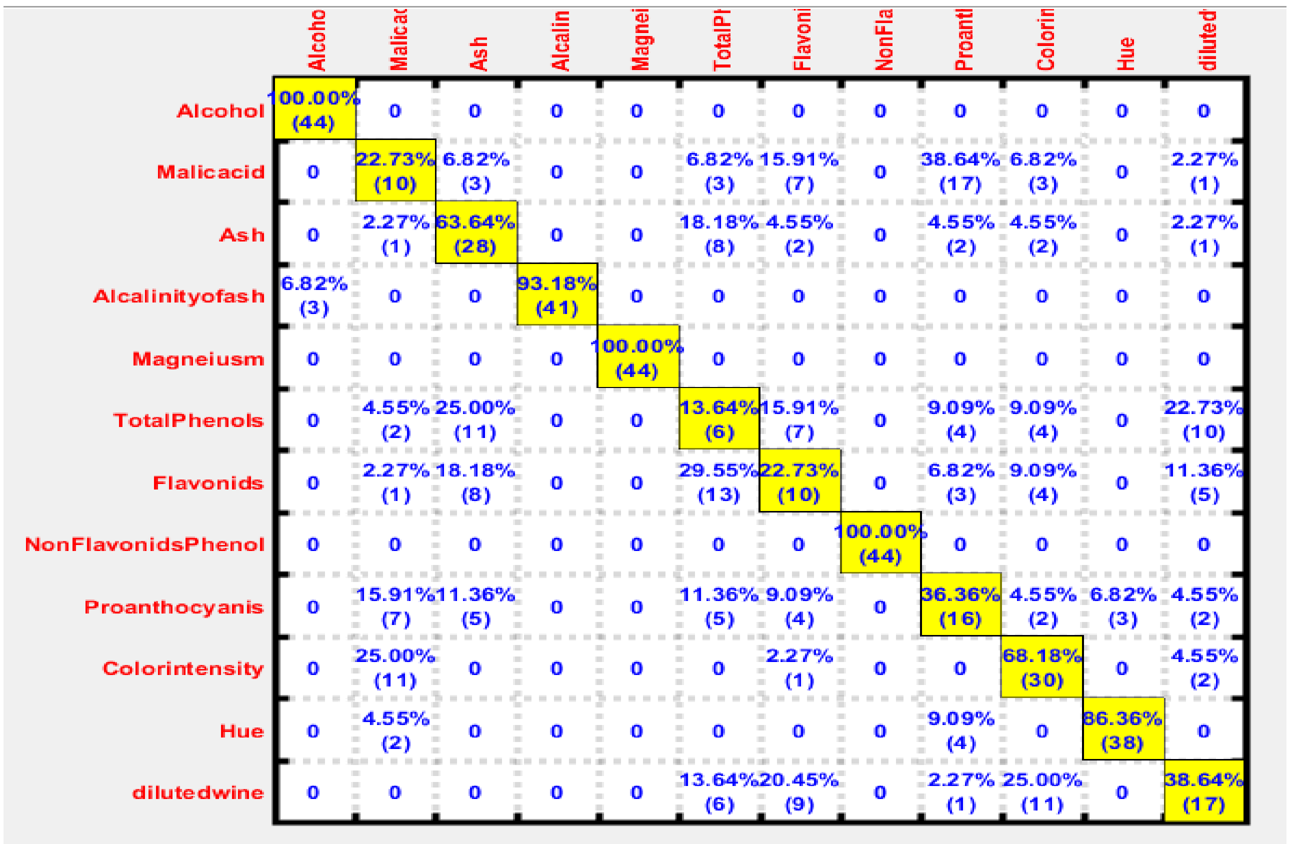

Fig. 9 Evaluation of Wine Data for KNN Classifier with output results

Fig. 10 Overall Evaluation of Wine Data for KNN Classifier

Specific representation of output that are compared with data's within its attributes on classes which is shown on One Vs. One Wine Dataset with KNN classifier; On All Vs. All, all the data's from all the attributes are taken in account were the average of the values are taken which gives the accuracy of all the attributes.

The Outputs shows the Overall evaluation of the Wine Data set using KNN type of Classifier.

Thus give the accuracy of

OVO: 88.1816%

AVA: 62.12%

Majority: 100%

Algorithm 2. Support Vector Machine (SVM)

Dataset 1. SVM Breast Cancer:

Output Results:

A1 | f_x | SVM(BreastCancer)

	A	B	C	D	E	F	G	H	I	J	K	L
1	SVM(BreastCan		One Vsone		ALL VS ALL		Maximum Voting			Mis classification		
2												
3	Radius		100%		27.59%		58(100%)			0(0%)		
4	Texture		96.55%				56(96.55%)			2(3.45%)		
5	Perimeter		17.24%				10(17.24%)			48(82.76%)		
6	Area		70.69%				41(70.69%)			17(29.31%)		
7	Smoothness		15.52%				9(15.52%)			49(84.48%)		
8	Compactness		48.28%				28(48.28%)			30(51.72%)		
9	Concavity		98.28%				57(98.28%)			1(1.72%)		
10	ConcavePoints		27.59%				16(27.59%)			42(72.41%)		
11	Symmetry		0(No data Variable)				0(No data Variable)			58(100%)		
12												
13												

