



# Indian Liver Patient Analysis

---

Group 6:  
Chih Hao Yuan  
Ching Yu Hsu



# Table of contents

01

**Visualization**

02

**Correlation**

03

**Logit**

04

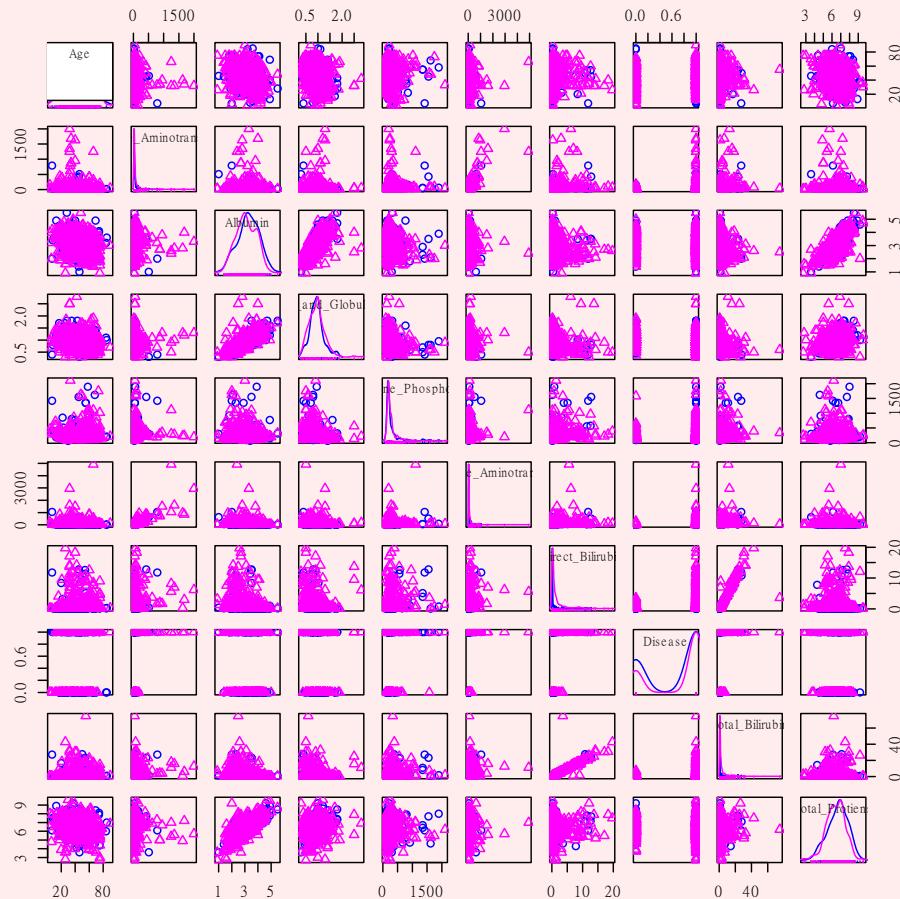
**Neural network – binary output**

05

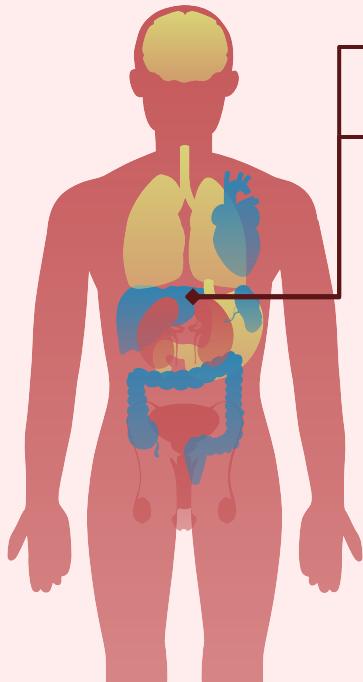
**List of lesson learned**



## 1(a)- Overall view of relationship of Disease with all continuous variables



# 1(b)-Highlight at least two graphs where there are strong relationships between pairs of continuous variables

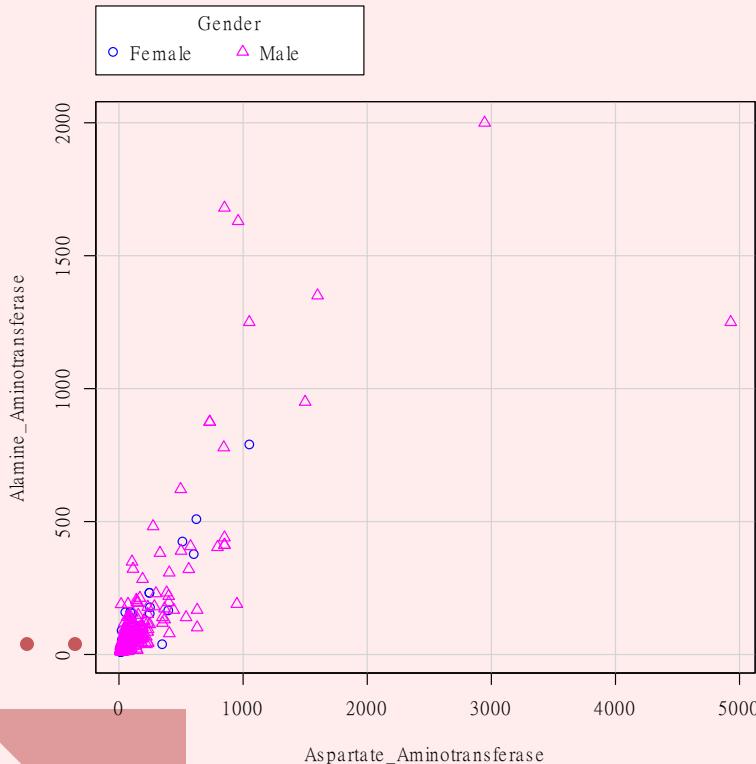


	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens	Albumin	Albumin_and_Globulin_Ratio	Disease
Age	1										
Gender	0.0559	1									
