A Critical Comparative Study of Liver Patients from USA and INDIA: An Exploratory Analysis

Bendi Venkata Ramana[1], Prof. M. Surendra Prasad Babu[2], Prof. N. B. Venkateswarlu[3]

[1] Associate Professor, Dept.of IT, AITAM,
Tekkali, Andhra Pradesh-532201, India.

[2] Dept. of CS&SE, AUCE,
Visakhapatnam-530 003, A.P, India.,

[3] Dept. of CSE, AITAM, Tekkali,
Andhra Pradesh-532201, India.

Abstract

Recent research studies on liver diagnosis indicated difference in classification accuracy of various classifiers with different data sets. K-Nearest Neighbor classifier is observed to be giving best results with India liver patients' data set with all feature set combinations. Performance is better for the India Liver dataset compared to UCLA liver dataset with all the selected algorithms [1]. In order to envisage the reason for this difference, we propose to analyze the liver patients' populations of both USA and India. We have carried out extensive ANOVA, MANOVA analysis on these data sets to observe any significant difference among the groups. It has observed that liver patients of both the countries are having significant difference which is the reason for difference in classifiers performance. Results of this study are very important for the development of automatic medical diagnosis system and the need for its localization settings based on the geographical region.

Keywords: *ANOVA, MANOVA, Classification, Liver diagnosis*

1. Introduction

Two data sets were evaluated using analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA). First dataset is taken from University of California at Irvine (UCI) Machine Learning Repository [2], which contains 345 records with 6 attributes as shown in Table 1. Second dataset contains 583 liver patient records from north east, Andhra Pradesh, India with 10 attributes as shown in Table 2. Alkphos, SGPT and SGOT are the common attributes from the two data sets, they are taken for the purpose of comparison.

Group 1 indicates UCI data set and Group 2 indicates INDIA data set. In UCI data set, 145 patients are labelled as liver patients others are not. Similarly, in Indian data set 416 patients are labelled as liver patients and remaining as non-liver patients.

In this paper, Standard statistical methods One-way Analysis of Variance (ANOVA) and Multivariate Analysis of Variance (MANOVA) are applied to evaluate the significance between two populations for better classification [5]. One-way Analysis of Variance (ANOVA) is used to test the significant difference in a single dependent variable among two or more groups formed by a single independent or classification variable, whereas Multivariate Analysis of Variance (MANOVA) is used to test the significant difference in more than one dependent variable and several independent variables.

Two Liver patient datasets were used in this study, one is collected from Andhra Pradesh state of India and the second one is BUPA Liver Disorders datasets taken from University of California at Irvine (UCI) Machine Learning Repository [4]. The attributes of Indian data set were Age, Gender, Total_Bilirubin, Direct_Bilirubin, Alkphos, SGPT, SGOT, Total_Protiens, Albumin and A/G ratio. The attributes of UCI data set were Mcv, Alkphos, SGPT, SGOT and Gammagt. The common liver functional tests from both the data sets were Alkphos, SGPT and SGOT [1][2][3].

2. Related work

Mireille Tohm´et al [7] proposed an alternative to usual multiclass multivariate group comparison tests such as

Hypothesis tests are used to compare and show the efficiency of drugs. Junning Li et al.[8] proposed a Dynamic Bayesian Networks (DBN)-based group-analysis which combines the DBN approach and the multivariate analysis of variance (MANOVA). Neven Cukrov et al.[9] was applied multivariate statistical analysis to the measured physico-chemical parameters to estimate anthropogenic and natural influences to water system of the Krka River. Z. Haddi et al.[10] proposed Multivariate Analysis of Variance (MANOVA) to test the significance of the differences between cheeses groups. Z. A. Dastgheib et al. [11] applied multivariate analysis of variance (MANOVA) to select pairs of features showing the most significant differences between the groups to get more classifier accuracy. S. Dimitrova [12] conducted MANOVA to check the significance of the influence of three different factors namely 1 planetary gcomagnctic activity Ievel estimated by Ap-index and divided into five levels, 2. gender - males and females and 3. the presence of medication. Paulo Ricardo Galhanone et al. [13] applied MANOVA and Discriminate Analysis to Spectral analysis of the multichannel EEG of neonates is carried out with a view to determining differences in characteristics of High-Voltage-Slow, Low-Voltage-Irregular and Mixed EEG patterns. Diego Moitre, and Fernando Magnago [14] presented the application of the methodology of analysis of variance of multivariate data (MANOVA) to detect the impact of the fuel consumption on the market price. B.Surendiran et al. [15] proposed an Univariate Analysis of Variance (ANOVA) and Discriminate Analysis (DA) classifier for classifying the masses present in mammogram. Martha L. Zequera et al. [16] was designed to assess the effect of time on the repeatability of the LorAn pressure distribution measurement system, and evaluate the variability of plantar pressure and postural balance, during barefoot standing in diabetic and non-diabetic subjects, for future diabetic foot clinical evaluation. Benjamin F et al. [17] presented Directed canonical analysis as an extension of the general form of canonical analysis, which is a method for reducing the dimensionality of multivariate data sets with minimum loss of discriminatory variance. Aleksandar Jeremic et al. [18] developed a frequency-domain channel estimation algorithm for single-user multiantenna orthogonal frequency division multiplexing (OFDM) wireless systems in the presence of synchronous interference.