Tab.6 Output on Tabular Representation

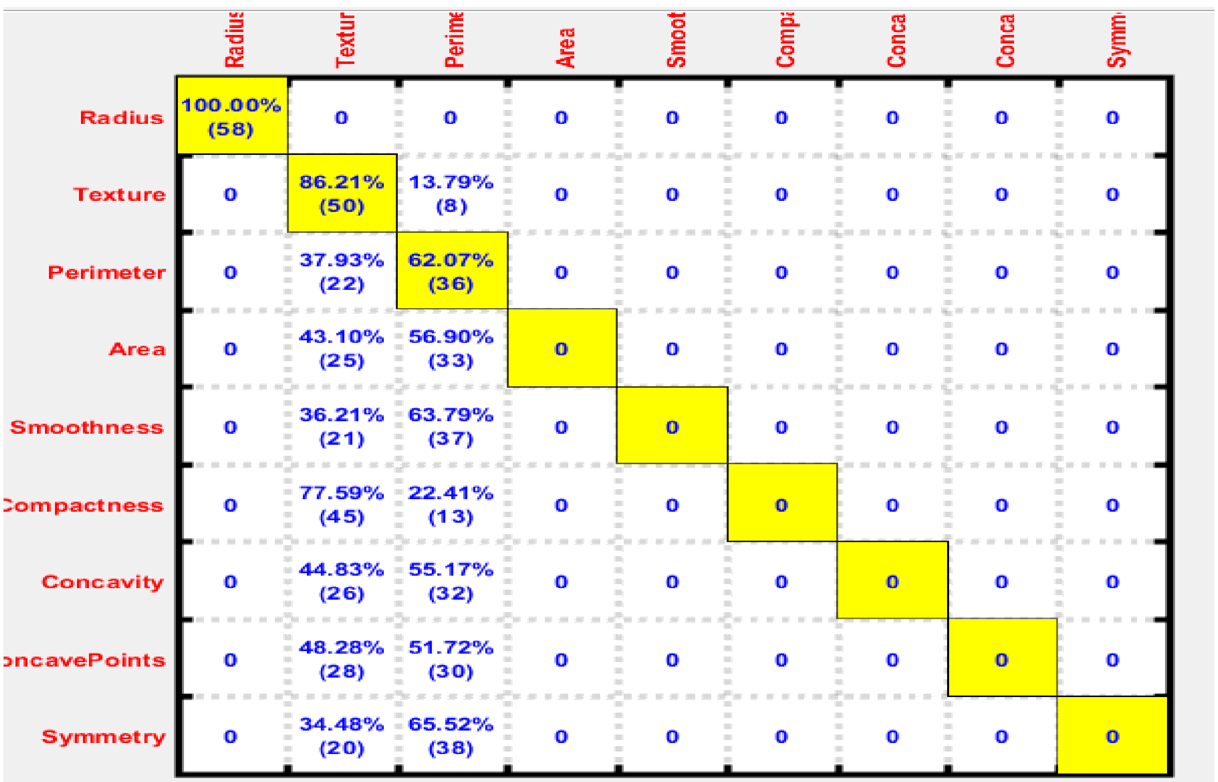

Fig.11 Evaluation of Breast Cancer Data for SVM Classifier with output results

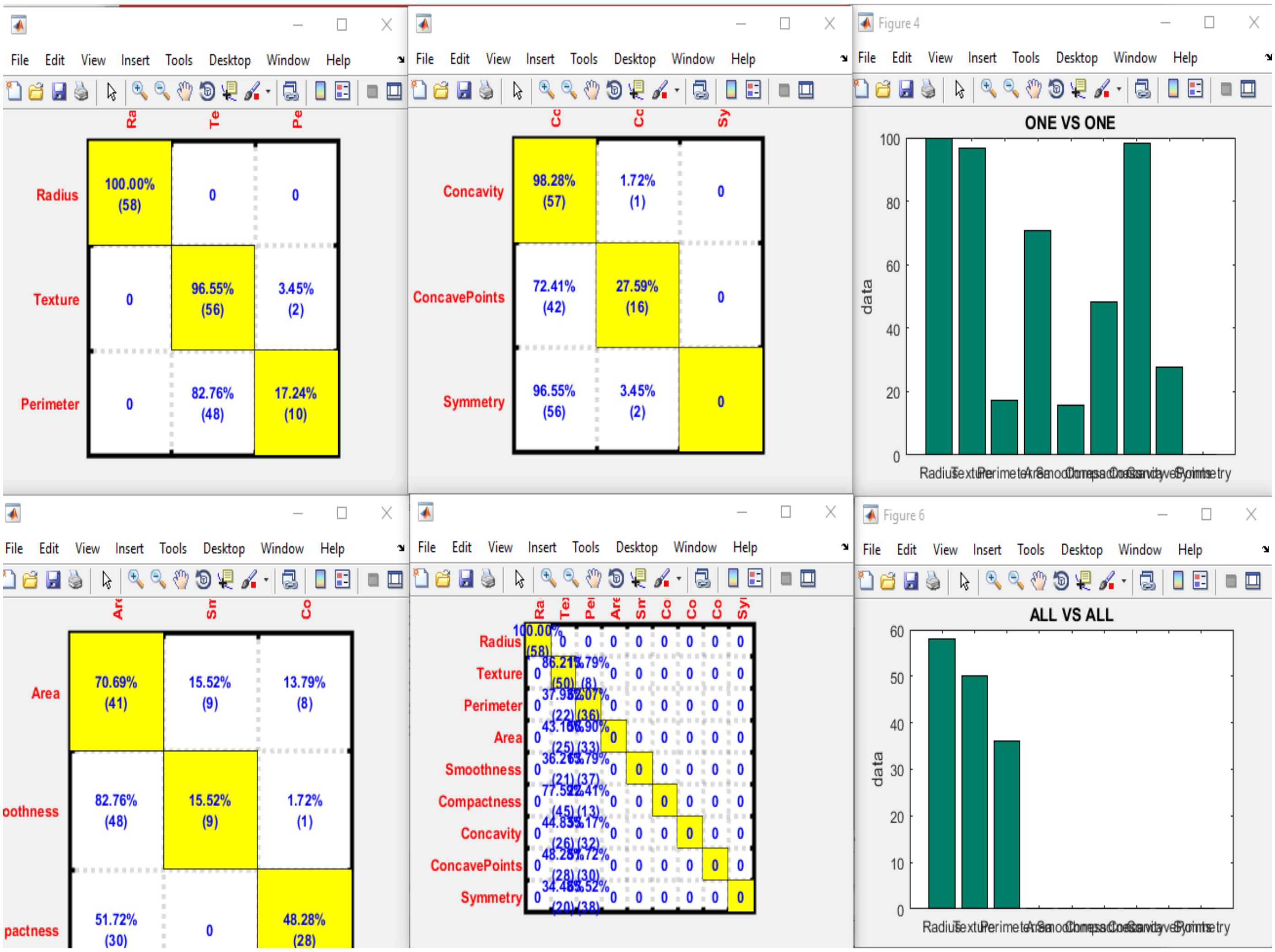

Fig.12 Overall Evaluation of Breast Cancer Data for SVM Classifier

Specific representation of output that are compared with data's within its attributes on classes which is shown on One Vs. One Breast Cancer Dataset with SVM classifier; On All Vs. All, all the data's from all the attributes are taken in account were the average of the values are taken which gives the accuracy of all the attributes.

The Outputs shows the Overall evaluation of the Breast Cancer Data set using SVM type of Classifier.

Thus give the accuracy of

OVO: 52.683%

AVA: 27.59%

Majority: 100%

Dataset 2. SVM Diabetics:

Output Results:

A1 SVM(Diabetics)

SVM(Diabetics)	One Vsone	ALL VS ALL	Maximum Voting	Mis classification
EncounterID	100%	49.06%	666(100%)	0(0%)
Patientbr	98.20%		654(98.20%)	12(1.80%)
Admission typeid	100%		666(100%)	0(%)
Discharge positionid	100%		666(100%)	0(%)
admissionsourceid	2.70%		18(2.70%)	648(97.30%)
Timein hospital	0(No data Variaable)		0(No data Variaable)	666(100%)
numlab procedure	99.25%		661(99.25%)	5(0.75%)
numprocedure	100.00%		666(100%)	0(0%)
nummedications	98.95%		659(98.95%	7(1.05%)
NumoutPatient	0.30%		2(0.30%)	664(99.70%)
numberemergency	1.05%		7(1.05%)	659(98.95%)
numberinpatient	99.85%		665(99.85%)	1(0.015%)

Tab.7 Output on Tabular Representation

	encour	patient	admiss	discha	admiss	timeinh	numlab	numpr	numme	numou	numbe	numbe
encounterid	100.00% (666)	0	0	0	0	0	0	0	0	0	0	0
patientnbr	1.50% (10)	98.20% (654)	0	0	0	0	0.30% (2)	0	0	0	0	0
admissiontypeid	0	0	100.00% (666)	0	0	0	0	0	0	0	0	0
discharegepositionid	0	0	68.02% (453)	31.53% (210)	0.45% (3)	0	0	0	0	0	0	0
admissionsourceid	0	0	56.76% (378)	10.81% (72)	31.23% (208)	0	0	0	1.20% (8)	0	0	0
timeinhosiptal	0	0	75.98% (506)	9.61% (64)	14.41% (96)	0	0	0	0	0	0	0
numlabprocedure	0	0	0	0	0.30% (2)	0	98.95% (659)	0	0.75% (5)	0	0	0
numprocedure	0	0	72.07% (480)	1.50% (10)	0	0	0	0	0	25.98% (173)	0.30% (2)	0.15% (1)
nummedications	0	0	1.05% (7)	0.45% (3)	59.31% (395)	0	9.76% (65)	0	29.43% (196)	0	0	0
numoutpatient	0	0	0.30% (2)	0.30% (2)	0	0	0	0	0	99.40% (662)	0	0
numberemergency	0	0	0.15% (1)	0	0	0	0	0	0	99.85% (665)	0	0
numberinpatient	0	0	17.57% (117)	2.85% (19)	0.15% (1)	0	0	0	0	79.43% (529)	0	0

Fig. 13 Evaluation of Diabetes Data for SVM Classifier with output results

Fig. 14 Overall Evaluation of Diabetes Data for SVM Classifier

Specific representation of output that are compared with data's within its attributes on classes which is shown on One Vs. One Wine Dataset with KNN classifier; On All Vs. All, all the data's from all the attributes are taken in account were the average of the values are taken which gives the accuracy of all the attributes.

The Outputs shows the Overall evaluation of the Wine Data set using KNN type of Classifier.

Thus give the accuracy of

OVO: 66.691%

AVA: 49.06%

Majority: 100%

Dataset 3. SVM Fisher Iris:

Output Results:

SVM(Fisheriris)	One Vsone	ALL VS ALL	Maximum Voting	Mis classification
SETOSA	100%	94%	16(100%)	0(0%)
VERSICOLOR	88%		14(87.50%)	2(12.50%)
VIRGINICA	94%		15(93.75%)	1(6.25%)

Tab.8 Output on Tabular Representation

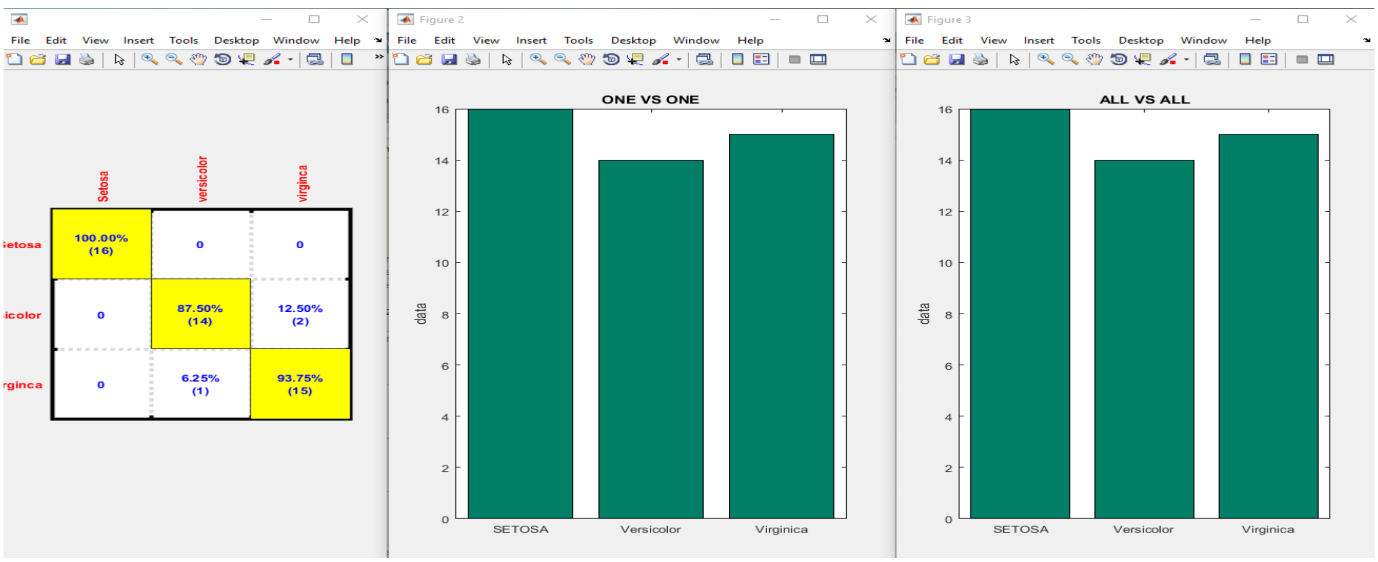

Fig. 15 Overall Evaluation of Fisher Iris Data for SVM Classifier

Specific representation of output that are compared with data's within its attributes on classes which is shown on One Vs. One for Fisher Iris Dataset with SVM classifier; The Outputs shows the Overall evaluation of the Fisher Iris Data set using SVM type of Classifier.