Total_Bilirubin	0.011	0.0881	1								
Direct_Bilirubin	0.0068	0.0992	0.8745	1							
Alkaline_Phosphotase	0.0789	-0.0294	0.2057	0.2332	0.1248	1					
Alamine_Aminotransferase	-0.0878	0.0813	0.2134	0.2332	0.1248	0.1666	1				
Aspartate_Aminotransferase	-0.0205	0.0794	0.2373	0.257	0.1666	0.7919	1				
Total_Protiens	-0.1862	-0.0951	-0.0079	3E-05	-0.0271	-0.0424	-0.0258	1			
Albumin	-0.2642	-0.0956	-0.2221	-0.2284	-0.1634	-0.0287	-0.0849	0.7831	1		
Albumin_and_Globulin_Ratio	-0.2164	-0.0034	-0.2063	-0.2001	-0.2342	-0.0024	-0.07	0.2349	0.6896	1	
Disease	0.1332	0.0813	0.2202	0.2463	0.1834	0.1631	0.1518	-0.0336	-0.1598	-0.1631	1

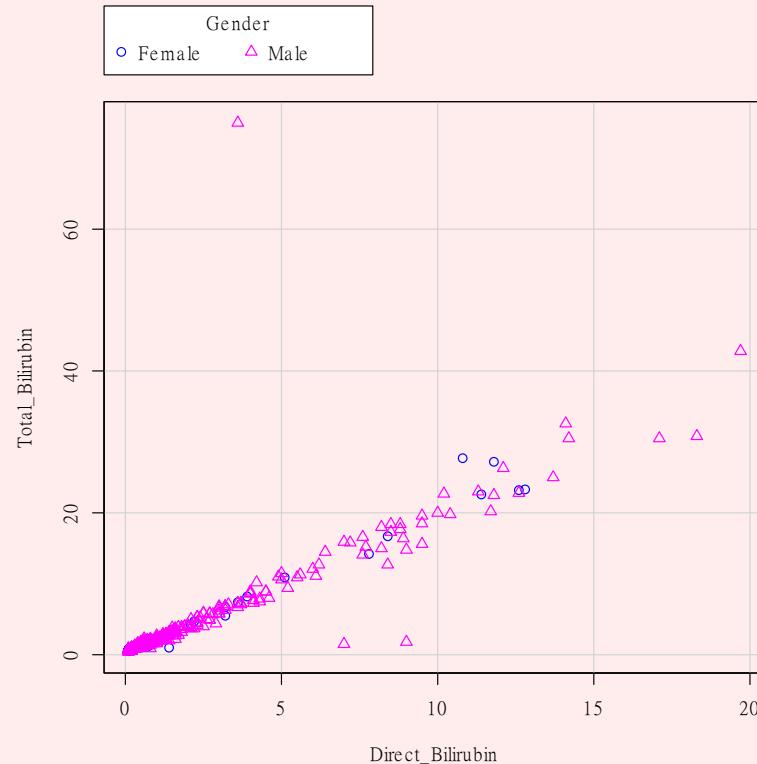
## Strong Relationship :