2. One way analysis of variance (ANOVA)

The F statistics obtained from ANOVA only tell us whether there is any significant difference in the mean values of the two groups. In this ALKPHOS, SGPT and SGOT were considered as dependent variables and Group was considered as factoring variable.

The results of ANOVA were represented in three rows.

1. Between Groups:- Between groups indicates the variability due to the place of data (between $\left(\overline{x}_i - \overline{x}\right)^2$ groups variability)
2. Within Groups:-With in groups indicates variability due to random error $\left(x_{ij} - \overline{x}_i\right)^2$
3. Total:- Indicates total variability

The ANOVA F-statistic is a ratio of the Between Group Variation divided by the Within Group Variation

3. Multivariate analysis of variance (MANOVA)

Multivariate analysis of variance is a way to test the hypothesis that one or more independent variables, or factors, have an effect on a set of two or more dependent variables. The goal of our analysis is to look for an effect of one or more IVs on several DVs at the same time. Four different multivariate tests were considered to identify the significant effect of the IVs on all of the DVs, as a group.

4. Results and Discussion

Our analysis includes population comparisons based on the common attributes, Alkphos, SGPT and SGOT and their combinations. Total we will have $3C_1 + 3C_2 + 3C_3$ combinations totaling 7 for experiment 1, experiment 2 and experiment 3.

Table 1: UCA Liver dataset and attributes available

Attribute	Type
Mcv	Integer
Alkphos	Integer
SGPT	Integer
SGOT	Integer
Gammagt	Real number
Drinks	Real number

Table 2: INDIA dataset and attributes

Attribute	Type
Gender	Categorical
Age	Real number
Total_bilirubin	Real number
Direct_ bilirubin	Real number
Total_protiens	Real number
Albumin	Real number
A/G ratio	Real number
SGPT	Integer
SGOT	Integer
Alkphos	Integer

Experiment 1

Experiment 1 includes the analysis of all Patients that means both liver and non liver patients of UCI and India

(Pooled analysis). UCI data set contains 345 patient records and INDIA data set contains 583 patient records. Total records are 928.

The analysis reported from Table 3 to Table 16 for the UCI & INDIA data sets for the liver data with both liver patients and non liver patients, Table 17 to Table 30 for the UCI & INDIA data sets for the liver data with only liver patients and Table 31 to Table 44 for the UCI & INDIA data sets for the liver data with only non liver patients.

Table 3, Table 5 & Table 7 shows descriptive statistics that are no of records, mean standard deviation, standard error etc. for the individual attributes ALKPHOS, SGPT and SGOT respectively.

Table 4, Table 6 & Table 8 shows one way analysis of variance for the attributes ALKPHOS, SGPT and SGOT respectively. The results reported in Table 4, Table 6 & Table 8 indicates the significant difference between groups of data sets.

P-value (significance value) in table 4 Indicates the probability of getting a mean difference between the groups as high as what is observed by chance. The lower the p-value, the more significant the difference between the groups. The p-value in table 4 is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on ALKPHOS.

Significant value in table 6 is 0.000 which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on SGPT.

Significant value in table 8 is 0.000 which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on SGOT.

Table 3:Descriptive Statistics of ALKPHOS

ALKPHOS 3								
Group	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	345	69.87	18.348	.988	67.93	71.81	23	138
2	583	290.58	242.938	10.061	270.82	310.34	63	2110
Total	928	208.52	220.381	7.234	194.33	222.72	23	2110

Table 4: One Way ANOVA on ALKPHOS between UCI & INDIA datasets

ALKPHOS					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10557739.946	1	10557739.946	283.665	.000
Within Groups	34464783.484	926	37218.989		
Total	45022523.430	927			

Table 5: Descriptive Statistics of SGPT

SGPT								
Group	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	345	30.41	19.512	1.051	28.34	32.47	4	155
2	583	80.71	182.620	7.563	65.86	95.57	10	2000
Total	928	62.01	147.212	4.832	52.53	71.49	4	2000

Table 6: ANOVA on SGPT between UCI & INDIA datasets

SGPT					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	548541.541	1	548541.541	25.994	.000
Within Groups	19540784.351	926	21102.359		
Total	20089325.892	927			

Table 7:Descriptive Statistics of SGOT

SGOT								
Group	N	Mean	Standard Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	345	24.64	10.064	.542	23.58	25.71	5	82
2	583	109.91	288.919	11.966	86.41	133.41	10	4929
Total	928	78.21	232.691	7.638	63.22	93.20	5	4929

Table 8: ANOVA on SGOT between UCA & INDIA datasets

SGOT					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1575814.094	1	1575814.094	30.014	.000
Within Groups	48616664.510	926	52501.798		
Total	50192478.603	927			

Table 9, Table 11, Table 13 & Table 15 shows the descriptive statistics for the combination of attributes ALKPHOS, SGPT, ALKPHOS, SGOT, SGPT, SGOT and ALKPHOS, SGPT ,SGOT respectively.

