Critical Analysis of Indian Liver Patients Dataset using ANOVA Method

K.Swapna[1], Prof. M.S. Prasad Babu[2]
[1]*Research Scholar, Department of Computer Science and Systems Engineering, AU College of Engineering (A), Visakhapatnam, INDIA. E-mail: pentaswapna@yahoo.co.in*
[2]*Professor, Department of Computer Science and Systems Engineering, AU College of Engineering (A), Andhra University, Visakhapatnam, INDIA.E-mail: profmspbabu@gmail.com*

***Abstract* - Automatic Liver Diagnosis System helps liver patients to detect their diseases in an early stage so that they can take appropriate treatments for their healthy life. During the author's recent studies on clustering and classification algorithms for liver patient datasets, it has been observed that there is a difference in liver function test values regarding age and gender. Hence, the authors turn their attention to analyze the liver patient dataset with respect to gender and age population. The authors collected 500 records of liver patient data from pathological laboratories in southern India and clubbed their earlier collection of 583 records which are kept as ILPD dataset in UCI Machine learning repository. The aim of this paper is to enhance the ILPD dataset with a total of 1083 records and to conduct one way Analysis of Variance (ANOVA) method to examine the significant differences amongst the two groups. It is found that there are significant differences in the ranges of values with respect to the gender and age attributes.**

***Index Term*-- Gender differences, Age differences, ANOVA, Liver Datasets**

I. INTRODUCTION

Liver disease patients are growing year by year for various reasons like drinking of contaminated water and food, irregular eating habits, consumption of pickles, alcoholism, infections, usage of medicines for longer duration, stress etc. I.Liver diseases vary with respect to geographical location and various factors such as socio-economic status, life style, and other endemic diseases. Most of the liver diseases, even in advance stages, may not show prominent clinical symptoms and are diagnosed or found incidentally during general checkups, investigations for other diseases [2]. Automatic Liver Diagnosis System enables the Gastroenterologists to detect liver diseases accurately which helps to reduce the increasing deaths of liver patients. The present analysis helps to identify the potential liver disease patients whose blood tests/ attributes lie in the ranges of values as found in this analysis. In this study first, the datasets is pre –processed then applied the standard statistical method One-way Analysis of Variance (ANOVA). ANOVA is an extremely useful technique for researches in the fields of economics, biology, business, education, medical, and several other disciplines. It is very important in all those situations where comparison of more than two populations is needed. This technique is used to examine the significance of the differences amongst more than two sample means at the same time.

II. RELATED WORKS

Various researchers have used ANOVA and MANOVA methods for their analyses in different fields. ANOVA method based analysis revealed that women and younger have higher levels of dental anxiety than men and elderly and therefore suffer more from dental problems. It implies that fear of pain was found to be the most important predictor of dental anxiety [6]. ANOVA analysis is also conducted in house sparrow eggs to examine the relationship between female egg provisioning (egg size) and the quality of her male partner (i.e. badge size) of sexual dimorphism. It reveals that male attributes may not affect egg size with respect to either male eggs or all eggs in the clutch [8]. ANOVA Analysis is also conducted for the comparison of students success in developmental Math course offered in three different learning environments (online, blended, and face-to-face). ANOVA method on online study shows that there are significant differences between learning environments. It reveals that 1. High performance success for online students 2. Least success performance for the blended courses have the, and finally 3. Most poorly performance success for the face-to-face students [7]. ANOVA method is also applied to the Liver patient's specimen's data to establish the presence of liver diseases and also to find out the types of liver diseases. Maximum number of cases are in the age group of 41-50 years (29%). Males are predominant in case of gender, in the study, with a male: to female ratio of 6:1. Liver Cirrhosis was the commonest disease found, comprising of 25% of the cases [3]. ANOVA method based estimation of reference values change in laboratories [4]. ANOVA and MANOVA based analysis was conducted [5] for population comparison between ILPD data set and USA datasets. It revealed that more significant differences exists in the above two datasets with all the possible attribute combinations, except analysis on SGPT between non liver patients of USA and INDIA data sets.

III. PROBLEM DEFINITION

In the present study 500 records are collected personally from various pathological labs with 13 attributes and 583 liver patient's records with 10 attributes are taken from ILPD dataset from UCI machine learning repository. A new dataset is formed with 1083 records combing the above two datasets with common attributes. The new dataset is

consisting of 634 are liver patients and 449 are non-liver patients with 10 attributes. Pre-processing is conducted on the new data set for eliminating the missing values. Then ANOVA test is conducted on liver dataset to examine the influence of gender and age differences. There are two parts in the analysis. The first part is related to gender factor and second part is related to age factor combined with gender factor. In the first part, three experiments were conducted: 1. First experiment is aimed at finding the differences in gender among the liver disease and non-liver disease patients; 2. Second experiment is aimed at finding the significant differences in gender among the liver disease patients exclusively and 3. Third experiment is aimed at finding the significant differences in gender from the non-liver disease patients exclusively. In the second part, twelve experiments are conducted treating the dataset into five age groups, namely 0 - 20, 21 - 40,41 - 60, 61-80, 81 - 100 years. Here both gender and age parameters are taken together to find the significant difference. The details of these experiments are: 4. Fourth experiment is aimed at finding the significant differences in gender with age group (0-20) among liver disease and non-liver disease patients; 5. Fifth experiment is aimed at finding the significant differences in gender with age (0-20) group among liver disease patients exclusively and 6. Sixth experiment is aimed at finding the significant differences in gender combined with age group (0-20) from non-liver disease patients exclusively; 7. Seventh experiment is aimed at finding the significant differences in gender with age group ((21-40)) among liver disease and non-liver disease patients; 8. Eighth experiment is aimed at finding the significant differences in gender with age ((21-40)) group among liver disease patients exclusively and 9. Ninth experiment is to find the significant differences in gender combined with age group (20- 40) from non-liver disease patients exclusively; 10. Tenth experiment is to find the significant differences in gender with age group (41-60)among liver disease and non-liver disease patients. 11. Eleventh experiment is to find the significant differences in gender with age (41-60)group among liver disease patients exclusively and 12. Twelve experiment is to find the significant differences in gender combined with age group (41-60)from non-liver disease patients exclusively; 13. Thirteenth experiment is to find the significant differences in gender with age group (61-80)among liver disease and non-liver disease patients; 14. Fourteenth experiment is to find the significant differences in gender with age (61-80)group among liver disease patients exclusively and 15. Fifteenth experiment is to find the significant differences in gender combined with age group (61-80)from non-liver disease patients exclusively; In the age group (80-100) have very less records and hence it is not considered for the separate experiment.

IV. EXPERIMENTAL DATASET

Pre-processing of 1083 records yields 1080 records out of which 634 are liver patients and 446 are non-liver patients. The attributes in this dataset are Age, Gender, TB, DB, Alkphos, SGPT, SGOT, TP, ALB, and A/G Ratio. Out of these attributes TB (Total Bilirubin), DB (Direct Bilirubin), TP (Total Proteins), ALB (albumin), A/G Ratio, SGPT, SGOT and Alkphos are tests related to liver functioning. They are used to measure the levels of enzymes, proteins and bilirubin levels which helps for the diagnosis of liver disease. The description of the liver dataset attributes and normal values are shown in Table I.

TABLE I
ATTRIBUTES IN LIVER DATASET

Attributes	Information(Normal Value)
Age	Age of the patient
Gender	Gender of the patient
TB(LFT)	Total- Bilirubin (0.22-1.0 mg/dl)
DB(LFT)	Direct- Bilirubin (0.0-0.2mg/dl)
Alkphos(LFT)	Alkaline Phosphates (110-310U/L)
SGPT(LFT)	AlamineAminotransferase (545U/L)
SGOT(LFT)	AspartateAminotransferase(540U/L)
TP(LFT)	Total Protiens (5.5-8gm/dl)
ALB(LFT)	Albumin (3.5-5gm/dl)
A/GRatio(LFT)	Albumin and Globulin Ratio (>=1)

V. METHODOLOGY

In almost every real world data the missing values are common problem. The presence of missing values in data results in datasets being "incomplete" i.e., since some information will not be available [12]. Pre-processing techniques are applied on raw data to make the data clean, noise free, and consistent [13]. It is observed that among 10 attributes, the results has 3 missing values. Those records are removed and the ANOVA method is applied on the clean dataset.

One way analysis of variance (ANOVA)

A one way analysis of variance or simply one-way ANOVA [10] is used to compare two means of independent (unrelated) groups using F-variance. *F-variance* is used to test the null hypothesis, depend upon null hypothesis value (p–value) decide that groups are significant or not; Usually ANOVA is used to test whether the mean of two or more populations (levels) are equal or not. Therefore he null hypothesis becomes that the means of the different populations are the same and the alternate hypothesis is atleast one sample's mean is different from the others. The steps of ANOVA procedures are as follows.

Steps in ANOVA Process:

Let $n_1, n_2, \ldots . n_k$ represent the sample sizes of k samples of the population and $\mu_1, \mu_2, \ldots . \mu_k$ represent mean of each sample.