1. Direct Bilirubin & Total Bilirubin
2. Aspartate Aminotransferase & Alamine Aminotransferase

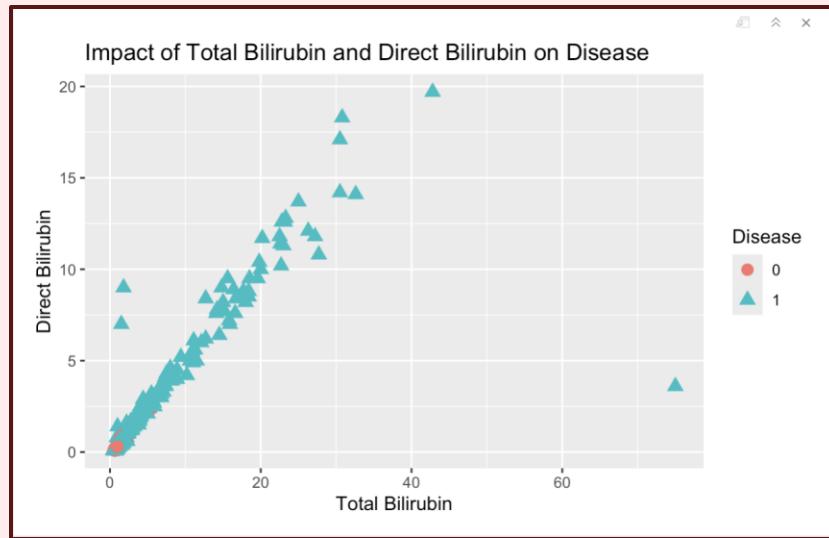
## Alamine Aminotransferase & Aspartate Aminotransferase