Thus give the accuracy of

OVO: 94%

AVA: 94%

Majority: 100%

Dataset 4. SVM Forest Fire Data:

Output Results:

A1 | f_x SVM(ForestFire)

	A	B	C	D	E	F	G	H	I	J	K	L
1	SVM(ForestFire)		One Vsone		ALL VS ALL		Maximum Voting			Mis classification		
2												
3	Xaxis		79%		76.94%		34(79.07%)			9(20.93%)		
4	Yaxis		83.72%				36(83.72%)			7(16.28%)		
5	Month		95%				41(95.35%)			2(4.65%)		
6	Day		100%				43(100%)			0(0%)		
7	FFMC		100.00%				43(100%)			0(0%)		
8	DMC		95%				41(95.35%)			2(4.65%)		
9	DC		100%				43(100%)			0(0%)		
10	ISI		97.67%				42(97.67%)			1(2.33%)		
11	TEMP		93.02%				40(93.02%)			3(6.98%)		
12	RTI		100%				43(100%)			0(0%)		
13	WIND		100%				43(100%)			0(0%)		
14	RAIN		100%				43(100%)			0(0%)		
15												

Tab.9 Output on Tabular Representation

	Xaxis	Yaxis	Month	Day	FFMC	DMC	DC	ISI	TEMP	RTI	WIND	RAIN
Xaxis	44.19% (19)	13.95% (6)	4.65% (2)	23.26% (10)	0	0	0	0	0	0	13.95% (6)	0
Yaxis	6.98% (3)	88.37% (38)	0	4.65% (2)	0	0	0	0	0	0	0	0
Month	4.65% (2)	0	88.37% (38)	0	0	0	0	6.98% (3)	0	0	0	0
Day	27.91% (12)	4.65% (2)	0	55.81% (24)	0	0	0	0	0	0	11.63% (5)	0
FFMC	0	0	0	0	100.00% (43)	0	0	0	0	0	0	0
DMC	0	0	0	0	9.30% (4)	88.37% (38)	0	0	0	2.33% (1)	0	0
DC	0	0	0	0	0	2.33% (1)	97.67% (42)	0	0	0	0	0
ISI	2.33% (1)	4.65% (2)	32.56% (14)	0	0	0	0	58.14% (25)	2.33% (1)	0	0	0
TEMP	0	0	0	0	0	0	0	6.98% (3)	93.02% (40)	0	0	0
RTI	0	0	0	0	0	0	0	0	0	100.00% (43)	0	0
WIND	23.26% (10)	44.19% (19)	0	23.26% (10)	0	0	0	0	0	0	9.30% (4)	0
RAIN	0	0	0	0	0	0	0	0	0	0	0	100.00% (43)

Fig. 16 Evaluation of Forest Fire Data for SVM Classifier with output results

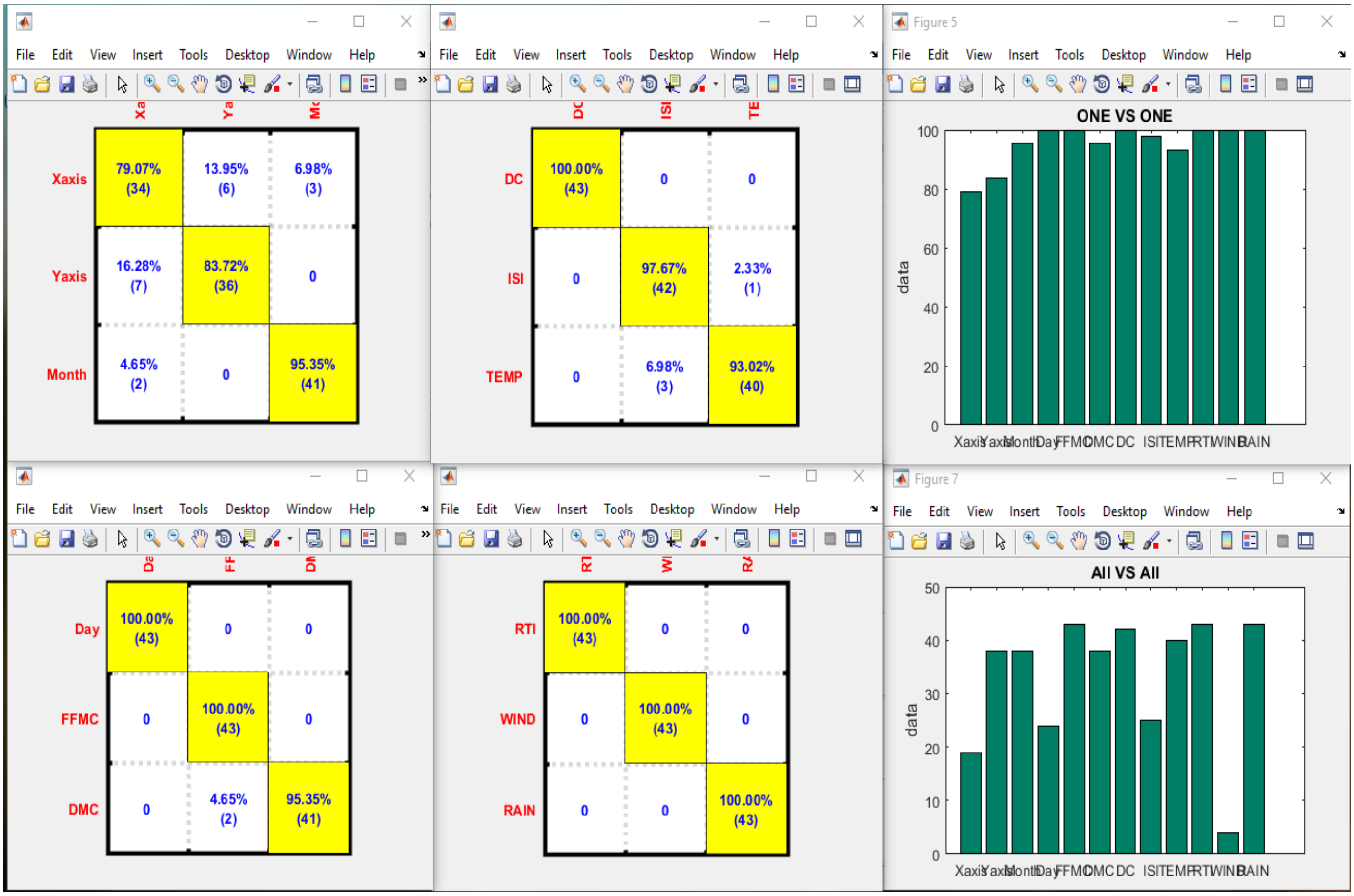

Fig. 17 Overall Evaluation of Forest Fire Data for SVM Classifier

Specific representation of output that are compared with data's within its attributes on classes which is shown on One Vs. One Forest Fire Dataset with SVM classifier; On All Vs. All, all the data's from all the attributes are taken in account were the average of the values are taken which gives the accuracy of all the attributes.

The Outputs shows the Overall evaluation of the Forest Fire Data set using SVM type of Classifier.

Thus give the accuracy of

OVO: 95.28%

AVA: 76.94%

Majority: 100%

Dataset 5. SVM Wine:

Output Results:

A1 SVM(Wine) Attributes

	A	C	E	G	J
1	SVM(Wi	One Vsone	ALL VS ALL	Maximum Voting	Mis classification
2					
3	alcohol	100%	85%	14(100%)	0(0%)
4	Malic Acid	78.57%		11(78.57%)	3(21.43%)
5	Ash	100%		14(100%)	0(0%)
6	Alcalinityofash	100%		14(100%)	0(0%)
7	Magnesium	100%		14(100%)	0(0%)
8	Total Phenols	100%		14(100%)	0(0%)
9	Flavonids	79%		11(78.57%)	3(21.43%)
10	NonFlavonoidsPhen	100%		14(100%)	0(0%)
11	Proanthocyanis	100.00%		14(100%)	0(%)
12	ColorIntensity	100%		14(100%)	0(%)
13	Hue	100%		14(100%)	0(%)
14	DilutedWine	100%		14(100%)	0(%)
15					