Step1: First $F-value$ is to be computed by using the formula $F-value = \frac{MS_{between}}{MS_{with\,in}}$

Where $MS_{between}$ is the Mean Square between the samples and $MS_{with\,in}$ is the Mean Square with in the samples

Step 2: Compute $p-value$ using $F-distribution$ table that corresponds to $(n-k)$ and $(k-1)$ degrees of freedom at different levels of significance (5%) where k is the number of samples and n is the total number of items (x_{ij} 's) in all the samples.

***Step 3*:** Compare $F-value$ and $p-value$

1. If $(p-value > F-value)$ then the difference is considered as insignificant i.e. null-hypothesis (H_0) accepted.
2. If $(p-value \le F-value)$ then the difference is considered as significant i.e. the null-hypothesis (H_0) will be **rejected** with a conclusion that the samples could not have come from the same universe.

Note: 1. Generally in a statistical sample distribution, if there are n variables and m constraints on the distribution then there are n − m degrees of freedom**.**

Note: 2.The general guidelines to decide the Significance of null hypothesis:

a. If $p\_value > 0.10$ then null hypothesis not significant.
b. If $p\_value \le 0.10$ then marginally significant.
c. If $p\_value \le 0.05$ then "significant"
d. If $p\_value \le 0.01$ then "highly significant."

Computational Procedure to Calculating $MS_{between}$ **:**

The value of $MS_{between}$ is calculated using the formula

$$MS_{between} = \frac{SS_{between}}{(k-1)}$$

Where $(k-1)$ represents degrees of freedom (df) between the samples and $SS_{between}$ is sum of squares for variance between the samples.

i.e. $SS_{between} = n_1(\mu_1 - \bar{\mu})^2 + n_2(\mu_2 - \bar{\mu})^2 + \ldots\ldots\ldots + n_k(\mu_k - \bar{\mu})^2$

Where $(k-1)$ represents degrees of freedom (df) between the samples and $SS_{between}$ is sum of squares for variance between the samples.

i.e. $SS_{between} = n_1(\mu_1 - \bar{\mu})^2 + n_2(\mu_2 - \bar{\mu})^2 + \ldots\ldots\ldots + n_k(\mu_k - \bar{\mu})^2$

Computational Procedure to Calculating MS_{within} **:**

To obtain the deviations of the values of the sample items for all the samples from corresponding means of the samples and calculate the squares of such deviations and then obtain their total. The variance or mean square (*MS*) within samples is calculated using the formula

$$MS_{with\,in} = \frac{SS_{with\,in}}{(n-k)}$$

Where $(n-k)$ represents degrees of freedom within samples, n is the total number of items in all the samples. and SS_{within} represents sum of squares for variance within samples and it is given by the equation.

i.e

$$SS_{within} = \sum_{i=1,2,3.....}(x_{1i} - \mu_1)^2 + \sum(x_{2i} - \mu_2)^2 + \ldots\ldots\ldots + \sum(x_{ki} - \mu_k)^2.$$

***Analysis of variance table*:**

For the sake of convenience the information obtained through various steps stated above can be tabulated as shown in the table II.

TABLE II
Analysis of Variance Table

Source of variation	*Sum of Squares(SS)*	*Degrees of freedom*	*Mean Square(MS) (This is SS divided by df)*	*F Ratio*
Between samples or categories	$SS_{between} = n_1(\mu_1 - \bar{\mu})^2 + n_2(\mu_2 - \bar{\mu})^2 + \ldots\ldots\ldots + n_k(\mu_k - \bar{\mu})^2$	$k-1$	$\frac{SS_{between}}{(k-1)}$	$\frac{MS_{between}}{MS_{within}}$
Within samples or categories	$SS_{within} = \sum(x_{1i} - \mu_1)^2 + \sum(x_{2i} - \mu_2)^2 + \ldots\ldots\ldots\ldots \sum(x_{ki} - \mu_k)^2$ $i = 1,2,3.....$	$n-k$	$\frac{SS_{within}}{(n-k)}$	
Total	$\sum(x_{ij} - \bar{\mu})^2$ where $i = 1,2,3,\ldots..,\ j = 1,2,3,\ldots$	$(n-1)$		

VI. RESULTS AND DISCUSSION

*A. **ANOVA Analysis with respect to Gender***

In this study analysis is done for Gender and Age groups, gender difference find out among liver & non-liver disease patients , liver disease patients , non-liver disease patients based on the following common attributes, Total Bilirubin, Direct Bilirubin, Alkphos, SGPT SGOT Total_Protien, Serum albumin and A/G Ratio . The dataset of 1080 records is analyzed in three experiments. One is, the total dataset of liver & Non -liver disease patients is divided into male and female groups; second is, only liver disease patients are divided into male and female groups; third is, non-liver patients are divided into male and female groups.

Experiment1

Liver & non-liver disease patients with gender

This analysis includes both the liver disease and non- liver disease patients as one dataset which contains 1080 records (as three missing values are found out of total dataset of 1083 records), out of this 728 are male and 352 are female. ANOVA analysis for each attributes is given in Table I. Age, Total Bilirubin, Direct Bilirubin etc…After the analysis, the results are obtained for all attributes. In that

some attributes are significant, those significant attribute results are given in Table III to table XVIII. By ANOVA analysis significant attributes found are Total Bilirubin, Direct Bilirubin, Alkphos, SGPT, SGOT, Total_Protien, Serum albumin and A/G Ratio. In that some attributes are significant, those significant attribute results are given in Table III to Table XVIII. By ANOVA analysis significant attributes found are Total Bilirubin, Direct Bilirubin, Alkphos, SGPT, SGOT, Total_Protien, Serum albumin and A/G Ratio.

The significant values are in Table IV,VI,VIII,X,XII, XIV, XVI ,XVIII of Total Bilirubin, Direct Bilirubin, Alkphos, SGPT, SGOT, Total_Protien, Serum Albumin and A/G Ratio are 0.002 ,0.002, 0.012, 0.001, 0.002, 0.015, 0.000, 0.001 which is less than 0.05 at $p < 0.05$. Thus, the null hypothesis cannot be rejected. This confirms that there is significant difference among the groups, that is, it is observed that Total Bilirubin, Direct Bilirubin, Alkphos, SGPT, SGOT, Total_Protien, Serum albumin and A/G Ratio attributes are instrumental in diagnosing liver disease patients Table III, V, VII, IX, XI, XIII, XV, XVI contains the ranges are Total_Bilirubin(0.400, 75.00 mg/dl), Direct_Bilirubin(0.100, 19.7 mg/dl), Alkphos(11, 2110 U/L), SGPT(10, 200 U/L), SGOT (11, 4929 U/L), Total Protiens(2.700, 9.600 gm/dl), Serum Albumin (0.900,5.800 gm/dl), A/G Ratio(0.300, 2.800) ranges of male liver disease and non-liver disease patients and Total Bilirubin (0.400,27.70 mg/dl), Direct Bilirubin (0.100, 12.8 mg/dl), Alkphos(63, 1896 U/L), SGPT(10, 790 U/L), SGOT(10,1050 U/L), Total_Protiens(3.600, 9.200 gm/dl), Serum Albumin(1.000, 5500 gm/dl), A/G Ratio(0.300, 1.800) ranges of female liver disease and non-liver disease patients. In each table SD stands for standard deviation and SE stands for standard error. The ranges are different in male to female. This corroborates that gender is important in diagnosing liver disease, gender is instrumental with above all attributes.

The ANOVA analysis done in this group provides non-significant attributes: Age is 0.996, the value is greater than 0.05 at $p < 0.05$.Thus the null hypothesis is rejected This confirms that there is no significance among the groups, that is, it can be said that Age attribute is not instrumental in diagnosis.

This Experiment 1 confirms that there is a significant gender difference on TB, DB, ALKPHOS, SGPT, SGOT, Total_Protien, Serum albumin, A/G ratio accepts age attribute with liver and non-liver disease male and female population. The result confirms that gender difference is plays an instrumental role in diagnosis process with above attributes.

TABLE III
DESCRIPTIVE STATISTIC OF TOTAL_BILIRUBIN

TB	N	Mean	SD	SE	Min	Max
1Male	728	3.598753E0	5.5774751	.2064316	.4000	75.0000
2 fema	352	2.576663E0	3.8515894	.2047097	.4000	27.7000
Total	1080	3.264972E0	5.0995269	.1548871	.4000	75.0000

TABLE IV
ANOVA ANALYSIS OF TB BETWEEN GENDER

TB	SS	Df	MS	F	Sig.
Between groups	249.042	1	249.042	9.635	.002
Within groups	27914.562	1078	25.799		
Total	28163.604	1079			

TABLE V
DESCRIPTIVE STATISTIC OF DIRECT_BILIRUBIN

DB	N	Mean	SD	SE	Min	Max
1 male	728	1.591466E0	2.4870934	.0920515	.1000	19.7000
2 female	352	1.066893E0	1.8271635	.0971126	.1000	12.8000
Total	1080	1.420157E0	2.3048841	.0700059	.1000	19.7000

TABLE VI
ANOVA ANALYSIS OF DB BETWEEN GENDER

DB	SS	Df	MS	F	Sig.
Between groups	65.601	1	65.601	12.479	.000
Within groups	5687.827	1078	5.257		
Total	5753.427	1079			

TABLE VII
DESCRIPTIVE STATISTIC OF ALKPHOS

ALKPHOS	N	Mean	SS	SE	Min	Max
1 male	728	222.55	191.549	7.090	11	2110
2 female	352	192.24	211.970	11.266	63	1896
Total	1080	212.00	198.930	6.042	11	2110