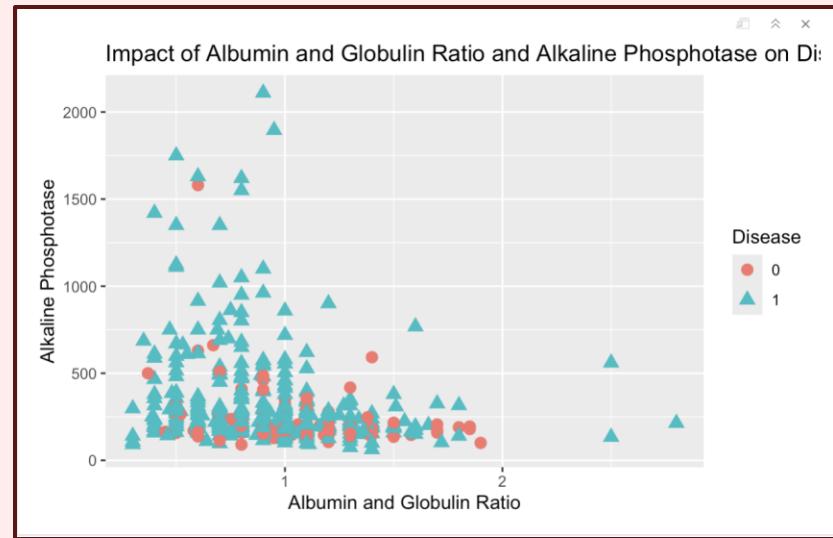
## Total Bilirubin & Direct Bilirubin



# 1(c)- The graph with pairs continuous variables and plot Disease



Total Proteins & Total Bilirubin



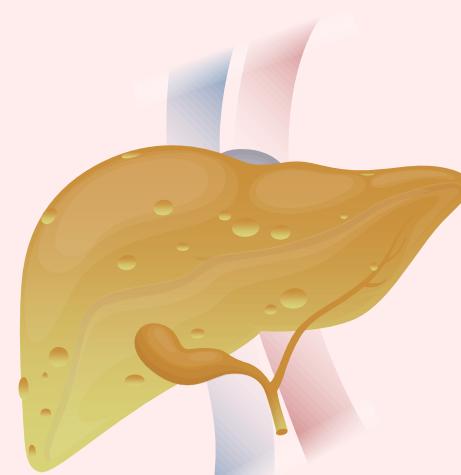
Albumin and Globulin Ratio & Alkaline Phosphotase

## 2(a)- Correlation Analysis on all variables

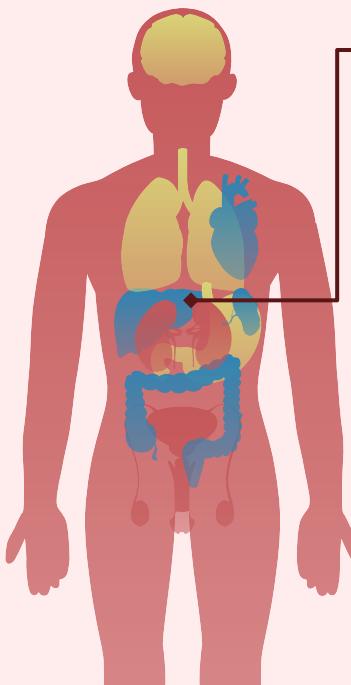
	Age	Alamine_Aminotransferase	Albumin	Albumin_and_Globulin_Ratio
Age	1.000000000	-0.08779916	-0.26421094	-0.216408346
Alamine_Aminotransferase	-0.087799162	1.000000000	-0.02865750	-0.002374990
Albumin	-0.264210935	-0.02865750	1.000000000	0.689632342
Albumin_and_Globulin_Ratio	-0.216408346	-0.00237499	0.68963234	1.000000000
Alkaline_Phosphotase	0.078878350	0.12477671	-0.16341865	-0.234166499
Aspartate_Aminotransferase	-0.020498946	0.79186215	-0.08491457	-0.070039828
Direct_Bilirubin	0.006784303	0.23318008	-0.22840915	-0.200124685
Disease	0.133163583	0.16311694	-0.15976956	-0.163131363
Gender	0.055880710	0.08133871	-0.09557894	-0.003424034
Total_Bilirubin	0.011000374	0.21337549	-0.22208657	-0.206267186
Total_Protiens	-0.186248122	-0.04243210	0.78311217	0.234887181

	Alkaline_Phosphotase	Aspartate_Aminotransferase	Direct_Bilirubin	Disease
Age	0.07887835	-0.02049895	0.00678430344	0.13316358
Alamine_Aminotransferase	0.12477671	0.79186215	0.23318008266	0.16311694
Albumin	-0.16341865	-0.08491457	-0.22840915308	-0.15976956
Albumin_and_Globulin_Ratio	-0.23416650	-0.07003983	-0.20012468525	-0.16313136
Alkaline_Phosphotase	1.000000000	0.16657999	0.23400757200	0.18336330
Aspartate_Aminotransferase	0.16657999	1.000000000	0.25702239483	0.15183446
Direct_Bilirubin	0.23400757	0.25702239	1.00000000000	0.24627293
Disease	0.18336330	0.15183446	0.24627292852	1.00000000
Gender	-0.02936768	0.07942099	0.09916008108	0.08134945
Total_Bilirubin	0.20573917	0.23732305	0.87448096885	0.22021794
Total_Protiens	-0.02706202	-0.02575101	0.00003270877	-0.03361377