The results reported in Table 10, Table 12, Table 14 & Table 16 are the four different multivariate tests and their significant values(p) for the combination of attributes ALKPHOS, SGPT , ALKPHOS, SGOT , SGPT, SGOT and ALKPHOS, SGPT ,SGOT respectively.

Table 9:Descriptive Statistics of ALKPHOS & SGPT

	GROUP	Mean	Std. Deviation	N
ALKPHOS	1	69.87	18.348	345
	2	290.58	242.938	583
	Total	208.52	220.381	928
SGPT	1	30.41	19.512	345
	2	80.71	182.620	583
	Total	62.01	147.212	928

Table 10: Multivariate Tests[a] on ALKPHOS & SGPT between UCA & INDIA datasets

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power
Intercept	Pillai's Trace	.469	408.849b	2.000	925.000	.000	.469	817.698b	1.000
	Wilks' Lambda	.531	408.849b	2.000	925.000	.000	.469	817.698b	1.000
	Hotelling's Trace	.884	408.849b	2.000	925.000	.000	.469	817.698b	1.000
	Roy's Largest Root	.884	408.849b	2.000	925.000	.000	.469	817.698b	1.000
GROUP	Pillai's Trace	.240	146.205b	2.000	925.000	.000	.240	292.410b	1.000
	Wilks' Lambda	.760	146.205b	2.000	925.000	.000	.240	292.410b	1.000
	Hotelling's Trace	.316	146.205b	2.000	925.000	.000	.240	292.410b	1.000
	Roy's Largest Root	.316	146.205b	2.000	925.000	.000	.240	292.410b	1.000

Significant value in table 10 is 0.000 which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on ALKPHOS and SGPT.

Table 11:Descriptive Statistics of ALKPHOS & SGOT

	GROUP	Mean	Std. Deviation	N
ALKPHOS	1	69.87	18.348	345
	2	290.58	242.938	583
	Total	208.52	220.381	928
SGOT	1	24.64	10.064	345
	2	109.91	288.919	583
	Total	78.21	232.691	928

Table 12: Multivariate Tests[a] on ALKPHOS & SGOT between UCA & INDIA datasets

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power
Intercept	Pillai's Trace	.455	386.308b	2.000	925.000	.000	.455	772.615b	1.000
	Wilks' Lambda	.545	386.308b	2.000	925.000	.000	.455	772.615b	1.000
	Hotelling's Trace	.835	386.308b	2.000	925.000	.000	.455	772.615b	1.000
	Roy's Largest Root	.835	386.308b	2.000	925.000	.000	.455	772.615b	1.000
GROUP	Pillai's Trace	.239	145.327b	2.000	925.000	.000	.239	290.655b	1.000
	Wilks' Lambda	.761	145.327b	2.000	925.000	.000	.239	290.655b	1.000
	Hotelling's Trace	.314	145.327b	2.000	925.000	.000	.239	290.655b	1.000
	Roy's Largest Root	.314	145.327b	2.000	925.000	.000	.239	290.655b	1.000

Significant value in table 12 is 0.000 which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on ALKPHOS and SGOT.

Significant value in table 14 is 0.000 which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on SGOT and SGPT.

Table 13:Descriptive Statistics of SGPT & SGOT

	GROUP	Mean	Std. Deviation	N
SGOT	1	24.64	10.064	345
	2	109.91	288.919	583
	Total	78.21	232.691	928
SGPT	1	30.41	19.512	345
	2	80.71	182.620	583
	Total	62.01	147.212	928

Table 14: Multivariate Tests[a] on SGPT & SGOT between UCA & INDIA datasets

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power
Intercept	Pillai's Trace	.121	63.431b	2.000	925.000	.000	.121	126.861b	1.000
	Wilks' Lambda	.879	63.431b	2.000	925.000	.000	.121	126.861b	1.000
	Hotelling's Trace	.137	63.431b	2.000	925.000	.000	.121	126.861b	1.000
	Roy's Largest Root	.137	63.431b	2.000	925.000	.000	.121	126.861b	1.000
GROUP	Pillai's Trace	.033	15.775b	2.000	925.000	.000	.033	31.549b	1.000
	Wilks' Lambda	.967	15.775b	2.000	925.000	.000	.033	31.549b	1.000
	Hotelling's Trace	.034	15.775b	2.000	925.000	.000	.033	31.549b	1.000
	Roy's Largest Root	.034	15.775b	2.000	925.000	.000	.033	31.549b	1.000

Table 15:Descriptive Statistics of ALKPHOS, SGPT & SGOT

	GROUP	Mean	Std. Deviation	N
SGOT	1	24.64	10.064	345
	2	109.91	288.919	583
	Total	78.21	232.691	928
SGPT	1	30.41	19.512	345
	2	80.71	182.620	583
	Total	62.01	147.212	928
ALKPHOS	1	69.87	18.348	345
	2	290.58	242.938	583
	Total	208.52	220.381	928

Significant value in table 16 that is multivariate analysis on ALKPHOS, SGPT and SGOT is 0.000 which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on ALKPHOS, SGPT and SGOT.