TABLE VIII
ANOVA ANALYSIS OF ALKPHOS BETWEEN GENDER

ALKPHOS	SS	Df	MS	F	Sig.
Between groups	248792.067	1	248792.067	6.318	.012
Within groups	4.261E7	1078	39379.780		
Total	4.286E7	1079			

TABLE IX
DESCRIPTIVE STATISTIC OF SGPT

SGPT	N	Mean	SD	SE	Min	Max
1 male	728	72.17	159.936	5.919	10	2000
2 female	352	43.51	61.887	3.289	10	790
Total	1080	62.81	136.556	4.148	10	2000

TABLE X
ANOVA ANALYSIS OF SGPT BETWEEN GENDER

SGPT	SS	Df	MS	F	Sig.
Between groups	195814.395	1	195814.395	10.594	.001
Within groups	2.000E7	1078	18483.746		
Total	2.020E7	1079			

TABLE XI
DESCRIPTIVE STATISTIC OF SGOT

SGOT	N	Mean	SD	SE	Min	Max
1 male	728	94.92	254.004	9.401	11	4929
2 female	352	52.22	84.025	4.466	10	1050
Total	1080	80.97	214.782	6.524	10	4929

TABLE XII
ANOVA ANALYSIS OF SGOT BETWEEN GENDER

SGOT	SS	Df	MS	F	Sig.
Between groups	434668.904	1	434668.904	9.496	.002
Within groups	4.953E7	1078	45772.402		
Total	4.996E7	1079			

TABLE XIII
DESCRIPTIVE STATISTIC OF TOTALPROTIEN

TP	N	Mean	SD	SE	Min	Max
1 male	728	6.557123E	.9011561	.0333533	2.70	9.60
2female	352	6.693785E	.7857555	.0417624	3.60	9.20
Total	1080	6.601753E	.8671744	.0263385	2.70	9.60

TABLE XIV
ANOVA ANALYSIS OF TP BETWEEN GENDER

TP	SS	Df	MS	F	Sig.
Between groups	4.452	1	4.452	5.948	.015
Within groups	809.954	1078	.749		
Total	814.407	1079			

TABLE XV
DESCRIPTIVE STATISTIC OF SERUM ALBUMIN

SALB	N	Mean	SD	SE	Min	Max
1 male	728	3.364521E	.7319157	.0270924	.90	5.80
2female	352	3.560452E	.6242192	.0331769	1.00	5.50
Total	1080	3.428506E	.7043652	.0213936	.90	5.80

TABLE XVI
ANOVA ANAYSIS SALB BETWEEN GENDER

SALB	SS	Df	MS	F	Sig.
Between groups	9.152	1	9.152	18.749	.000
Within groups	528.157	1078	.488		
Total	537.309	1079			

TABLE XVII
DESCRIPTIVE STATISTIC OF A/G RATIO

A/G	N	Mean	SD	SE	Min	Max
1 male	728	1.062624E0	.3169531	.0117471	.30	2.80
2female	352	1.126051e0	.2686953	.0143215	.30	1.80
Total	1080	1.083296e0	.3034092	.0092324	.30	2.80

TABLE XVIII
ANOVA ANALYSIS A/G RATIO BETWEEN GENDER

A/G RATIO	SS	Df	MS	F	Sig.
Between groups	.955	1	0.955	10.460	.001
Within groups	98.375	1078	0.091		
Total	99.330	1079			

Experiment 2

Liver disease patients with gender

This analysis includes the liver disease patients selectively from both data sets. ANOVA analysis compared male and female liver disease patients. For this analysis, the data contains 634 total records, in that 465 male liver disease is done for each of attributes as given in Table I. After the analysis the results are obtained for 9 attributes.In that some attributes are significant, those significant attributes' results are given in Table XIX to Table XXIV by ANOVA analysis significant attributes found are SGPT, SGOT, and there Serum albumin. The Significant values are shown in Table XX, XXII,XXIV of SGPT, SGOT, and Serum albumin are 0.024, 0.041, 0.020 which is less than0.05at $p < 0.05$ Thus, the null hypothesis is cannot be rejected.

TABLE XIX
DESCRIPTIVE STATISTIC OF SGPT

SGPT	N	Mean	SD	SE	Min	Max
1 Male	465	93.85	196.855	9.129	10	2000
2 Female	169	58.28	85.986	6.614	10	790
Total	634	84.36	174.974	6.949	10	2000

TABLE XX
ANOVA ANALYSIS SGPT BETWEEN GENDER

SGPT	SS	Df	MS	F	Sig.
Between groups	156800.054	1	156800.054	5.155	.024
Within groups	1.922E7	632	30416.084		
Total	1.938E7	633			

TABLE XXI
DESCRIPTIVE STATISTIC OF SGOT

SGOT	N	Mean	SD	SE	Min	Max
1 male	465	129.42	312.962	14.513	11	4929
2 female	169	78.81	115.257	8.866	10	1050
Total	634	115.93	275.359	10.936	10	4929

TABLE XXII
ANOVA ANALYSIS OF SGOT BETWEEN GENDER

SGOT	SS	Df	MS	F	Sig.
Between groups	317414.680	1	317414.680	4.207	.041
Within groups	4.768E7	632	75440.362		
Total	4.800E7	633			

TABLE XXIII
DESCRIPTIVE STATISTIC OF SERUM ALBUMIN

SALB	N	Mean	SD	SE	Min	Max
1 Male	465	3.192258E	.7816485	.03624	.900	5.80
2Female	169	3.354438E	.7488848	.05760	1.00	4.90
Total	634	3.235489E	.7757904	.03081	.900	5.80

TABLE XXIV
ANOVA ANALYSIS OF SALB BETWEEN GENDER

SALB	SS	Df	MS	F	Sig.
Between groups	3.260	1	3.260	6.455	0.020
Within groups	377.711	632	0.598		
Total	380.971	633			

This confirms that is significance difference among the groups that is, we can say that, SGPT, SGOT and Serum albumin attributes are instrumental in diagnosing liver disease patients in male and female population. The Significant attributes ranges are shown in Table XIX, XXI,

XXIII contains the ranges are SGPT(10, 200U/L), SGOT(11, 4929 U/L), Serum Albumin(0.900, 5.800 gm/dl)) ranges of male liver disease patients and SGPT (10 ,790U/L) SGOT (10, 1050U/L) Serum Albumin (1.0000, 4.9000 gm/dl) ranges of female liver disease patients. The ranges are different in male to female liver disease patients. This corroborates that gender is important in diagnosing liver disease, gender is instrumental with above all attributes.

The ANOVA analysis done in this group provides non-significant attributes: Age, Total_Bilirubin ,Direct_Bilirubin, Alkphos, Total_Bilirubin , A/G Ratio. The significant values are 0.124, 0.292, 0.140, 0.675, 0.201, 0.148 respectively groups that is, it can be said that Age, Total_Bilirubin ,Direct_Bilirubin, Alkphos, Total_Bilirubin , A/G Ratio are not instrumental in diagnosing liver disease patients in male and female population the above those values are greater than 0.05 at $p < 0.05$.Thus, the null hypothesis is rejected. This confirms that there is no significance among the groups that is, we can say that Age, Total_Bilirubin, Direct_Bilirubin, Alkphos, Total_Bilirubin, A/G Ratio is not instrumental in diagnosing of liver disease male and female population.

The Experiment 2confirms that there is significant difference among the groups considered , that is, it can be said that SGOT, SGPT, Serum Albumin are instrumental in diagnosing liver disease in male and female population. The ranges are different in male to female liver disease patients. This corroborates that gender is important in diagnosing liver disease, gender is instrumental with SGOT, SGPT and Serum Albumin.

Experiment 3

Non-liver disease patients with gender

This analysis includes the non-liver disease patients selectively from both data sets and compared male and female non- liver patients. For this analysis, both data sets, contains 449 total records; in that 264 male and 185 female non-liver disease patients. ANOVA analysis is done for each of attributes as given in Table I. After the analysis, the results are obtained for 9 attributes. In that some attributes are significance, those significance attribute results are given in Table XXV to Table XXXIV. By ANOVA analysis significant attributes found are TB, DB, ALKPHOS, SGPT, and SGOT. The Significant values are shown in Table XXVI XXVIII, XXX, XXXIV of TB, DB, ALKPHOS, SGPT, SGOT values are 0.002,0.041,0.001,0.005,0.000 which is less than0.05at $p < 0.05$. Thus the null hypothesis cannot be rejected. This confirms that there is significant difference among the groups, that is, it can be said that, TB, DB, ALKPHOS, SGPT, SGOT attributes are instrumental in diagnosing non-liver disease patients in male and female population.

The Significant attributes ranges are shown in Tables XXV, XXVII, XIX, XXXI, XXXIII contains the ranges are TB(0.40, 5.00 mg/dl) of male and (0.40, 2.20mg/dl) female , DB(0.10, 3.89mg/dl) of male and (0.10, 0.88mg/dl) of female, ALKPHOS(63,901U/L) of male and (63, 406 U/L) of female SGPT(12,119U/L) of male and (10,77 U/L) of female SGOT(12,125 U/L) of male and (10, 78 U/L) of female ranges of non-liver disease patients. The ranges are different in male to female liver disease patients. This corroborates that gender is important in diagnosing non-liver disease, gender is instrumental with above all significant attributes.