Tab.10 Output on Tabular Representation

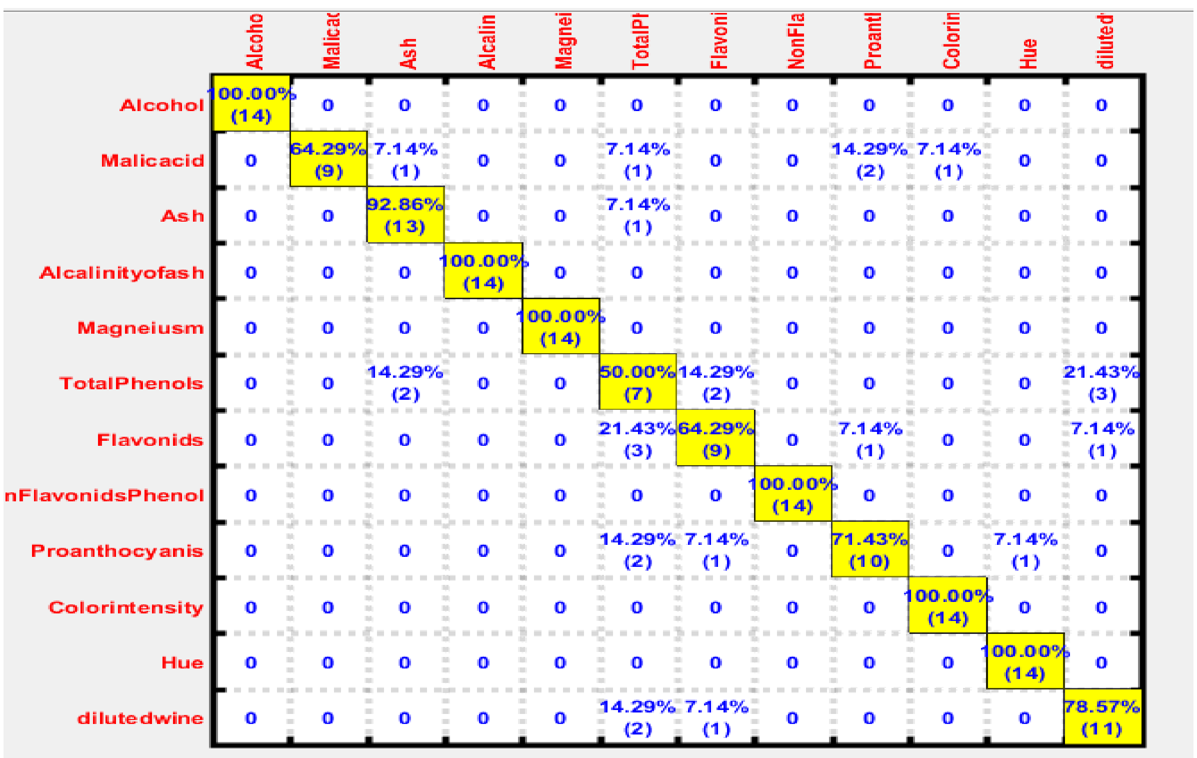

Fig. 18 Evaluation of Wine Data for SVM Classifier with output results

Fig. 19 Overall Evaluation of Wine Data for SVM Classifier

Specific representation of output that are compared with data's within its attributes on classes which is shown on One Vs. One Wine Dataset with SVM classifier; On All Vs. All, all the data's from all the attributes are taken in account were the average of the values are taken which gives the accuracy of all the attributes.

The Outputs shows the Overall evaluation of the Wine Data set using SVM type of Classifier.

Thus give the accuracy of

OVO: 96.46%

AVA: 85%

Majority: 100%

Algorithm 3. Bayes Classifier

Dataset 1. Bayes Breast Cancer:

Output Results:

A1 — Bayes(BreastCancer)

Bayes(BreastCa	One Vsone	ALL VS ALL	Maximum Voting	Mis classification
Radius	100%	32.91%	34(100%)	0(0%)
Texture	88.57%		31(88.57%)	4(11.43%)
Perimeter	31.43%		11(31.43%)	24(68.57%)
Area	64.71%		22(64.71%)	12(35.29%)
Smoothness	28.57%		10(28.57%)	25(71.43%)
Compactness	60.00%		21(60%)	14(40%)
Concavity	55.88%		19(55.88%)	15(44.12%)
ConcavePoints	100.00%		35(100%)	0(0%)
Symmetry	0(no data Value)		0(no data Value)	35(100%)

Tab.11 Output on Tabular Representation

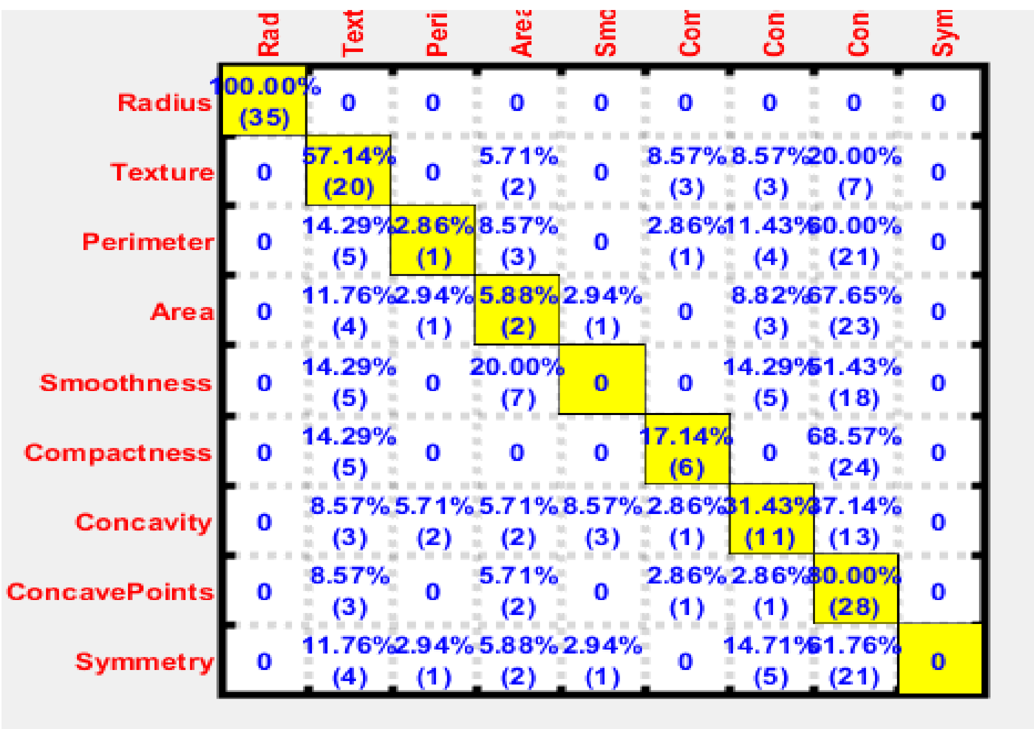

Fig. 20 Evaluation of Breast Cancer Data for Bayes Classifier with output results

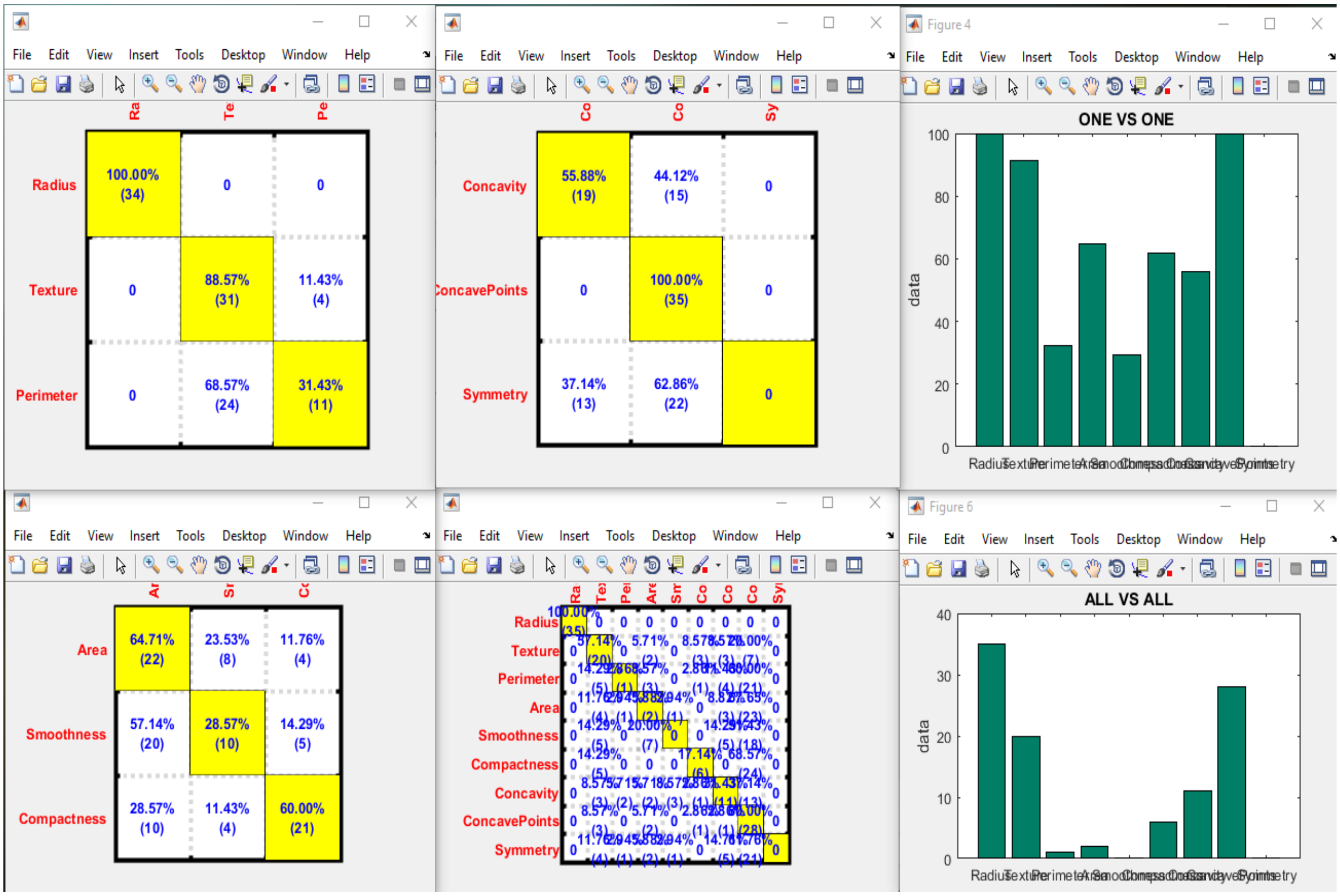

Fig. 21 Overall Evaluation of Breast Cancer Data for Bayes Classifier

Specific representation of output that are compared with data's within its attributes on classes which is shown on One Vs. One Breast Cancer Dataset with Bayes classifier; On All Vs. All, all the data's from all the attributes are taken in account were the average of the values are taken which gives the accuracy of all the attributes.