The significant values are less than 0.05 ($p < 0.05$) for four different multivariate tests for all the combination of attributes. This indicates that there is a significant effect of the independent variables on all of the dependent variables considered as a group.

Table 16: Multivariate Tests[a] on ALKPHOS, SGPT & SGOT between UCA & INDIA datasets

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power
Intercept	Pillai's Trace	.473	276.082b	3.000	924.000	.000	.473	828.245b	1.000
	Wilks' Lambda	.527	276.082b	3.000	924.000	.000	.473	828.245b	1.000
	Hotelling's Trace	.896	276.082b	3.000	924.000	.000	.473	828.245b	1.000
	Roy's Largest Root	.896	276.082b	3.000	924.000	.000	.473	828.245b	1.000
GROUP	Pillai's Trace	.240	97.462b	3.000	924.000	.000	.240	292.386b	1.000
	Wilks' Lambda	.760	97.462b	3.000	924.000	.000	.240	292.386b	1.000
	Hotelling's Trace	.316	97.462b	3.000	924.000	.000	.240	292.386b	1.000
	Roy's Largest Root	.316	97.462b	3.000	924.000	.000	.473	828.245b	1.000

Experiment 2

Experiment 2 includes the analysis of liver Patients of UCI and India. UCI data set contains 145 liver patient records and INDIA data set contains 416 liver patient records. Total records are 561.

Table 17, Table 19 & Table 21 shows descriptive statistics that are no of records, mean standard deviation, standard error etc. for the individual attributes ALKPHOS, SGPT and SGOT respectively.

Table 18, Table 20 & Table 22 shows one way analysis of variance for the attributes ALKPHOS, SGPT and SGOT respectively. The results reported in Table 18,

Table 20 & Table 22 indicates the significant difference between groups of data sets.

Significant value in table 18 is 0.000 which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on ALKPHOS.

Significant value in table 20 is 0.000 which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on SGPT.

Table 17:Descriptive Statistics of ALKPHOS

ALKPHOS								
Group	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	145	71.98	18.591	1.544	68.93	75.03	23	138
2	416	319.01	268.308	13.155	293.15	344.87	63	2110
Total	561	255.16	255.254	10.777	233.99	276.33	23	2110

Table 18: ANOVA on ALKPHOS between UCI & INDIA datasets

ALKPHOS					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6561308.964	1	6561308.964	122.564	.000
Within Groups	29925259.916	559	53533.560		
Total	36486568.881	560			

Table 19:Descriptive Statistics of SGPT

SGPT								
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	145	31.21	15.778	1.310	28.62	33.80	10	103
2	416	99.61	212.768	10.432	79.10	120.11	12	2000
Total	561	81.93	185.771	7.843	66.52	97.33	10	2000

Table 20: ANOVA on SGPT between UCI & INDIA datasets

SGPT					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	503032.864	1	503032.864	14.939	.000
Within Groups	18823073.139	559	33672.761		
Total	19326106.004	560			

Table 21:Descriptive Statistics of SGOT

SGOT								
Group	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	145	22.79	7.738	.643	21.52	24.06	5	57
2	416	137.70	337.390	16.542	105.18	170.22	11	4929
Total	561	108.00	294.802	12.447	83.55	132.45	5	4929

Table 22: ANOVA on SGOT between UCI & INDIA datasets

SGOT					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1419839.186	1	1419839.186	16.798	.000
Within Groups	47248901.812	559	84523.975		
Total	48668740.998	560			

Significant value in table 22 is 0.000 which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that population differ a lot on on SGOT.

Table 23, Table 25, Table 27 & Table 29 shows the descriptive statistics for the combination of attributes ALKPHOS, SGPT , ALKPHOS, SGOT , SGPT, SGOT and ALKPHOS, SGPT ,SGOT respectively. The results reported in Table 24, Table 26, Table 28 & Table 30 are the four different multivariate tests and their significant values (p) for the combination of attributes ALKPHOS, SGPT , ALKPHOS, SGOT , SGPT, SGOT and ALKPHOS, SGPT ,SGOT respectively.

Table 23:Descriptive Statistics of ALKPHOS & SGPT

	GROUP	Mean	Std. Deviation	N
ALKPHOS	1	71.98	18.591	145
	2	319.01	268.308	416
	Total	255.16	255.254	561
SGPT	1	31.21	15.778	145
	2	99.61	212.768	416
	Total	81.93	185.771	561

Table 24: Multivariate Tests[a] on ALKPHOS & SGPT between UCA & INDIA datasets

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent Parameter	Observed Power
Intercept	Pillai's Trace	.378	169.812b	2.000	558.000	.000	.378	339.623b	1.000
	Wilks' Lambda	.622	169.812b	2.000	558.000	.000	.378	339.623b	1.000
	Hotelling's Trace	.609	169.812b	2.000	558.000	.000	.378	339.623b	1.000
	Roy's Largest Root	.609	169.812b	2.000	558.000	.000	.378	339.623b	1.000
GROUP	Pillai's Trace	.189	65.173b	2.000	558.000	.000	.189	130.346b	1.000
	Wilks' Lambda	.811	65.173b	2.000	558.000	.000			
	Hotelling's Trace	.234	65.173b	2.000	558.000	.000			
	Roy's Largest Root	.234	65.173b	2.000	558.000	.000			

Significant value in table 24 is 0.000 which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on ALKPHOS and SGPT.