The ANOVA analysis done in this group provides non-significant attributes: Age, Total_Protien, Serum albumin A/G Ratio ,the significant values are 0.453,0.224,0.058,0.300 respectively the above those values are greater than 0.05 at $p < 0.05$.Thus, the null hypothesis is rejected .This confirms that there is no significance among the groups that is, we can say that Age, Total_Protien, Serum albumin, A/G ratio is not instrumental in diagnosing non- liver disease in male and female population.

In the Experiment 3 confirms that there is a significance gender difference on TB,DB ALKPHOS, SGPT, SGOT attributes with selectively non-liver disease patients, remaining attributes are not significance, the result, confirms there is a gender difference on TB, DB, ALKPHOS, SGPT and SGOT attributes with non-liver disease patients. The above attributes are Instrumental in diagnosing non-liver disease in male and female population.

TABLE XXV
DESCRIPTIVE STATISTIC OF TOTAL_BILIRUBIN

TB	N	Mean	SD	SE	Min	Max
1 Male	261	.825682	.3761000	.0231474	.4000	5.0000
2Female	185	.730324	.1872945	.0137702	.4000	2.2000
Total	446	.786392	.3156815	.0148979	.4000	5.0000

TABLE XXVI
ANOVA ANALYSIS OF TB BETWEEN GENDER

TB	SS	Df	MS	F	Sig.
Between groups	.989	1	.989	10.127	.002
Within groups	43.656	444	.098		
Total	44.645	445			

TABLE XXVII
DESCRIPTIVE STATISTIC OF DIRECT_BILIRUBIN

DB	N	Mean	SD	SE	Min	Max
1 Male	261	.287159	.2939709	.180927	.1000	3.890
2Female	185	.240973	.1007641	.0074083	.1000	.8800
Total	446	.268129	.2354158	.0111100	.1000	3.890

TABLE XXVIII
ANOVA ANALYSIS OF DB BETWEEN GENDER

DB	SS	Df	MS	F	Sig.
Between groups	.232	1	.232	4.217	.041
Within groups	24.56	444	.055		
Total	24.828	445			

TABLE XXIX
DESCRIPTIVE STATISTIC OF ALKPHOS

ALKPHOS	N	Mean	SD	SE	Min	Max
1 Male	261	157.82	87.288	5.372	63	901
2 Female	185	134.89	57.022	3.751	63	406
Total	446	148.37	75.297	3.553	63	901

TABLE XXX
ANOVA ANALYSIS OF ALKPHOS BETWEEN GENDER

ALKPHOS	SS	Df	MS	F	Sig.
Between groups	57173.776	1	57173.776	10.293	.001
Within groups	2482839.111	444	5554.450		
Total	2540012.886	445			

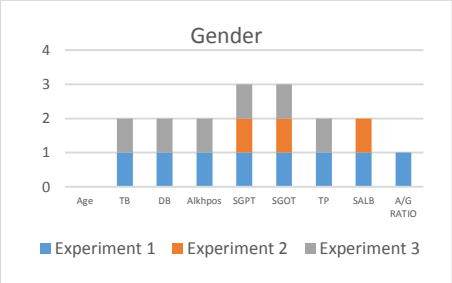

Fig. 1. Significant attributes with respect to Gender

As given in the above graph Fig.1, it is observed that the common significant attributes in the all three experiments by gender differences are SGOT, SGP. This corroborates that gender is important in diagnosing liver diseases with respective to above attributes.

B. ANOVA Analysis on Age factor with respect to Gender

In this study of liver disease & non-liver disease patients of both datasets all age groups is done .Both datasets records are divided into five groups according to age as given under.

Group1 - (0-20) years
Group2 - (21-40) years
Group3 - (41-60) years
Group4 - (61-80) years
Group 5 - (81-100) years

The above groups consist of 71, 417, 455, 130 and 7 records respectively. Since the *Group5* has a very few records, it is not considered for analysis. Hence only *Group1, 2, 3, 4* are considered and these four groups have a total records of 1073.

In the above ANOVA analysis of Gender three experiments results are shows that age is not instrumental, in Age is taken as one interval group. But we hope to move for analysis the age is divided particular age groups with gender difference.

In ANOVA analysis of Age and Gender the aim is to find out the gender and age difference significance.

TABLE XXXI
DESCRIPTIVE STATISTIC OF SGPT

SGPT	N	Mean	SD	SE	Min	Max
1 Male	261	34.16	15.721	.968	12	119
2 Female	185	30.06	14.601	1.074	10	77
Total	446	32.47	15.386	.726	10	119

TABLE XXXII
ANOVA ANALYSIS OF SGPT BETWEEN GENDER

SGPT	SS	Df	MS	F	Sig.
Between groups	1831560	1	1831.560	7.855	.005
Within groups	104226.342	444	233.169		
Total	106057.902	445			

TABLE XXXIII
DESCRIPTIVE STATISTIC OF SGOT

SGOT	N	Mean	SD	SE	Min	Max
1 Male	261	34.36	15.409	.948	12	125
2Female	185	28.02	13.145	.966	10	78
Total	446	31.75	14.836	.700	10	125

TABLE XXXIV
ANOVA ANALYSIS OF SGOT BETWEEN GENDER

SGOT	SS	Df	MS	F	Sig.
Between groups	4364.612	1	4364.612	20.702	.000
Within groups	94242.444	444	210.833		
Total	98607.056	445			

Every age group is repeated for the below three experiments and find out which attributes are significant with respect to the age groups. Experiment- 4 Liver and non -liver disease patients, Experiment- 5 Liver disease patient, Experiment- 6 Non -liver disease patients.

Experiment- 4
Liver &non-liver disease patients with gender (0-20) age group

This *Experiment- 4* includes the liver disease and non-liver disease patients from both data sets and compare male and female liver patients with non- liver disease patients. In the analysis of both data sets 71 records, are considered (45 male, and 26 female liver & non- liver disease Patient's records), ANOVA analysis is carried out for each of attributes as given in given Table I. After the *Experiment 4*, the results are obtained for 8 remaining attributes. In that all attributes are non-significance of age group (0-20) with respective to Gender, results are obtained in tables, but the space consumption only significance tables are presented in this paper.

ANOVA analysis done in this groups provides non- significant attributes: TB, DB Alkphos, SGOT,SGPT, Total_Protien, Serum albumin , A/G Ratio..The insignificant values are0.387, 0.311, 0.290, 0.942, 0.942,0.306,0.337,0.502 respectively Those values

are greater than 0.05 at $p < 0.05$.Thus, the null hypothesis is rejected .This confirms that there is no significance among the groups that is, it can be said that TB, DB, Alkphos, SGPT, SGOT, Total_Protien, Serum albumin, A/G Ratio is not instrumental in diagnosing liver & non-liver disease in male and female population.

Experiment- 5

Liver disease patients with gender (0-20) age group

This *Experiment- 5* includes the liver disease Patients selectively from both data sets and compare male liver disease patients and female liver disease patients. In the analysis (0-20) age group of dataset contains 40 total records, in that 29 male liver disease patients, 11 female liver disease patients' records. ANOVA analysis is carried out for each of attributes as given in given Table I. After the *Experiment-5* the results are obtained in all attributes. In that some attributes are significance, those significance attribute results are given in Table XXXV to Table XXXVI. It shows ANOVA analysis of significant attribute Direct_Bilirubin.

TABLE XXXV
DESCRIPTIVE STATISTIC OF DIRECT_BILIRUBIN

DB	N	Mean	SD	SE	MIn	Max
1	29	1.274138E	1.3593053	.2524167	.10	4.28
2	11	2.970000E	3.7692970	1.1364858	.10	11.8
Total	40	1.740500E	2.3574627	.3727476	.10	11.8

TABLE XXXVI
ANOVA ANALYSIS OF DB BETWEEN GENDER

DB	SS	df	MS	F	Sig.
Between Groups	22.936	1	22.936	4.497	.041
Within Groups	193.812	38	5.100		
Total	216.748	39			

Tables XXXV, XXXVI contains ANOVA result of Direct_Bilirubin The significant value 0.041 in Table[36] is smaller than 0.05 at $p < 0.05$.Thus, the null hypothesis is cannot be rejected. This confirms that there is significance difference among the groups that is, we can say that Direct_Bilirubin is instrumental in diagnosing liver disease in male and female population. Table XXXV contains the ranges (0.1 - 4.28 mg/dl) of male and (0.1-11.8mg/dl) of female liver disease patients. The ranges are different in male to female liver disease patients. This corroborates that gender important in diagnosing liver disease, gender with age group is instrumental with Direct_Bilirubin.

ANOVA analysis of non-significant attributes are, Total bilirubin, Alkphos, SGOT, SGPT, Total_Protien, Serum albumin, A/G Ratio the insignificant values 0.55, 0.493, 0.622, 0.444, 0.291, 0.884, 0.631.Respectively the above those values are greater than 0.05 at $p > 0.05$.Thus, the null hypothesis is rejected .This confirms that there is no significance among the groups that is, we can say that the attributes are not instrumental in diagnosing liver disease in male and female population.