The Outputs shows the Overall evaluation of the Breast Cancer Data set using Bayes type of Classifier.

Thus give the accuracy of

OVO: 58.7955%

AVA: 32.91%

Majority: 100%

Dataset 2. Bayes Diabetics:

Output Results:

	A	B	C	D	E	F	G	H	I	J	K	L
1	**Bayes(Diabetics**		**One Vsone**		**ALL VS ALL**		**Maximum Voting**			**Mis classification**		
2												
3	EncounterID		100%		63.19%		400(100%)			0(0%)		
4	Patientbr		100.00%				400(100%)			0(0%)		
5	Admission typeid		100%				400(100%)			0(0%)		
6	Discharge positionid		23.75%				95(23.75%)			305(76.25%)		
7	admissionsourceid		49.00%				196(49%)			204(51%)		
8	Timein hospital		96.25%				385(96.25%)			15(3.75%)		
9	numlab procedure		99.00%				396(99%)			4(1%)		
10	numprocedure		100.00%				400(100%)			0(0%)		
11	nummedications		98.50%				394(98.50%)			6(1.50%)		
12	NumoutPatient		2.00%				8(2%)			392(98%)		
13	numberemergency		99.75%				399(99.75)			1(0.25%)		
14	numberinpatient		3.25%				13(3.25)			387(96.75%)		
15												

Tab.12 Output on Tabular Representation

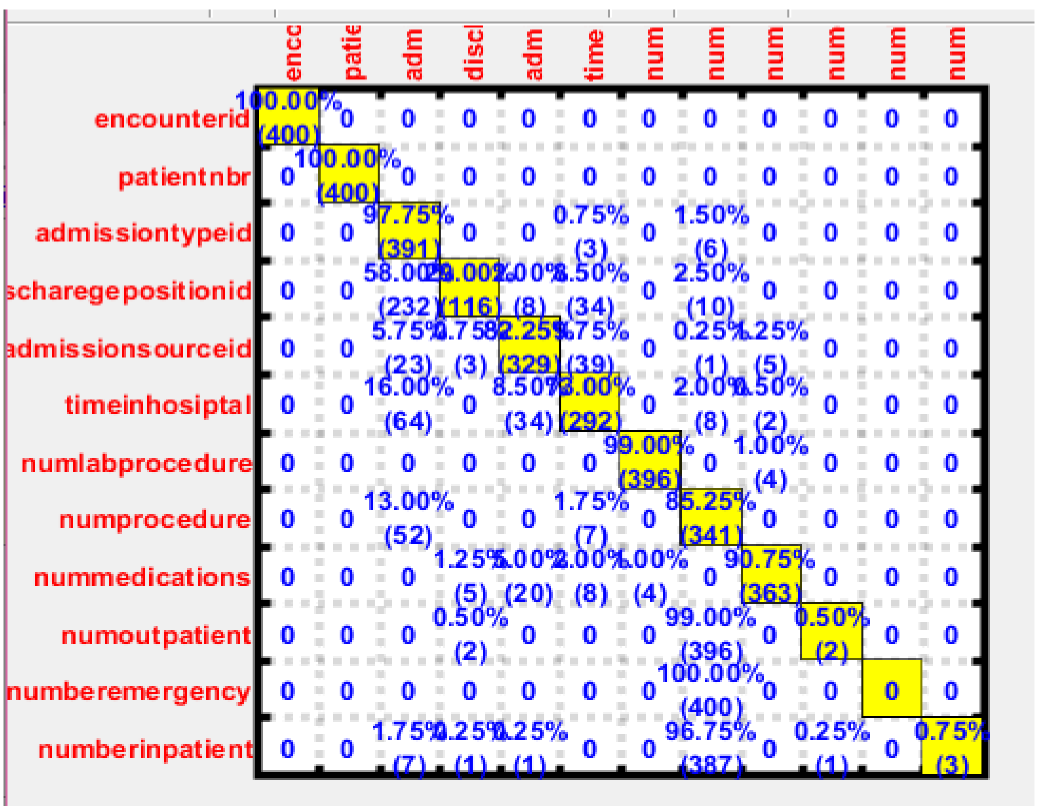

Fig. 22 Evaluation of Diabetes Data for Bayes Classifier with output results

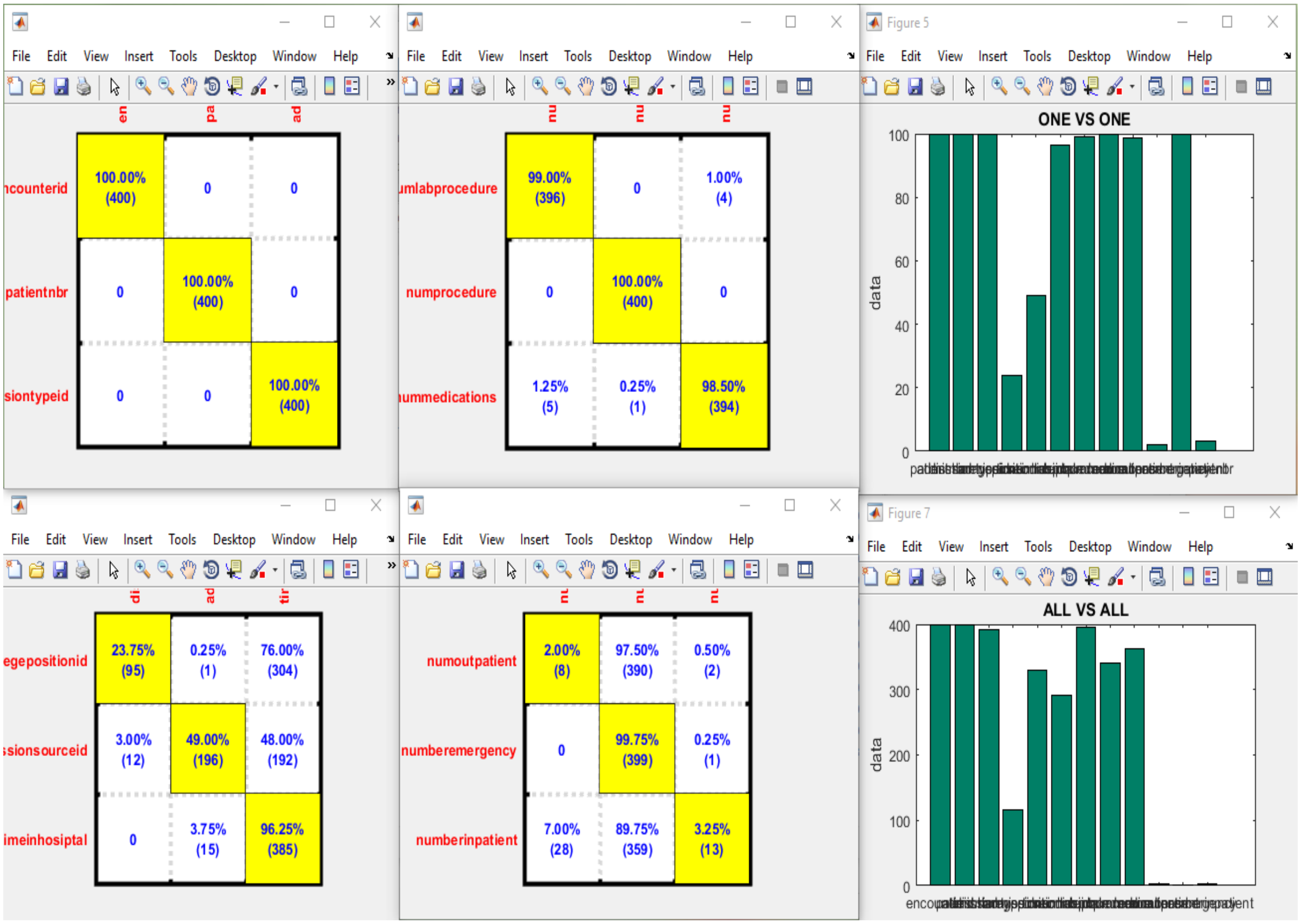

Fig. 23 Overall Evaluation of Diabetes Data for Bayes Classifier

Specific representation of output that are compared with data's within its attributes on classes which is shown on One Vs. One Diabetics Dataset with Bayes classifier; On All Vs. All, all the data's from all the attributes are taken in account were the average of the values are taken which gives the accuracy of all the attributes.