Table 25:Descriptive Statistics of ALKPHOS & SGOT

	GROUP	Mean	Std. Deviation	N
ALKPHOS	1	71.98	18.591	145
	2	319.01	268.308	416
	Total	255.16	255.254	561
SGOT	1	22.79	7.738	145
	2	137.70	337.390	416
	Total	108.00	294.802	561

Table 26: Multivariate Tests[a] on ALKPHOS & SGOT between UCA & INDIA datasets

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power
Intercept	Pillai's Trace	.362	158.401b	2.000	558.000	.000	.362	316.802b	1.000
	Wilks' Lambda	.638	158.401b	2.000	558.000	.000	.362	316.802b	1.000
	Hotelling's Trace	.568	158.401b	2.000	558.000	.000	.362	316.802b	1.000
	Roy's Largest Root	.568	158.401b	2.000	558.000	.000	.362	316.802b	1.000
GROUP	Pillai's Trace	.187	64.337b	2.000	558.000	.000	.187	128.673b	1.000
	Wilks' Lambda	.813	64.337b	2.000	558.000	.000	.187	128.673b	1.000
	Hotelling's Trace	.231	64.337b	2.000	558.000	.000	.187	128.673b	1.000
	Roy's Largest Root	.231	64.337b	2.000	558.000	.000	.187	128.673b	1.000

Significant value in table 26 is 0.000 which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on ALKPHOS and SGOT.

Table 27:Descriptive Statistics of SGOT & SGPT

	GROUP	Mean	Std. Deviation	N
SGOT	1	22.79	7.738	145
	2	137.70	337.390	416
	Total	108.00	294.802	561
SGPT	1	31.21	15.778	145
	2	99.61	212.768	416
	Total	81.93	185.771	561

Table 28: Multivariate Tests[a] on SGOT & SGPT between UCA & INDIA datasets

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power
Intercept	Pillai's Trace	.089	27.283b	2.000	558.000	.000	.089	54.566b	1.000
	Wilks' Lambda	.911	27.283b	2.000	558.000	.000	.089	54.566b	1.000
	Hotelling's Trace	.098	27.283b	2.000	558.000	.000	.089	54.566b	1.000
	Roy's Largest Root	.098	27.283b	2.000	558.000	.000	.089	54.566b	1.000
GROUP	Pillai's Trace	.031	8.921b	2.000	558.000	.000	.031	17.841b	.973
	Wilks' Lambda	.969	8.921b	2.000	558.000	.000	.031	17.841b	.973
	Hotelling's Trace	.032	8.921b	2.000	558.000	.000	.031	17.841b	.973
	Roy's Largest Root	.032	8.921b	2.000	558.000	.000	.031	17.841b	.973

Significant value in table 28 is 0.000 which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on SGPT and SGOT.

Significant value is 0.000 in table 30 that is multivariate analysis on ALKPHOS, SGPT and SGOT is which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on ALKPHOS, SGPT and SGOT.

Table 29:Descriptive Statistics of ALKPHOS, SGPT & SGOT

	GROUP	Mean	Std. Deviation	N
SGOT	1	22.79	7.738	145
	2	137.70	337.390	416
	Total	108.00	294.802	561
SGPT	1	31.21	15.778	145
	2	99.61	212.768	416
	Total	81.93	185.771	561
ALKPHOS	1	71.98	18.591	145
	2	319.01	268.308	416
	Total	255.16	255.254	561

Table 30: Multivariate Tests[a] on ALKPHOS, SGPT & SGOT between UCA & INDIA datasets

Effect 30		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power
Intercept	Pillai's Trace	.381	114.487b	3.000	557.000	.000	.381	343.461b	1.000
	Wilks' Lambda	.619	114.487b	3.000	557.000	.000	.381	343.461b	1.000
	Hotelling's Trace	.617	114.487b	3.000	557.000	.000	.381	343.461b	1.000
	Roy's Largest Root	.617	114.487b	3.000	557.000	.000	.381	343.461b	1.000
GROUP	Pillai's Trace	.190	43.446b	3.000	557.000	.000	.190	130.339b	1.000
	Wilks' Lambda	.810	43.446b	3.000	557.000	.000	.190	130.339b	1.000
	Hotelling's Trace	.234	43.446b	3.000	557.000	.000	.190	130.339b	1.000
	Roy's Largest Root	.234	43.446b	3.000	557.000	.000	.190	130.339b	1.000

Experiment 3

Experiment 3 includes the analysis of non liver Patients of UCI and India. UCI data set contains 200 non liver patient records and INDIA data set contains 167 non liver patient records. Total records are 367.

Table 31 Table 33 & Table 35 shows descriptive statistics that are no of records, mean standard deviation, standard error etc. for the individual attributes ALKPHOS, SGPT and SGOT respectively.

Table 32, Table 34 & Table 36 shows one way analysis of variance for the attributes ALKPHOS, SGPT and SGOT respectively. The results reported in Table 32, Table 34 & Table 36 indicates the significant difference between groups of data sets.