Experiment- 6

Non-liver disease patients with gender (0-20) age group

This *Experiment-6* includes the non-liver disease patients selectively from both data sets and comparison of male liver disease patients with female liver disease patients. In the analysis of age group (0-20), the dataset contains 31total records (16 male and 15 female liver disease patients' records). ANOVA analysis for each of 8 attributes is given in Table I. After the *Experiment-6* the results are obtained for 8 attributes. In that some attributes are significant, those significant attribute results are given in Table XXXVII. Table XXXVIII. By it shows ANOVA analysis significant attributes found are Alkphos.

TABLE XXXVII
DESCRIPTIVE STATISTIC OF ALKPHOS

ALKPHOS	N	Mean	SD	SE	Min	Max
1	16	305.94	148.221	37.055	96	592
2	15	144.20	74.745	19.299	85	350
Total	31	227.68	142.627	25.617	85	592

TABLE XXXVIII
ANOVA ANALYSIS OF ALKPHOS BETWEEN GENDER

ALKPHOS	SS	df	MS	F	Sig.
Between Groups	202521.437	1	202521.43	14.404	.001
Within Groups	407755.338	29	14060.52		
Total	610276.774	30			

Tables XXXVII, XXXVIII contains ANOVA result of Alkphos. The significant value 0.001 in Table XXXVIII is smaller than 0.05 at $p < 0.05$.Thus, the null hypothesis is cannot be rejected. This confirms that there is significance difference among the groups that is, we can say that Alkphos is instrumental in diagnosing non-liver disease in male and female population. Table XXXVII contains the ranges (96-592 U/L) of B and (85-350 U/L) of female liver disease patients. The ranges are different in male to female non-liver disease patients. This corroborates that gender with age group is important in diagnosing liver Disease, gender in 0-20 age group instrumental with Alkphos.

Analysis of non-significant attributes are Total_Bilirubin, Direct- Bilirubin, SGOT, SGPT, Total_Protien,, Serum albumin , A/G ratio. The insignificant values are 0.248, 0.805, 0.706, 0.335,0327, 0.710, 0.296 respectively the above those values are greater than 0.05 at $p > 0.05$.Thus, the null hypothesis is rejected diagnosing non-.This confirms that there is insignificance among the groups that is, we can say that the attributes are not instrumental in liver disease in male and female population.

In the above three Experiments of (liver & non-liver disease, liver disease, non-liver disease patients) it is observed that in (0 -20) age group patients , the common significant attributes are DB, Alkhpos. This corroborates that gender difference with age is important in diagnosing liver disease. The gender with age group is instrumental with DB, Alkhpos.

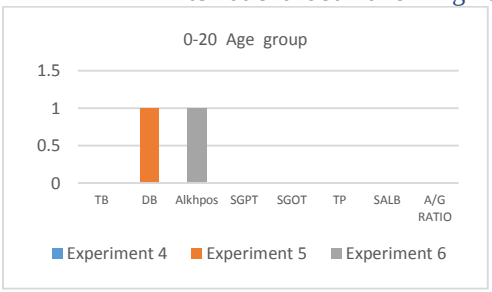

Fig. 2. significant attributes of 0-20 age group with gender

Experiment-7

Liver & non-liver disease patients with gender (21-40) age group

This *Experiment-7* includes the liver disease and non-liver disease patients selectively from both data sets and comparison of male liver disease patients and female liver disease patients. In this analysis (21-40) age group dataset contains 417 total records (278 male and 139 female liver disease patients' records). ANOVA analysis for each of 8 attributes is given in Table I. After the *Experiment-7* the results are obtained for 8 attributes. In that some attributes are significance and those significant attribute results are given in Table XXXIX to Table L. By ANOVA analysis significant attributes found are TOTAL_BILIRUBIN, DIRECT_BILIRUBIN, SGOT, SGPT, Serum albumin, A/G Ratio.

TABLE XXXIX
DESCRIPTIVE STATISTIC OF TOTAL_BILIRUBIN

TB	N	Mean	SD	SE	Min	Max
1	278	3.901942E	5.761024	.3455233	.46	42.80
2	139	2.581223E	2.743768	.2327234	.40	10.28
Total	417	3.461703E	4.998558	.2447805	.40	42.80

TABLE XXXX
ANOVA ANALYSIS OF TB BETWEEN GENDER

TB	SS	df	MS	F	Sig.
Between Groups	161.638	1	161.638	6.556	.011
Within Groups	10232.366	415	24.656		
Total	10394.005	416			

TABLE XLI
DESCRIPTIVE STATISTIC OF DIRECT_BILIRUBIN

DB	N	Mean	SD	SE	Min	Max
1	278	1.782050E0	2.9149188	.1748252	.10	19.70
2	139	1.024748E0	1.2016247	.1019205	.10	4.69
Total	417	1.529616E0	2.5028847	.1225668	.10	19.70

TABLE XLII
ANOVA ANALYSIS OF DB BETWEEN GENDER

DB	SS	df	MS	F	Sig.
BetweenGroups	53.145	1	53.145	8.639	.003
Within Groups	2552.859	415	6.151		
Total	2606.004	416			

TABLE XLIII
DESCRIPTIVE STATISTIC OF SGPT

SGPT	N	Mean	SD	SE	Min	Max
1	278	94.33	235.323	14.114	.1000	19.7000
2	139	40.63	35.130	2.980	.1000	4.6900
Total	417	76.43	194.744	9.537	.1000	19.7000

TABLE XLIV
ANOVA ANALYSIS OF SGPT BETWEEN GENDER

SGPT	SS	df	MS	F	Sig.
BetweenGroups	267236.260	1	267236.260	7.151	.008
Within Groups	1.551E7	415	37372.703		
Total	1.578E7	416			

TABLE XLV
DESCRIPTIVE STATISTIC OF SGOT

SGOT	N	Mean	SD	SE	Min	Max
1	278	110.29	259.608	15.570	.1000	19.7000
2	139	44.99	39.118	3.318	.1000	4.6900
Total	417	88.53	215.254	10.541	.1000	19.7000

TABLE XLVI
ANOVA ANALYSIS OF SGOT BETWEEN GENDER

SGOT	SS	df	MS	F	Sig.
Between Groups	395252.201	1	395252.201	8.688	.003
Within Groups	1.888E7	415	45493.715		
Total	1.928E7	416			

TABLE XLVII
DESCRIPTIVE STATISTIC OF SERUM ALBUMIN

SALB	N	Mean	SD	SE	Min	Max
1	278	3.534532E0	.7022617	.0421189	.10	19.700
2	139	3.704317E0	.5155369	.0437273	.10	4.6900
Total	417	3.591127E0	.6503645	.0318485	.10	19.700

TABLE XLVIII
ANOVA ANALYSIS OF SALB BETWEEN GENDER

SALB	SS	df	MS	F	Sig.
Between Groups	2.671	1	2.671	6.397	.012
Within Groups	173.286	415	.418		
Total	175.957	416			

TABLE XLIX
DESCRIPTIVE STATISTIC OF A/G RATIO

A/G	N	Mean	SD	SE	Min	Max
1	277	1.120217E0	.2979446	.0179018	.30	2.50
2	138	1.180580E0	.2403458	.0204596	.60	1.80
Total	415	1.140289E0	.2812600	.0138065	.30	2.50

TABLE L
ANOVA ANALYSIS OF A/G RATIO BETWEEN GENDER

SGPT	SS	df	MS	F	Sig.
Between Groups	.336	1	.336	4.276	.039
Within Groups	32.415	413	.078		
Total	32.750	414			

The Significant values are shown in Table XL, XLII, XLIV, XLVI, XLVIII, L of Total_Bilirubin, DIRECT_BILIRUBIN, SGOT, SGPT, Serum albumin, A/G ratio. are 0.011,0.003,0.008,0.003,0.012,0.039 which is less than0.05at $p < 0.05$ Thus, the null hypothesis cannot be rejected. This confirms that there is significant difference among the groups, that is, it can be said that Total_Bilirubin, Direct_Bilirubin, SGOT, SGPT, Serum albumin, A/G ratio attributes are instrumental in diagnosing liver disease and non-liver disease patients in (21-40) age group male and female population.

The Significant attributes ranges are shown in Tables XXXIX, XLI, XLIII, XLV, XLVII,XLIX contains the ranges are TB(0.46-42.8 mg/dl) of male and (0.40-10.28 mg/dl) of female, DB(0.1-19.7 mg/dl) of male and(0.1-4.6 mg/dl) of female SGPT (11-2000 U/L) of male and (11- 232 U/L) of female SGOT (0.1-19.7 mg/dl) of male and(0.1-4.6 mg/dl) of female, Serum albumin (0.9-5.8 gm/dl) of male and (1.9-5.5 gm/dl) of female A/G Ratio (0.3-2.5) of male and 0.6-1.8) of female ranges of liver disease and non-liver disease patients. The ranges are different in male to female liver disease patients. This corroborates that gender is important in diagnosing liver disease or non-liver disease, gender with age group (21-40) instrumental with above all significant attributes

The ANOVA analysis is done in this group provides non-significant attributes: Alkphos, Total_Protien. The insignificant values are 0.057, 0.232 respectively groups that is, it can be said that the above those values are greater than 0.05 at $p > 0.05$.Thus, the null hypothesis is rejected .This confirms that there is no significance among the groups that is, it can be said that Alkphos, Total_Protien attributes are not instrumental in diagnosing liver disease or non- liver disease in male and female population.