The Outputs shows the Overall evaluation of the Diabetics Data set using Bayes type of Classifier.

Thus give the accuracy of

OVO: 72.625%

AVA: 63.19%

Majority: 100%

Dataset 3. Bayes Fisher Iris:

Output Results:

Bayes(Fisheriris	One Vsone	ALL VS ALL	Maximum Voting	Mis classification
SETOSA	100%	96%	15(100%)	0(0%)
VERSICOLOR	93%		14(93.33%)	1(6.67%)
VIRGINICA	93%		14(93.33%)	1(6.67%)

Tab.13 Output on Tabular Representation

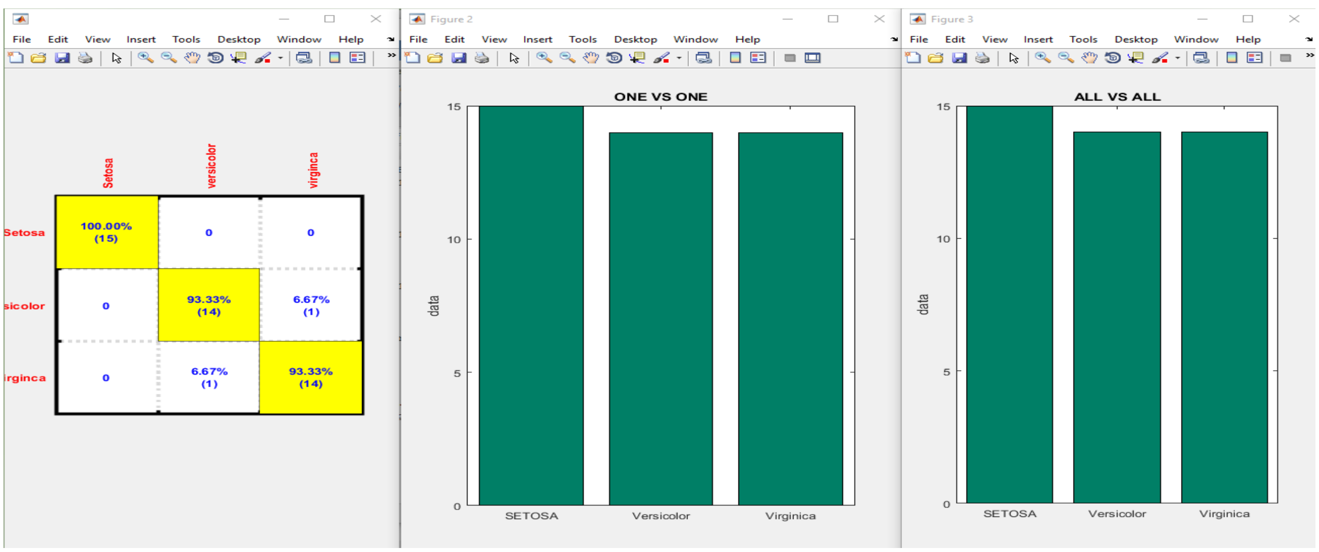

Fig. 24 Overall Evaluation of Fisher Iris Data for Bayes Classifier

Specific representation of output that are compared with data's within its attributes on classes which is shown on One Vs. One for Fisher Iris Dataset with Bayes classifier; The Outputs shows the Overall evaluation of the Fisher Iris Data set using Bayes type of Classifier.

Thus give the accuracy of

OVO: 96%

AVA: 96%

Majority: 100%

Dataset 4. Bayes Forest Fire Data:

Output Results:

A1 | Bayes(ForestFire)

	A	B	C	D	E	F	G	H	I	J	K	L
1	**Bayes(ForestFir**		**One Vsone**		**ALL VS ALL**		**Maximum Voting**			**Mis classification**		
2												
3	Xaxis		72%		77.35%		18(72%)			7(28%)		
4	Yaxis		92.31%				24(92.31%)			2(7.70%)		
5	Month		100%				26(100%)			0(0%)		
6	Day		100%				25(100%)			0(0%)		
7	FFMC		96.15%				25(96.15%)			1(3.85%)		
8	DMC		100%				26(100%)			0(0%)		
9	DC		100%				25(100%)			0(0%)		
10	ISI		96.15%				25(96.15%)			1(3.85%)		
11	TEMP		96.15%				25(96.15%)			1(3.85%)		
12	RTI		100%				25(100%)			0(0%)		
13	WIND		100%				26(100%)			0(0%)		
14	RAIN		100%				26(100%)			0(0%)		
15												

Tab.14 Output on Tabular Representation

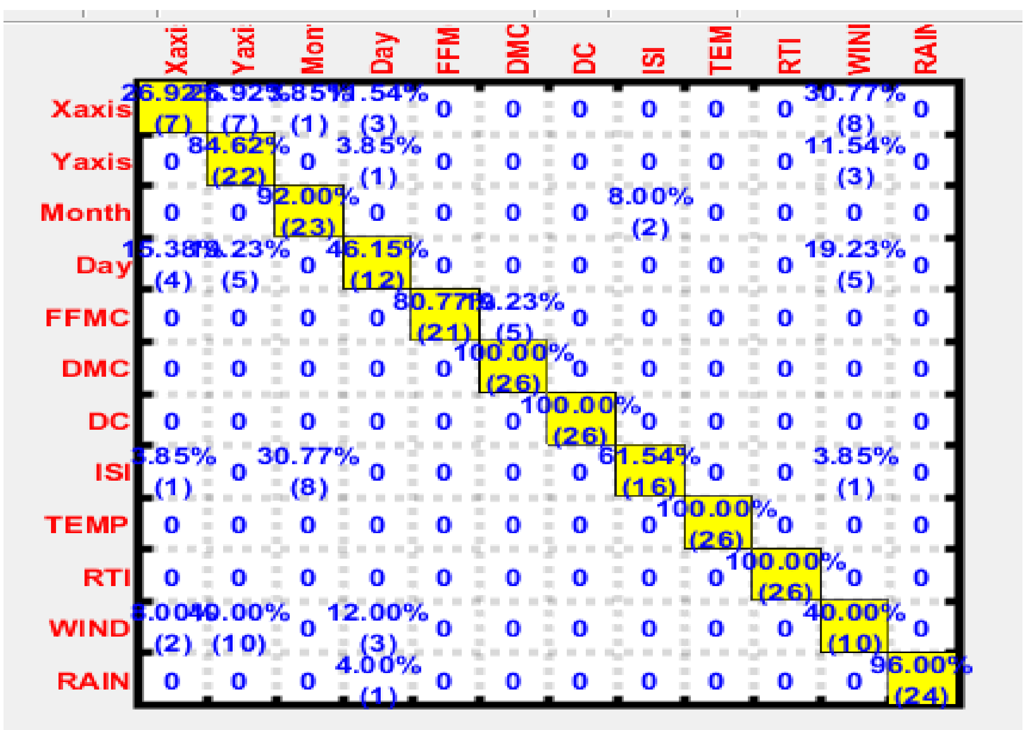

Fig. 25 Evaluation of Forest Fire Data for Bayes Classifier with output results

Fig. 26 Overall Evaluation of Forest Fire Data for Bayes Classifier

Specific representation of output that are compared with data's within its attributes on classes which is shown on One Vs. One Forest Fire Dataset with Bayes classifier; On All Vs. All, all the data's from all the attributes are taken in account were the average of the values are taken which gives the accuracy of all the attributes.

The Outputs shows the Overall evaluation of the Forest Fire Data set using Bayes type of Classifier.