Table 31:Descriptive Statistics of ALKPHOS

ALKPHOS								
Group	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	200	68.34	18.062	1.277	65.82	70.86	37	134
2	167	219.75	140.986	10.910	198.21	241.29	90	1580
Total	367	137.24	122.039	6.370	124.71	149.77	37	1580

Table 32: One Way ANOVA on ALKPHOS between UCI & INDIA datasets

ALKPHOS					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2086485.085	1	2086485.085	226.352	.000
Within Groups	3364523.814	365	9217.873		
Total	5451008.899	366			

Table 33:Descriptive Statistics of SGPT

SGPT								
Group	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	200	29.83	21.845	1.545	26.78	32.87	4	155
2	167	33.65	25.060	1.939	29.82	37.48	10	181
Total	367	31.57	23.408	1.222	29.16	33.97	4	181

Table 34: One Way ANOVA on SGPT between UCI & INDIA datasets

SGPT					
	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	1333.383	1	1333.383	2.443	.119
Within Groups	199214.731	365	545.794		
Total	200548.114	366			

Significant value in table 32 is 0.000 which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on ALKPHOS.

Significant value in table 34 is 0.000 which is greater than 0.119 ($p > 0.05$) can accept the null hypothesis that indicates there is no significant difference between groups. Then we can say that there is no populations differ on SGPT.

Significant value in table 36 is 0.000 which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on SGOT.

Table 37, Table 39, Table 41 & Table 43 shows the descriptive statistics for the combination of attributes ALKPHOS, SGPT , ALKPHOS, SGOT , SGPT, SGOT and ALKPHOS, SGPT ,SGOT respectively.

Table 35:Descriptive Statistics of SGOT

SGOT								
Group	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	200	25.99	11.289	.798	24.42	27.56	8	82
2	167	40.69	36.412	2.818	35.13	46.25	10	285
Total	367	32.68	26.913	1.405	29.92	35.44	8	285

Table 36: One Way ANOVA on SGOT between UCI & INDIA datasets

SGOT 36					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	19662.272	1	19662.272	29.240	.000
Within Groups	245443.788	365	672.449		
Total	265106.060	366			

Table 37:Descriptive Statistics of ALKPHOS & SGPT

	GROUP	Mean	Std. Deviation	N
ALKPHOS	1	68.34	18.062	200
	2	219.75	140.986	167
	Total	137.24	122.039	367
SGPT	1	29.82	21.845	200
	2	33.65	25.060	167
	Total	31.57	23.408	367

The results reported in Table 38, Table 40, Table 42 and Table 44 are the four different multivariate tests and their significant values(p) for the combination of attributes ALKPHOS, SGPT , ALKPHOS, SGOT , SGPT, SGOT and ALKPHOS, SGPT ,SGOT respectively.

Table 38: Multivariate Tests[a] on ALKPHOS & SGPT between UCA & INDIA datasets

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power
Intercept	Pillai's Trace	.762	581.503b	2.000	364.000	.000	.762	1163.006b	1.000
	Wilks' Lambda	.238	581.503b	2.000	364.000	.000	.762	1163.006b	1.000
	Hotelling's Trace	3.195	581.503b	2.000	364.000	.000	.762	1163.006b	1.000
	Roy's Largest Root	3.195	581.503b	2.000	364.000	.000	.762	1163.006b	1.000
GROUP	Pillai's Trace	.391	116.721b	2.000	364.000	.000	.391	233.442b	1.000
	Wilks' Lambda	.609	116.721b	2.000	364.000	.000	.391	233.442b	1.000
	Hotelling's Trace	.641	116.721b	2.000	364.000	.000	.391	233.442b	1.000
	Roy's Largest Root	.641	116.721b	2.000	364.000	.000	.391	233.442b	1.000

Significant value in table 38 is 0.000 that is multivariate analysis on ALKPHOS and SGPT which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on ALKPHOS and SGPT.

Significant value in table 40 is 0.000 that is multivariate analysis on ALKPHOS and SGOT which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on ALKPHOS and SGOT.

Table 39:Descriptive Statistics of ALKPHOS & SGOT

	GROUP	Mean	Std. Deviation	N
ALKPHOS	1	68.34	18.062	200
	2	219.75	140.986	167
	Total	137.24	122.039	367
SGOT	1	25.99	11.289	200
	2	40.69	36.412	167
	Total	32.68	26.913	367

Table 40: Multivariate Tests[a] on ALKPHOS & SGOT between UCA & INDIA datasets

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power
Intercept	Pillai's Trace	.757	566.093b	2.000	364.000	.000	.757	1132.186b	1.000
	Wilks' Lambda	.243	566.093b	2.000	364.000	.000	.757	1132.186b	1.000
	Hotelling's Trace	3.110	566.093b	2.000	364.000	.000	.757	1132.186b	1.000
	Roy's Largest Root	3.110	566.093b	2.000	364.000	.000	.757	1132.186b	1.000
GROUP	Pillai's Trace	.385	114.136b	2.000	364.000	.000	.385	228.271b	1.000
	Wilks' Lambda	.615	114.136b	2.000	364.000	.000	.385	228.271b	1.000
	Hotelling's Trace	.627	114.136b	2.000	364.000	.000	.385	228.271b	1.000
	Roy's Largest Root	.627	114.136b	2.000	364.000	.000	.385	228.271b	1.000

Significant value in table 42 is 0.000 that is multivariate analysis on SGPT and SGOT which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on SGPT and SGOT.