Experiment- 8

Liver disease patients with gender (21-40) age group

This *Experiment- 8* includes selectively liver disease patients of both data sets and comparison of male and female liver disease patients. In the analysis, (21-40) age group dataset contains 246 total records (178 male and 68 female liver disease patients' records). ANOVA for each of attributes is given in Table I. After the *Experiment- 8* the results are obtained for 8 attributes. In that some attributes are significant, those significant attribute results are given in Table LI to Table LIV. By ANOVA analysis of significant attributes found are SGOT, SGPT. The Significant values are shown in Tables LII, LIV of SGOT, and SGPT are 0.031,0.023 which is less than0.05at $p <$ confirms that there is significance difference among the groups 0.05 Thus, the null hypothesis is cannot be rejected. This that is, we can say that SGOT, SGPT, attributes are Instrumental in diagnosing liver disease patients in 21-40 age group male and female population.

TABLE LII
ANOVA ANALYSIS OF SGPT BETWEEN GENDER

SGPT	N	Mean	SD	SE	Min	Max
1	178	127.37	288.863	21.651	11	2000
2	68	51.03	44.939	5.450	11	232
Total	246	106.26	249.008	15.876	11	2000

TABLE LI
DESCRIPTIVE STATISTIC OF SGPT

SGPT	SS	df	MS	F	Sig.
Between Groups	256714.620	1	286714.620	4.694	.031
Within Groups	1.490E7	244	61084.193		
Total	1.519E7	245			

TABLE LIII
DESCRIPTIVE STATISTIC OF SGOT

SGOT	N	Mean	SD	SE	Min	Max
1	178	153.08	316.611	23.731	16	2946
2	68	64.69	47.763	5.792	10	248
Total	246	128.65	273.154	17.416	10	2946

TABLE LIV
ANOVA ANALYSIS OF SGOT BETWEEN GENDER

SGOT	SS	df	MS	F	Sig.
Between Groups	384392.818	1	384392.818	5.241	.023
Within Groups	1.790E7	244	73343.481		
Total	1.828E7	245			

The Significant attributes ranges are shown in Table LI, LIII contains the ranges are SGPT(11, 20000 U/L) of male and (11, 232 U/L) of female SGOT(16, 2946 U/L) of male and (10, 248 U/L) of female ranges of liver disease patients. The ranges are different in male to female liver disease patients. This corroborates that gender is important in diagnosing liver disease patients, gender in 21-40 age group instrumental with above SGOT, SGPT significant attributes.

The Non-insignificant attributes are TB,DB, Alkphos, Serum albumin, A/G Ratio The insignificant values are 0.076, 0.647, 0.0547, 0.0598, 0.067 respectively the above those values are greater than 0.05 at $p > 0.05$.Thus, the null hypothesis is rejected. This confirms that there is insignificance among the groups that is, we can say that the attributes are not instrumental in diagnosing liver disease patients with 21-40 age group in male and female population.

Experiment- 9

Non-liver disease patients with gender (21-40) age group

This *Experiment- 9* includes the selectively non-liver disease patients of both data sets and compare male non-liver disease patients and female non-liver disease patients. In the analysis (21-40) age group of dataset contains 171 total records, in that 100 male non- liver disease patients, 71 female non- liver disease patients' records. We have carried out one way analysis of variance (ANOVA) for each of attributes in given Table I. after the *Experiment-9* results are obtained in all attributes. In that

some attributes are significance, those significance attribute results are given in Table LV to Table LVIII It shows ANOVA analysis of significant attributes is Total_Bilirubin, SGOT.

The Significant values are shown in Tables LVI, LVIII of TOTAL_BILIRUBIN, SGOT are 0.009,0.000 which is less than0.05at $p < 0.05$ Thus, the null hypothesis is cannot be rejected. This confirms that there is significance difference among the groups that is, we can say that TOTAL_BILIRUBIN, SGOT, attributes are instrumental in diagnosing Non -liver disease patients in (21-40) age group male and female population

TABLE LV
DESCRIPTIVE STATISTIC OF TOTAL_BILIRUBIN

TB	N	Mean	SD	SE	Min	Max
1	100	.802600	.2417262	.0241726	.4600	2.2000
2	71	.705493	.2254569	.0267568	.4000	2.2000
Total	171	.762281	.2392932	.0182992	.4000	2.2000

TABLE LVI
ANOVA ANALYSIS OF TB BETWEEN GENDER

TB	SS	df	MS	F	Sig.
Between Groups	.392	1	.392	7.082	.009
Within Groups	9.343	169	.055		
Total	9.734	170			

TABLE LVII
DESCRIPTIVE STATISTIC OF SGOT

SGOT	N	Mean	SD	SE	Min	Max
1	100	34.14	14.205	1.420	.1000	19.7000
2	71	26.11	9.736	1.155	.1000	4.6900
Total	171	30.81	13.125	1.004	.1000	19.7000

TABLE LVIII
ANOVA ANALYSIS OF SGOT BETWEEN GENDER

SGOT	SS	df	MS	F	Sig.
Between Groups	2675.493	1	2675.493	16.991	.000
Within Groups	26611.139	169	157.462		
Total	29286.632	170			

The Significant attributes ranges are shown in Table LVI, LVIII contain the ranges are TB (0.46-2.20 mg/dl) of male and (0.40-2.20mg/dl) of female SGOT (13-97 U/L) of male and (10-60 U/L) of female ranges of Non -liver disease patients. The ranges are different in male to female liver disease patients. This corroborates that gender is important in diagnosing Non liver disease patients, gender in (21-40) age group instrumental with above Total_Bilirubin, SGOT significant attributes. The insignificant attributes are DB, Alkphos, ,SGPT, Serum albumin ,A/G Ratio The insignificant values are 0.0547, 0.789 ,0.0647,0.0598,0.079 respectively the above those values are greater than 0.05 at $p > 0.05$.Thus, the null hypothesis is rejected. This confirms that there is insignificance among the groups that is, we can say that the attributes are not instrumental in diagnosing non- liver disease patients with (21-40) age group in male and female population

TABLE LIX
DESCRIPTIVE STATISTIC OF DIRECT_BILIRUBIN

DB	N	Mean	SD	SE	Min	Max
1	307	1.637687E0	2.3426046	.1336995	.10	19.70
2	148	1.161081E0	2.2201487	.1824951	.10	4.69
Total	455	1.482659E0	2.3118736	.1083823	.10	19.70

TABLE LX
ANOVA ANALYSIS OF DB BETWEEN GENDER

DB	SS	df	MS	F	Sig.
Between Groups	22.683	1	22.683	4.275	.039
Within Groups	2403.837	453	5.306		
Total	2426.521	454			

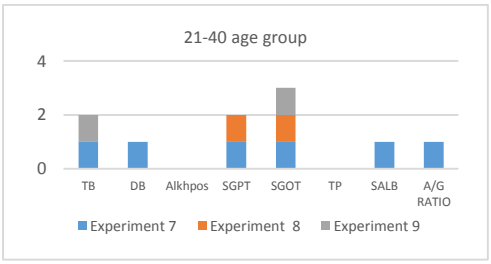

Fig. 3. Significant attributes of 21-40 age group with gender

In the above three experiments (liver &non-liver disease, liver disease, non-liver disease it is observed that (21-40) age group patients, the common significant attribute is SGOT. This corroborates that gender difference with age group is important in diagnosing liver disease, gender with (21-40) age group is instrumental with SGOT.

Experiment- 10

Liver & non-liver disease patients with gender (41-60)age group

This *Experiment- 10* includes the liver disease and non-liver disease patients from both data sets and comparison of male with female liver disease patients and non- liver disease patients. In the *Experiment- 10* (41-60)age group of dataset contains 445 total records (307 male and 148 female liver disease patients' records). ANOVA analysis for each of attributes is given in Table I. After the *Experiment- 10* the results are obtained for all attributes. In that some attributes are significant, those significant attribute results are given in Table LIX to Table LXIV. It shows ANOVA analysis of significant attributes is DIRECT_BILIRUBIN, Serum albumin, A/G ratio. After the *Experiment- 10* the results are obtained in all attributes. In that some attributes are significance, those significance attribute results are given in Table LIX to Table LXIV. It shows ANOVA analysis of significant attributes is Direct_Bilirubin, Serum Albumin, A/G ratio.