This gives the accuracy of

OVO: 96.063%

AVA: 77.35%

Majority: 100%

Dataset 5. Bayes Wine:

Output Results:

	A	B	C	D	E	F	G	H	I	J	K	L
1	**Bayes(W**		**One Vsone**		**ALL VS ALL**		**Maximum Voting**			**Mis classification**		
2												
3	alcohol		100%		90%		8(100%)			0(0%)		
4	Malic Acid		100.00%				9(100%)			0(0%)		
5	Ash		100%				9(100%)			0(0%)		
6	Alcalinityofash		100%				8(100%)			0(0%)		
7	Magnesium		100%				9(100%)			0(0%)		
8	Total Phenols		100%				9(100%)			0(0%)		
9	Flavonids		100%				8(100%)			0(0%)		
10	NonFlavonoidsPhen		100%				9(100%)			0(0%)		
11	Proanthocyanis		88.89%				8(88.89%)			1(11.11%)		
12	ColorIntensity		100%				8(100%)			0(0%)		
13	Hue		100%				9(100%)			0(0%)		
14	DilutedWine		100%				9(100%)			0(0%)		
15												

Tab.15 Output on Tabular Representation

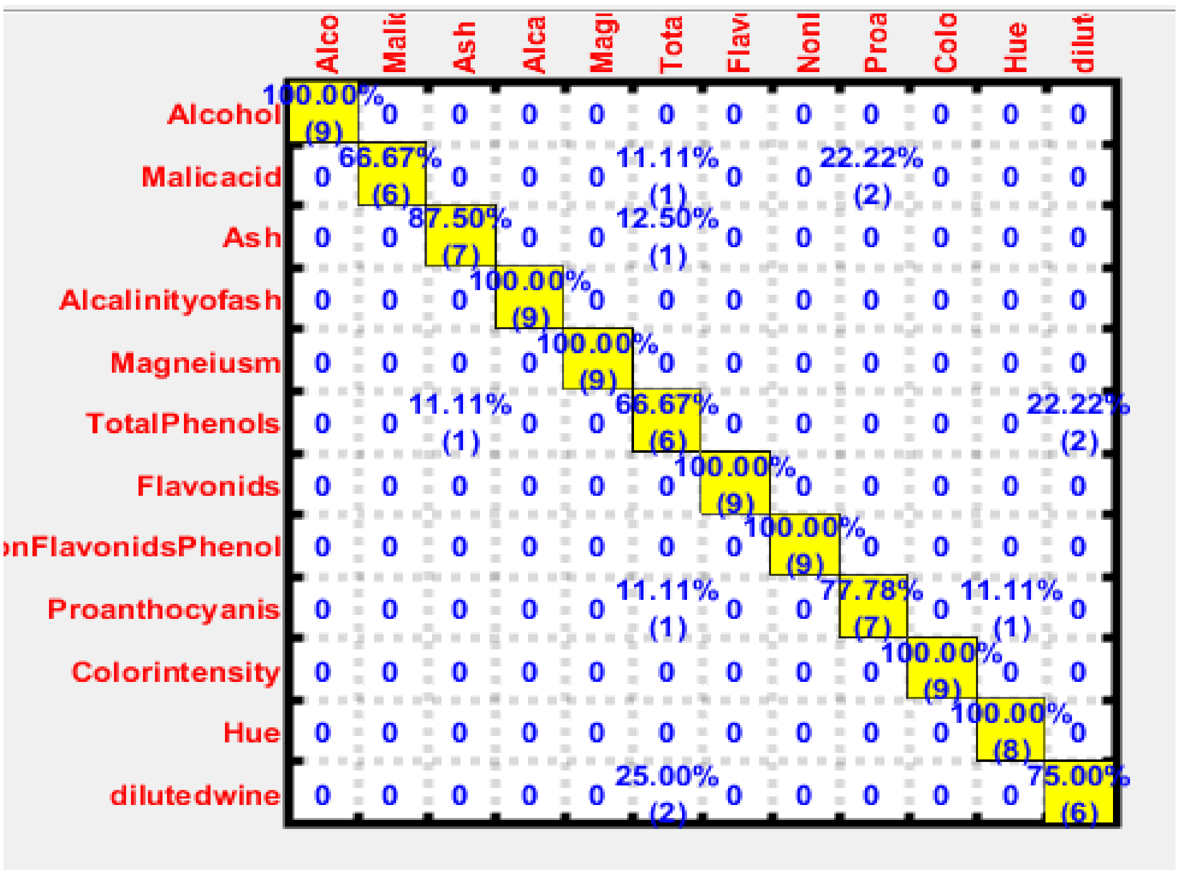

Fig. 27 Evaluation of Wine Data for Bayes Classifier with output results

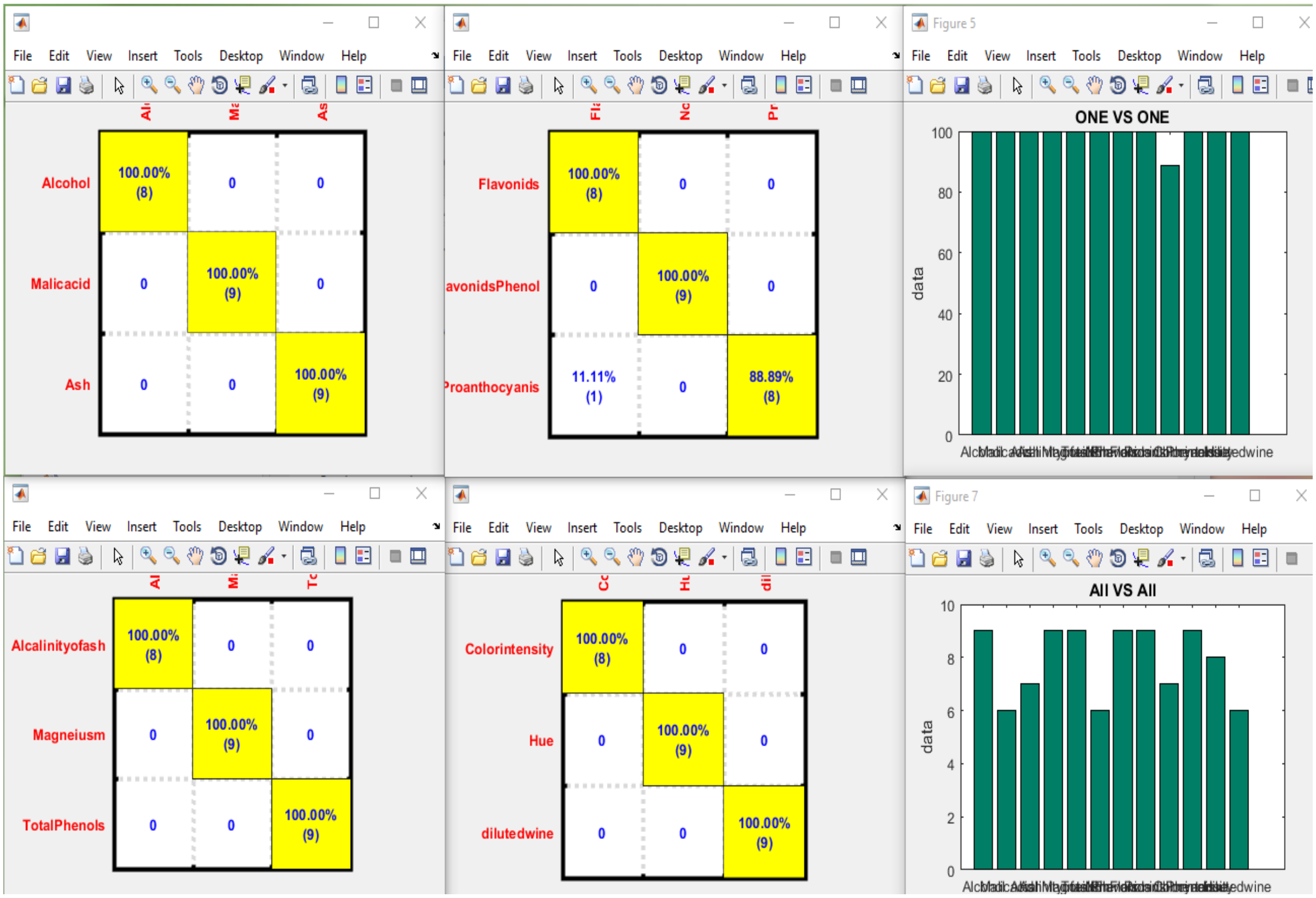

Fig. 28 Overall Evaluation of Wine Data for Bayes Classifier

Specific representation of output that are compared with data's within its attributes on classes which is shown on One Vs. One Wine Dataset with Bayes classifier; On All Vs. All, all the data's from all the attributes are taken in account were the average of the values are taken which gives the accuracy of all the attributes.

The Outputs shows the Overall evaluation of the Wine Data set using Bayes type of Classifier.