Significant value in table 44 is 0.000 that is multivariate analysis on ALKPHOS, SGPT and SGOT which is less than 0.05 ($p < 0.05$) can safely reject the null hypothesis that indicates there is more significant difference between groups. Then we can say that populations differ a lot on ALKPHOS, SGPT and SGOT.

Table 41:Descriptive Statistics of SGPT& SGOT

	GROUP	Mean	Std. Deviation	N
SGOT	1	25.99	11.289	200
	2	40.69	36.412	167
	Total	32.68	26.913	367
SGPT	1	29.82	21.845	200
	2	33.65	25.060	167
	Total	31.57	23.408	367

All of our tables are related to 95 % significant levels. We did investigate with 99 % and 90 % significant levels also. They also supports the groups are different in all the three experiments.

Table 42: Multivariate Tests[a] on SGPT & SGOT between UCA & INDIA datasets

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power
Intercept	Pillai's Trace	.682	391.098b	2.000	364.000	.000	.682	782.195b	1.000
	Wilks' Lambda	.318	391.098b	2.000	364.000	.000	.682	782.195b	1.000
	Hotelling's Trace	2.149	391.098b	2.000	364.000	.000	.682	782.195b	1.000
	Roy's Largest Root	2.149	391.098b	2.000	364.000	.000	.682	782.195b	1.000
GROUP	Pillai's Trace	.087	17.344b	2.000	364.000	.000	.087	34.689b	1.000
	Wilks' Lambda	.913	17.344b	2.000	364.000	.000	.087	34.689b	1.000
	Hotelling's Trace	.095	17.344b	2.000	364.000	.000	.087	34.689b	1.000
	Roy's Largest Root	.095	17.344b	2.000	364.000	.000	.087	34.689b	1.000

This study confirms the difference in liver patients of USA and India. Results of this study is very important while developing automatic medical diagnosis systems as it corroborates the necessity of localization of the software based on the geographical region. Also, liver specialists to be aware about these geographical differences among liver patients and prescribe any drugs accordingly.

Table 43:Descriptive Statistics of ALKPHOS, SGPT & SGOT

	GROUP	Mean	Std. Deviation	N
SGOT	1	25.99	11.289	200
	2	40.69	36.412	167
	Total	32.68	26.913	367
SGPT	1	29.82	21.845	200
	2	33.65	25.060	167
	Total	31.57	23.408	367
ALKPHOS	1	68.34	18.062	200
	2	219.75	140.986	167
	Total	137.24	122.039	367

Table 44: Multivariate Tests[a] on ALKPHOS, SGPT & SGOT between UCA & INDIA datasets

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power
Intercept	Pillai's Trace	.772	409.233b	3.000	363.000	.000	.772	1227.699b	1.000
	Wilks' Lambda	.228	409.233b	3.000	363.000	.000	.772	1227.699b	1.000
	Hotelling's Trace	3.382	409.233b	3.000	363.000	.000	.772	1227.699b	1.000
	Roy's Largest Root	3.382	409.233b	3.000	363.000	.000	.772	1227.699b	1.000
GROUP	Pillai's Trace	.407	83.103b	3.000	363.000	.000	.407	249.308b	1.000
	Wilks' Lambda	.593	83.103b	3.000	363.000	.000	.407	249.308b	1.000
	Hotelling's Trace	.687	83.103b	3.000	363.000	.000	.407	249.308b	1.000
	Roy's Largest Root	.687	83.103b	3.000	363.000	.000	.407	249.308b	1.000

5. Conclusions

In this study, the common attributes of the two data sets ALKPHOS, SGPT and SGOT are taken for One-way Analysis of Variance (ANOVA) and Multivariate Analysis of Variance (MANOVA). The analysis on data sets are in three ways. Experiment 1 shows that there exists more significant difference in the groups with all the possible attribute combinations. Experiment 2 also shows that there exists more significant difference in the groups with all the possible attribute combinations. Experiment 3 shows

that there exists more significant difference in the groups with all the possible attribute combinations except analysis on SGPT between non liver patients of UCI and INDIA data sets that indicates there is no significance difference between groups on SGPT for non liver patients of USA and INDIA.

6. Acknowledgements

The authors would like to thank Dr. S. Sridhar, Dr. Bevera Lakshmana Rao, K. Ravi Kanth and B. Sai Prasad for providing the data.

7. References

[1].Bendi Venkata Ramana, Prof. M.Surendra Prasad Babu and Prof. N. B. Venkateswarlu:" A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis". In Proceedings of the International Journal of Database Management Systems (IJDMS), Vol.3, No.2, pages 101- 114, May 2011.