TABLE LXI
DESCRIPTIVE STATISTIC OF SERUM ALBUMIN

SALB	N	Mean	SD	SE	Min	Max
1	307	3.293811E	.6983600	.0398575	1.50	5.50
2	148	3.475676E	.6429758	.0528523	1.00	4.70
Total	455	3.352967E	.6854586	.0321348	1.00	5.50

TABLE LXII
ANOVA ANALYSIS OF SALB BETWEEN GENDER

SALB	SS	df	MS	F	Sig.
Between Groups	3.303	1	3.303	7.124	.008
Within Groups	210.011	453	.464		
Total	213.313	454			

TABLE LXIII
DESCRIPTIVE STATISTIC OF A/G RATIO

A/G	N	Mean	SD	SE	Min	Max
1	306	1.033824E0	.3153048	.0180248	.1000	19.70
2	147	1.101020E0	.2708756	.0223414	.1000	4.69
Total	453	1.055629E0	.3029470	.0142337	.1000	19.70

TABLE LXIV
ANOVA ANALYSIS OF A/G RATIO BETWEEN GENDER

A/G RATIO	SS	df	MS	F	Sig.
Between Groups	.448	1	.448	4.928	.027
Within Groups	41.035	451	.091		
Total	41.483	452			

The Significant values are shown in Tables LX, LXII,LXIV of Direct_Bilirubin, Serum Albumin, A/G Ratio are 0.039, 0.008, 0.027 which is less than0.05at $p < 0.05$. Thus, the null hypothesis is cannot be rejected. This confirms that there is significance difference among the groups that is, we can say that, DIRECT_BILIRUBIN, SERUM ALBUMIN, A/G RATIO attributes are instrumental in diagnosing liver disease and Non-liver disease patients in (41-60)age group male and female population.

The Significant attributes ranges are shown in Table LIX, LXII, LXIII contains the ranges are DB(10-19.7 mg/dl) of male and(10-4.69 mg/dl) of female Serum albumin (1.5-5.5 mg/dl) of male and (1.0- 4.7 mg/dl) of female A/G Ratio (0.3- 2.8) of male and (0.3-1.8) of female ranges of liver disease and non-liver disease patients. The ranges are different in male to female liver disease patients. This corroborates that gender is important in diagnosing liver disease and non-liver disease patients, instrumental with above Direct_Bilirubin, Serum Albumin, and A/G Ratio significant attributes. The insignificant attributes are TB, Alkphos, SGOT, SGPT, and Total_Protien. The insignificant values are 0.0647, 0.0598, 0.079, 0.057 and 0.232 respectively the above those values are greater than 0.05 at $p > 0.05$.Thus, the null hypothesis is rejected .This confirms that there is insignificance among the groups that is, we can say that the attributes are not instrumental in diagnosing liver disease and non- liver disease in (41-60) age group male and female population.

Experiment-11

Liver disease patients with gender (41-60) age group

This *Experiment-11* includes the liver disease patients selectively from both data sets and compared of male with female liver disease patients.

In the *Experiment- 11* (41-60)age group of data set contains 260 total records (182 male and 78 female liver disease patients' records). ANOVA analysis for each of attributes is given in Table I. After the *Experiment- 11* results are obtained for 8 attributes. In that all attributes are not significance. This confirms that there is no significance difference among the groups that is it can be said that

gender with age group (41-60) is not instrumental in diagnosing liver disease in male and female population, gender with age is not instrumental with all 8 attributes (TB, DB, Alkphos, SGOT, SGPT, Serum albumin, Total_ protein, A/G Ratio) attributes.

Experiment- 12

Non-liver disease patients with gender (41-60) age group

This *Experiment- 12* includes the non-liver disease patients selectively from both data sets and compared of male with female non-liver disease patients. For this *Experiment- 12* (41-60) age group of dataset contains 195 total records (111 male and 84 female non- liver disease patients' records). ANOVA analysis for each of attributes is given in Table I. After the *Experiment- 12* the results are obtained in 8 attributes. In that some attributes are significance, those significance attribute results are given in Table LXV to Table LXXIV by ANOVA analysis significant attributes found are Total Bilirubin, Direct bilirubin, Alkphos, SGPT, and SGOT.

TABLE LXV
DESCRIPTIVE STATISTIC OF TOTAL_BILIRUBIN

TB	N	Mean	SD	SE	Min	Max
1	111	.820541	.320966	.0304648	.1000	19.70
2	84	.734405	.1550542	.0169178	.1000	4.69
Total	195	.783436	.2655708	.0190179	.1000	19.70

TABLE LXVI
ANOVA ANALYSIS OF TB BETWEEN GENDER

TB	SS	df	MS	F	Sig.
Between Groups	.355	1	.355	5.137	.025
Within Groups	13.328	193	.069		
Total	13.682	194			

TABLE LXVII
DESCRIPTIVE STATISTIC OF DIRECT_BILIRUBIN

DB	N	Mean	SD	SE	Min	Max
1	111	.283333	.1900016	.0180342	.10	1.60
2	84	.252143	.1077041	.0117515	.10	.76
Total	195	.269897	.1602253	.0114740	.10	1.60

TABLE LXVIII
ANOVA ANALYSIS OF DB BETWEEN GENDER

DB	SS	df	MS	F	Sig.
Between Groups	.047	1	.047	1.820	.179
Within Groups	4.934	193	.026		
Total	4.980	194			

TABLE LXIX
DESCRIPTIVE STATISTIC OF ALKPHOS

ALKPHOS	N	Mean	SD	SE	Min	Max
1	111	152.00	93.388	8.864	63	901
2	84	130.35	37.429	4.084	82	250
Total	195	142.67	75.233	5.388	63	901

TABLE LXX
ANOVA ANALYSIS OF ALKPHOS BETWEEN GENDER

ALKPHOS	SS	df	MS	F	Sig.
Between Groups	22422.007	1	22422.007	4.023	.046
Within Groups	1075612.988	193	5573.124		
Total	1098034.995	194			

TABLE LXXI
DESCRIPTIVE STATISTIC OF SGPT

SGPT	N	Mean	SD	SE	Min	Max
1	111	35.01	15.094	1.433	14	80
2	84	30.77	12.826	1.399	10	71
Total	195	33.18	14.282	1.023	10	80

TABLE LXXII
ANOVA ANALYSIS OF SGPT BETWEEN GENDER

SGPT	SS	df	MS	F	Sig.
BetweenGroups	857.660	1	857.660	4.275	.040
Within Groups	38715.693	193	200.599		
Total	39573.354	194			

TABLE LXXIII
DESCRIPTIVE STATISTIC OF SGOT

SGOT	N	Mean	SD	SE	Min	Max
1	111	35.23	17.803	1.690	13	125
2	84	29.98	16.459	1.796	12	78
Total	195	32.96	17.390	1.245	12	125

TABLE LXXIV
ANOVA ANALYSIS OF SGOT BETWEEN GENDER

SGOT	SS	df	MS	F	Sig.
Between Groups	1317.427	1	1317.427	4.434	.037
Within Groups	57349.322	193	297.147		
Total	58666.749	194			

The Significant values are shown in Tables LXVI, LXVIII ,LXX, LXXII,LXXIV of Total_Bilirubin, DIRECT_BILIRUBIN, ALKPHOS, SGPT, SGOT are 0.025,0.179,0.046,0.040,0.037 which is less than0.05at $p < 0.05$ Thus, the null hypothesis is cannot be rejected. This confirms that there is significance difference among the groups that is, we can say that, Total_Bilirubin, Direct_Bilirubin, , Alkphos, SGOT, SGPT attributes are instrumental in diagnosing non-liver disease patients in (41-60)age group male and female population.

The Significant attributes ranges are shown in tables LXV, LXVII, LXIX, LXXI, LXXIII contains the ranges are TB(0.4-2.9mg/dl) of male and (0.4-1.4 mg/dl) of female, DB(0.1-1.6mg/dl) of male and (0.1-0.76 mg/dl) of female ,Alkphos(63-901U/L) of male and (82-250 U/L) of female SGPT(14-80U/L) of male and (10-71 U/L) of female, SGOT(13-125 U/L) of male and (12-78 U/L) ranges of non-liver disease patients. The ranges are different in male to female liver disease patients. This corroborates that gender is important in diagnosing liver disease patients, gender in 41-60 age group instrumental with Above Total_Bilirubin, Direct_Bilirubin, Alkphos, SGPT, SGOT Significant attributes. The non-significant attributes are Total_Protien, Serum albumin, A/G Ratio. The insignificant values are 0.00610, 0.057, 0.232 respectively the above those values are greater than 0.05 at $p > 0.05$.Thus, the null hypothesis is rejected. This confirms that there is insignificance among the groups that is, we can say that the attributes are not instrumental in diagnosing non- liver disease in male and female population.

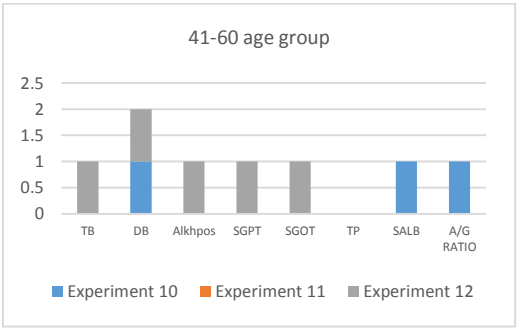

Fig. 4. Significant attributes of 41-60 age group with gender

In the above three Experiments (liver &non-liver disease, liver disease, non-liver disease) it is observed that in (41-60)age group patients, the common significant attribute is Direct_Bilirubin. This corroborates that gender difference with age is important in diagnosing liver disease, Gender with Age group (41-60) is instrumental with Direct_Bilirubin.