This gives the accuracy of

OVO: 99.0741%

AVA: 90%

Majority: 100%

5.1 Comparison of Different Classifiers with its corresponding Outputs:

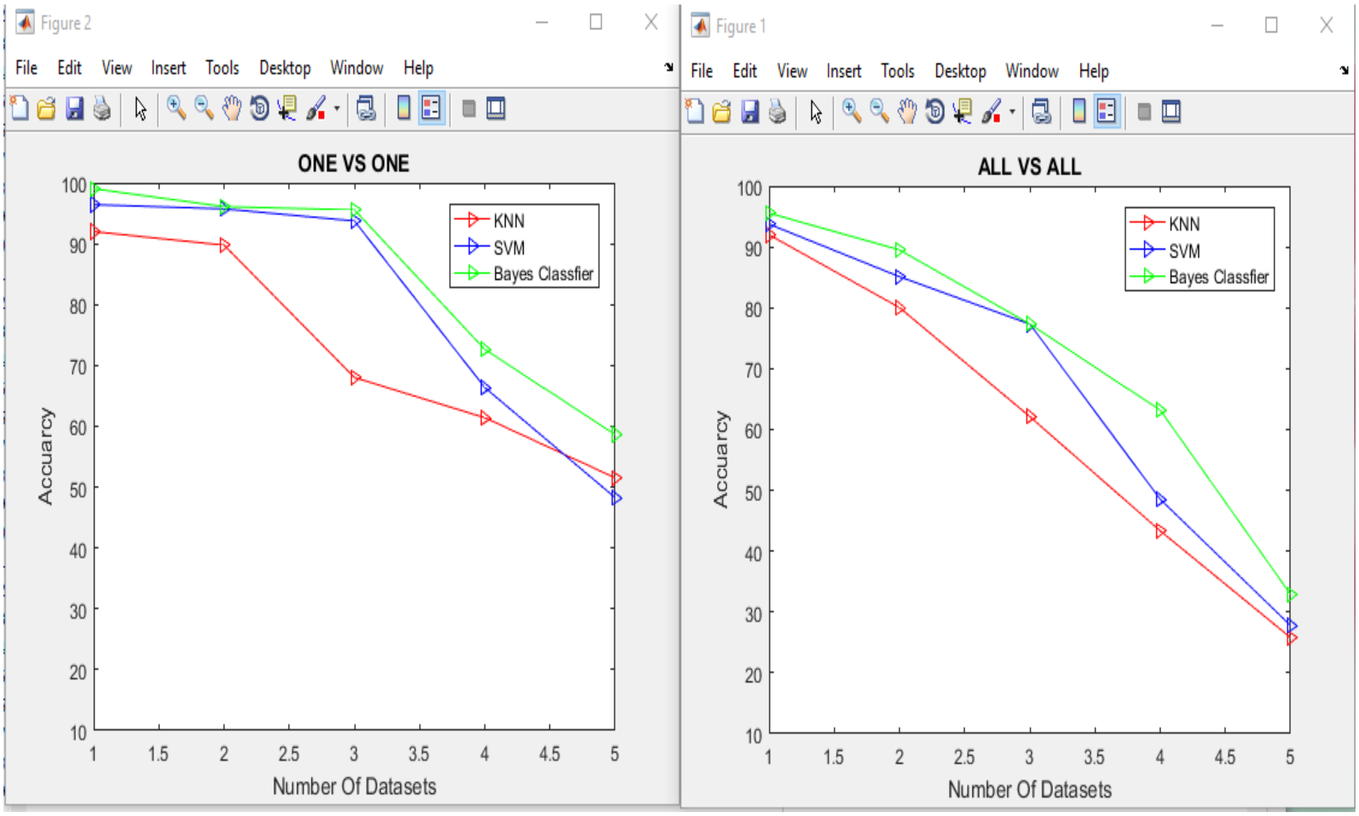

Fig. 29 Graphical representation shows the overall evaluation of the research work that includes 5 databases with 3 Classifiers.

Database are in order of

1. Fisher Iris
2. Wine
3. Forest
4. Diabetes
5. Breast Cancer

This comparison clearly shows that the Bayes type of Classifier give more accuracy and less misclassification rate when compared with SVM and KNN.

The Accuracy Percentage of all the datasets are more when using this type of classifier compared with SVM and KNN. Were as when we compare with SVM and KNN only on Breast Cancer data set KNN classifier gives more accuracy than SVM.

But on the overall work Bayes type of Classifier is the best used this Paper.

Overall output comparison table for datasets with its classifiers

	Breast cancer	Diabetes	Iris	Forest Fire	Wine
KNN OVO	51.5288%	69.75%	92%	68.0044%	88.1816%
KNN AVA	25.80%	43.35%	92%	80.04%	62.12%
SVM OVO	52.683%	66.691%	94%	95.28%	96.46%
SVM AVA	27.59%	49.06%	94%	76.94%	85%
Bayes OVO	58.7955%	72.625%	96%	96.063%	99.0741%
Bayes AVA	32.91%	63.19%	96%	77.35%	90%

Tab.16 Output on Tabular Representation

6. Conclusions & Future Improvement:

The goal of this method was to discover methods for building a generalized ensemble of classifiers . As we perform an empirical comparison of several multi classifier systems using several data sets those with different problems. Our experimental results shows that our new ensemble of classifier will show the outperform of best state-of-art standalone ensemble methods were we can clearly see that Bayes Classifier leads give more accuracy than SVM and KNN of out Datasets.

May be on practical work we might try to find the possible single method that is out performed when compared to other classifiers used in this method of data. This survey presented the different approaches employed to solve the problem of multiclass classification. The first approach relied on extending binary classification problems to handle the multiclass case directly. This included support vector machines, Bayes and k-nearest neighbors. The data's are been trained with different classifiers and the data's are ready for combination of classifiers with different other application techniques.

So here on this comparative study we say the best classifiers which would be Bayes classifier which has smallest misclassification rate when compared with the different classifiers on the datasets and also the combination method helps to improve the single classifiers result . Hence far more information that can be processed in this method to gain more accuracy.

7.References

1. Erin Allwein, Robert Shapire, and Yoram Singer. Reducing multiclass to binary: A unifying approach for margin classifiers. Journal of Machine Learning Research, pages 113–141, 2000.
2. Breiman, L., “Bagging Predictors,” Machine Learning, 24. 123-140. 1996.
3. http://imsecconference2014.files.wordpress.com/2014/02/unlicensed-submission49.pdf
4. Duo Ji, Huan-yu Zhao, Dong-feng Cai. Using the Multi-level Classification Method in the Patent Mining Task at NTCIR-7. In Proceedings of the 7th NTCIR Workshop Meeting. 2008.12. P362-364
5. Edwin Chen Math, stats, machine learning, human computation, visualization, data science. /2011/04/27/choosing-a-machine-learning-classifier/
6. N. Friedman, D. Geiger, and Goldszmidt M. Bayesian network classifiers. Machine Learning, 29:131–163, 1997.
7. Chapter Feature Extraction Volume 207 of the series Studies in Fuzziness and Soft Computing pp 343-358 Variable Selection using Correlation and Single Variable Classifier Methods: Applications Amir Reza Saffari Azar Alamdari
8. Advances in Pattern Recognition Volume 1876 of the series Lecture Notes in Computer Science pp 631-640 Date: 21 December 2000 The Role of Combining Rules in Bagging and Boosting Marina Skurichina , Robert P. W. Duin
9. Stephen D. Bay. Combining nearest neighbor classifiers through multiple feature subsets. In Proceedings of the 17th International Conference on Machine Learning, pages 37–45, Madison, WI, 1998.
10. Erin J. Bredensteiner and Kristin P. Bennett. Multicategory classification by support vector machines. Computational Optimization and Applications, 12:53–79, January 1999.
11. Corinna Cortes and Vladimir Vapnik. Support-vector networks. Machine Learning, pages 273–297, 1995.
12. http://archive.ics.uci.edu/ml/datasets/Wine
13. http://archive.ics.uci.edu/ml/datasets/Diabetes+130S+hospitals+for+years+1999 2008
14. http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Original)
15. http://archive.ics.uci.edu/ml/datasets/Iris
16. https://archive.ics.uci.edu/ml/datasets/Forest+Fires
17. S. Aeberhard, D. Coomans and O. leVel, "THE CLASSIFICATION PERFORMANCE OF RDA" Tech. Rep. no. 92-01, (1992), Dept. of Computer Science and Dept. of Mathematics and Statistics, James Cook University of North Queensland.
18. S.Aeberhard,D.Coomans and O. de Vel, Comparison of Classifiers in High Dimensional Settings, Tech. Rep. no. 92-02, (1992), Dept. of Computer Science and Dept. of Mathematics and Statistics, James Cook University of North Queensland.

19. Table 1 Beata Strack, Jonathan P. DeShazo, Chris Gennings, Juan L. Olmo, Sebastian Ventura, Krzysztof J. Cios, and John N. Clore, “Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records,” BioMed Research International, vol. 2014, Article ID 781670, 11 pages, 2014.

20. Beata Strack, Jonathan P. DeShazo, Chris Gennings, Juan L. Olmo, Sebastian Ventura, Krzysztof J. Cios, and John N. Clore, “Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records,” BioMed Research International, vol. 2014, Article ID 781670, 11 pages, 2014.
21. William H. Wolberg and O.L. Mangasarian: "Multisurface method of pattern separation for medical diagnosis applied to breast cytology", Proceedings of the National Academy of Sciences, U.S.A., Volume87, December 1990, pp 9193-9196.
22. T. G. Dietterich and G. Bakiri. Solving multiclass learning problems via error correcting output codes. Journal of Artificial Intelligence Research, 39:1–38, 1995.
23. Erin Allwein, Robert Shapire, and Yoram Singer. Reducing multiclass to binary: A unifying approach for margin classifiers. Journal of Machine Learning Research, pages 113–141, 2000.
24. Table 1“Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records,” BioMed Research International, vol. 2014, Article ID 781670, 11 pages, 2014.
25. http://www.mathworks.com/matlabcentral/fileexchange/28770-introduction-to-classification
26. Chapter Hybrid Artificial Intelligent Systems Volume 7209 of the series Lecture Notes in Computer Science pp 25-35 A First Study on Decomposition Strategies with Data with Class Noise Using Decision Trees José A. Sáez , Mikel Galar , Julián Luengo , Francisco Herrera.
27. Chapter Evolutionary Computation,Machine Learning and Data Mining in Bioinformatics Volume 4447 of the series Lecture Notes in Computer Science pp 47-56 One-Versus-One and One-Versus-All Multiclass SVM-RFE for Gene Selection in Cancer Classification Kai-Bo Duan , Jagath C. Rajapakse , Minh N. Nguyen
28. Chapter Multiple Classifier Systems Volume 2096 of the series Lecture Notes in Computer Science pp 399-408 Date: 22 June 2001 Application of the Evolutionary Algorithms for Classifier Selection in Multiple Classifier Systems with Majority Voting Dymitr Ruta Bogdan Gabrys