[2]. Bendi Venkata Ramana, Prof. M.Surendra Prasad Babu and Prof. N. B. Venkateswarlu:" "A Critical Evaluation of Bayesian Classifier For Liver Diagnosis Using Bagging and Boosting Methods", International Journal of Engineering Science and Technology (IJEST), Vol.3, No.4, pages 3422-3426, April 2011.

[3]. Bendi Venkata Ramana, Prof. M. S. Prasad Babu and B. R. Sarath Kumar: "New Automatic Diagnosis of Liver Status Using Bayesian Classification", IEEE International Conference on Intelligent Network and Computing (ICINC 2010), pages V2-385-V2-388, Kuala Lumpur, Malaysia, 26-29 Nov 2010.

[4]. BUPA Liver Disorders Dataset. UCI repository of machine learning databases. Available from ftp://ftp.ics.uci.edu/pub/machine-learningdatabases/ liverdisorders/bupa.data, last accessed: 07 October 2010.

[5]. 2nd edition "Statistical Methods for Practice and Research" A guide to data analysis using SPSS by Ajai S Gaur

[6].16th Edition HARRISON'S PRINCIPLES of Internal Medicine

[7]. Mireille Tohm´e, R´egis Lengell´e and Virginie Freytag:" A multi class multivariate Mireille group comparison test: Application to drug safety". In Proceedings of the 32nd IEEE International Conference on EMBS, 2006, pages 4711-4714, September 4, 2011.

[8]. Junning Li, Z. Jane Wang and Martin J. McKeown: "A Framework for Group Analysis of fMRI Data using Dynamic Bayesian Networks". In Proceedings of the 29th IEEE International Conference on EMBS pages 5991-5994, August 2007. [3].

[9]. Neven Cukrov, Nataša Tepi, Dario Omanovi, Sonja Lojen, Elvira Bura-Naki,Vjerocka Vojvodi and Ivanka Pižeta : "Anthropogenic and Natural Influences on the Krka River (Croatia) Evaluated by Multivariate Statistical Analysis". In Proceedings of the 31st IEEE International Conference on Information Technology Interfaces pages 219-224, June 2009.

[10]. Z. Haddi, F. Annanouch, A. Amari, A. Hadoune, B.Bouchikhi and N. El Bari: "Application of a Portable Electronic Nose Device to Discriminate and Identify Cheeses with Known Percentages of Cow's and Goat's Milk". In Proceedings of the IEEE International Conference on Sensors, pages 771-774, 2010.

[11]. Z. A. Dastgheib, B. Lithgowand Z. Moussavi: "Application of Fractal Dimension on Vestibular Response Signals for Diagnosis of Parkinson's Disease". In Proceedings of the 33rd IEEE International Conference on EMBS, pages 7892-7895, September 2011.

[12]. S. Dimitrova: "Investigations of Some Human Physiological Parameters in Relation to Geomagnetic Variations of Solar Origin and Meteorological Factors". In Proceedings of the 2nd IEEE International Conference on Recent Advances in Space Technologies, pages 728-733, 2005.

[13]. Paulo Ricardo Galhanone, David Martin Simpson, Antonio Fernando C. Infantosi Eduardo Faveret, Maria Alice Genofre, Helio Bello and Leonard de Azevedo : "Multivariate Analysis of Neonatal EEG in Different Sleep Stages:Methods and Preliminary ResuIts". In Proceedings of the 17th IEEE International Conference on Engineering in Medicine and Biology Society, vol 2, pages 1021-1022, 1995.

[14]. Diego Moitre, and Fernando Magnago : "Multivariate Analysis of Variance Applied to Competitive Electricity Markets: The Fixed Effects Model ". In Proceedings of the IEEE International Conference on Power Engineering Society General Meeting, pages 1-7, 2007.

[15]. B.Surendiran, Y.Sundaraiah, A.Vadivel: "Classifying Digital Mammogram Masses using Univariate ANOVA Discriminant Analysis ". In Proceedings of the IEEE International Conference on Advances in Recent Technologies in Communication and Computing , pages 175-177, 2009.

[16]. Martha L. Zequera, Leonardo Garavito, William Sandham, Jorge A. Alvarado, Ángela odríguez, Carlos A. Wilches, Ana C. Villa, Shirley V. Quintero and Juan C. Bernal: "Assessment of the effect of time in the repeatability of the stabilometric parameters in diabetic and non-diabetic subjects during bipedal standing using the LorAn pressure distribution measurement system". In Proceedings of the 33rd IEEE International Conference on EMBS, pages 8531-8534, September 2011.

[17]. Benjamin F. Merembeck and Brian J. Turner: "Directed Canonical Analysis and the Performance of Classifiers Under Its Associated Linear Transformation". In Proceedings of the IEEE Transactions on Geoscience and Remote Sensing, Vol. Ge-18, No. 2, pages 190-196, April 1980.

[18]. IBM Statistical Package for the Social Sciences (SPSS) 20.00: IBM® SPSS® Statistics.

[19]. Aleksandar Jeremic, Timothy A. Thomas, and Arye Nehorai: "OFDM Channel Estimation in the Presence of Interference". In Proceedings of the IEEE Transactions on Signal Processing, vol 52, pages 3429-3439, 2004.