Experiment- 13
Liver & non-liver disease patients with gender 61-80 age group

This *Experiment- 13* includes the liver disease and non-liver patients of both data sets and comparison of male liver disease patients with female liver disease patients. In the *Experiment-13* (61-80) age group dataset contains133 records (83 male and 50 female liver disease patients' records.) ANOVA analysis for each of attributes is given in Table I. After the *Experiment- 13* the results are obtained for all attributes. In that all attributes are not significant. This confirms that there is no significant difference among the groups that is gender is not instrumental in diagnosing liver disease and non-liver disease patents in male and female population. Especially during development of automatic medical diagnosis, gender is not instrumental with all attributes.

Experiment- 14
Liver disease patients with gender 61-80 age group

In this *Experiment- 14* includes the selectively liver disease patients of both data sets and compare

male liver disease patients and female liver disease patients. In the *Experiment- 14* (61-80) age group of dataset contains 87 total records, in that 55 male liver disease patients, 32 female liver disease patients' records. We have carried out one way analysis of variance (ANOVA) for each of attributes in given Table I. after the *Experiment- 14;* the results are obtained in all attributes. In that all attributes are not significance. This confirms that there is no significance difference among the groups that is we can say that gender is not instrumental in diagnosing liver disease in male and female population, gender is not instrumental with all attributes.

Experiment- 15

Non-liver disease patients with gender 61-80 age group

This *Experiment- 15* includes the selectively non-liver disease patients of both data sets and compare male non-liver disease patients and female non-liver disease patients. In the *Experiment- 15* (61-80)age group of dataset contains 45 total records, in that 29 male non- liver disease patients, 16 female non- liver disease patients' records. We have carried out one way analysis of variance (ANOVA) for each of attributes in given Table I. after the *Experiment- 15*; the results are obtained in all attributes. In that some attributes are significance, those significance attribute results are given in Table LXXV to Table LXXVI. It shows ANOVA analysis of significant attributes is Total_Protien.

TABLE LXXV
DESCRIPTIVE STATISTIC OF TOTAL_PROTIEN

TP	N	Mean	SD	SE	Min	Max
1	29	6.331034E0	.6772336	.1257591	.10	1.60
2	16	6.875000E0	.5170429	.1292607	.10	.76
Total	45	6.524444E0	.6725648	.1002600	.10	1.60

TABLE LXXVI
ANOVA ANALYSIS OF TP BETWEEN GENDER

TP	SS	df	MS	F	Sig.
Between Groups	3.051	1	3.051	7.785	.008
Within Groups	16.852	43	.392		
Total	19.903	44			

Tables LXXV, LXXVI contains ANOVA result of Total_Protien. The significant value 0.008 in Table LXXVI is smaller than 0.05 at $p < 0.05$.Thus, the null hypothesis is cannot be rejected. This confirms that there is significance difference among the groups that is, we can say that TOTAL_PROTIEN is instrumental in diagnosing non-liver disease in male and female population. Table LXXV contains the ranges (0.10- 1.60 gm/dl) of male and (0.10-0.76 gm/dl) of female non- liver disease patients. The ranges are different in male to female non-liver disease patients. This corroborates that gender is important in diagnosing liver disease, gender with age (61-80)group is instrumental with Total_Protien.

In the above three Experiments (liver &non-liver disease, liver disease, non-liver disease) it is observed that Total_Protien. is common significant attribute in (61-80) age group patients This corroborates that gender difference with age group is important in diagnosing liver disease, gender with age group (61-80) is instrumental with Total_Protien.

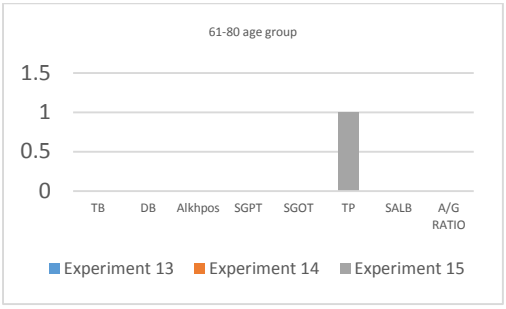

Fig. 5.significant attributes of 61-80 age group with gender

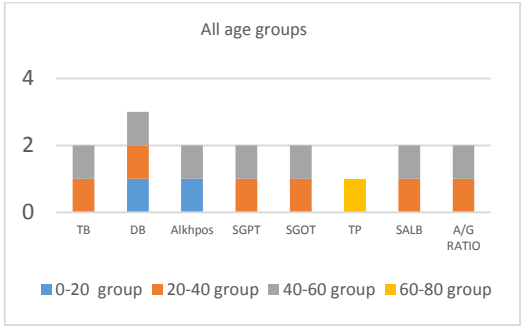

Fig. 6. Significant attributes of all age groups with gender

As given in the above graph fig.6, combine report of significant attributes of all above four age groups. This corroborates that gender is important in diagnosing liver diseases with respective attributes.

VII. CONCLUSIONS

In this study Age, Gender, Total Bilirubin, Direct Bilirubin, Alkphos, SGPT, SGOT, Total_Protiens, Albumin and A/G Ratio of the ILPD dataset are considered for one way Analysis of Variance (ANOVA). The dataset is analyzed to know the effect of gender and age disparities in causing liver disorders. Initially effect of gender disparity in three populations namely liver & non-liver patients, liver patients and non-liver patients have been considered and the results of ANOVA analysis on these three populations show that there exists more gender disparity. Later the effect of aging on liver functioning with respect to gender is considered and the dataset has been divided into four populations (0-20, 21-40, 41-60 and 61-80) based on the age. Each population is subdivided into three subpopulations namely liver & non-liver patients, liver patients and non-liver patients. Then effect of gender disparity in each subpopulation is analyzed using ANOVA. The results of the analysis on population (0-20) show that the values of DB and Alkphos are more significant to decide the liver functionality in that age group in male as well as female patients. Similarly values SGOT, DB and TP are more significant in the population (21-40), population (41-60) and population (61-80) respectively.

Results of this study are very important while developing automatic medical diagnosis systems as it validates the necessity of reference values with respect to gender and age group settings of the software based system. Also enables Gastroenterologists to be aware about these age and gender differences among liver patients and prescribe drugs accordingly.

ACKNOWLEDGMENTS

We sincerely thank the expert Gastroenterologists Dr.Srinivas Rao and Dr.SrinivasBaba for their highly valuable contribution and cooperation

REFERENCES

[1] Decision Factors on Effective Liver Patient Data Prediction" Hoon Jin1, Seoungcheon Kim2 and Jinhong Kim3, International Journal of Bio-Science and Bio-Technology Vol.6, No.4 (2014), pp.167-178.

[2] Sotoudehamanesh R, Sotoudeh M, Asgari A, Abedi-Ardakani B, Tavangar SM,Khakinejad A et al. Silent Liver Diseases in Autopsies from Forensic Medicine of Tehran. Archives of Iranian Medicine 2006 Oct; 9(4):324-28.

[3] Pathological Findings of Liver in Autopsy Cases A Study at Imphal"Ph. Madhubala Devi and Barida Ginia Myrthong, et al *Indian Acad Forensic Med. July-September 2013, Vol. 35, No.3*

[4] Sultana Furruqh, D. Anitha and T. Venkatesh." estimation of reference values in liver function test in health plan individuals of an urban south Indian population Indian Journal of Clinical Biochemistry, 2004, 19 (2) 72-79

[5] .A Critical Comparative Study of Liver Patients from USA and INDIA: An Exploratory Analysis" Bendi Venkata Ramana1, Prof. M. Surendra Prasad Babu2, Prof. N. B. Venkateswarlu3, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 2, May 2012, ISSN (Online): 16940814.

[6] Liddell A, Locker D: Gender and age differences in attitudes to dental pain and dental control. Community Detit Oral Epidemiol 1997; 25: 314-8. © Munksgaard 1997

[7] William A. Sadera and Scot W. McNary." Comparing student success between developmental math courses offered online, blended, and face-to-face." Volume 10, Number 3, Winter 2011 ISSN: 1541-4914.

[8] "Sexual dimorphism in house sparrow eggs" Pedro J. Cordero · Simon C. Griffith

[9] José M. Aparicio · David T. Parkin, Behav Ecol Sociobiology (2000) 48:353–357, © Springer-Verlag 2000.

[10] Benton, A.L. Differential behavioral effects in Frontal lobe disease. Neuropsychological, 6, 53–60. (1968).

[11] "Research Methodology, methods and techniques third edition" C.R Kothari Gaurav Garg

[12] "Hypothesis Testing - Analysis of Variance (ANOVA" tutorial from Boston University.

[13] " Impact of Outlier Removal and Normalization Approach in Modified k-Means Clustering Algorithm" Vaishali R. Patel1 and Rupa G. Mehta2- IJCSI International Journ8al of Computer Science Issues, Vol. 8, Issue 5, No 2, September 2011

[14] N. Karthikeyani Visalakshi and K. Thangavel,"Impact of Normalization in Distributed K-Means Clustering", Internationa l Journal of Soft Computing 4, Vol. 4, 2009, pp.168-172.

[15] The Indian liver patient dataset (ILPD)is from UCI machine repository in the area of life science. The ILPD data set is available in following hyper link http://archive.ics.uci.edu/datasets/ILPD+(indian+liver+patient+Dataset)

[16] Levine's" Guide to SPSS for Analysis of Variance 2nd Edition "Melanie C. Page Oklahoma State University Sanford L. Braver Arizona State University David P. MacKinnon Arizona State University.