	Gender	Total_Bilirubin	Total_Protiens
Age	0.055880710	0.011000374	-0.18624812163
Alamine_Aminotransferase	0.081338710	0.213375493	-0.04243209532
Albumin	-0.095578941	-0.222086570	0.78311217012
Albumin_and_Globulin_Ratio	-0.003424034	-0.206267186	0.23488718111
Alkaline_Phosphotase	-0.029367677	0.205739173	-0.02706201547
Aspartate_Aminotransferase	0.079420992	0.237323055	-0.02575100745
Direct_Bilirubin	0.099160081	0.874480969	0.00003270877
Disease	0.081349445	0.220217940	-0.03361376899
Gender	1.000000000	0.088067618	-0.09514928561
Total_Bilirubin	0.088067618	1.000000000	-0.00790592347
Total_Protiens	-0.095149286	-0.007905923	1.00000000000



## 2(b)- The strongest correlation with Disease



	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens	Albumin	Albumin_and_Globulin_Ratio	Disease
Age	1										
Gender	0.0559	1									
Total_Bilirubin	0.011	0.0881	1								
Direct_Bilirubin	0.0068	0.0992	0.8745	1							
Alkaline_Phosphotase	0.0789	-0.0294	0.2057	0.234	1						
Alamine_Aminotransferase	-0.0878	0.0813	0.2134	0.2332	0.1248	1					
Aspartate_Aminotransferase	-0.0205	0.0794	0.2373	0.257	0.1666	0.7919	1				
Total_Protiens	-0.1862	-0.0951	-0.0079	3E-05	-0.0271	-0.0424	-0.0258	1			
Albumin	-0.2642	-0.0956	-0.2221	-0.2284	-0.1634	-0.0287	-0.0849	0.7831	1		
Albumin_and_Globulin_Ratio	-0.2164	-0.0034	-0.2063	-0.2001	-0.2342	-0.0024	-0.07	0.2349	0.6896	1	
Disease	0.1332	0.0813	0.2202	0.2463	0.1834	0.1631	0.1518	-0.0336	-0.1598	-0.1631	1

**Strongest Relationship :**  
➤ Direct Bilirubin (0.2463)



## 3(a)- Logit Analysis on Disease with all explanatory variables

Call:

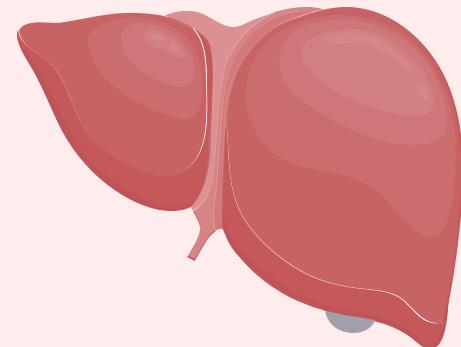
```
glm(formula = Disease ~ ., family = "binomial", data = indian_liver_patient)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-3.6046430	1.3296884	-2.711	0.00671 **
Age	0.0182922	0.0063783	2.868	0.00413 **
Gender	0.0309928	0.2328445	0.133	0.89411
Total_Bilirubin	0.0112271	0.0869355	0.129	0.89724
Direct_Bilirubin	0.4496843	0.2414926	1.862	0.06259 .
Alkaline_Phosphotase	0.0012701	0.0008079	1.572	0.11594
Alamine_Aminotransferase	0.0119298	0.0051001	2.339	0.01933 *
Aspartate_Aminotransferase	0.0024729	0.0031198	0.793	0.42798
Total_Protiens	0.9273824	0.3814252	2.431	0.01504 *
Albumin	-1.6951386	0.7480766	-2.266	0.02345 *
Albumin_and_Globulin_Ratio	1.8080515	1.1408815	1.585	0.11302
---				
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				

### 3(b)- The explanatory variables were significant

```
Call:  
glm(formula = Disease ~ Age + Alamine_Aminotransferase + Albumin +  
    Total_Protiens, family = "binomial", data = indian_liver_patient)  
  
Coefficients:  
Estimate Std. Error z value Pr(>|z|)  
(Intercept) -1.412616 0.759629 -1.860 0.062940 .  
Age 0.018687 0.006351 2.942 0.003258 **  
Alamine_Aminotransferase 0.020619 0.004073 5.063 4.13e-07 ***  
Albumin -0.988552 0.227535 -4.345 1.40e-05 ***  
Total_Protiens 0.576845 0.164064 3.516 0.000438 ***  
---  
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

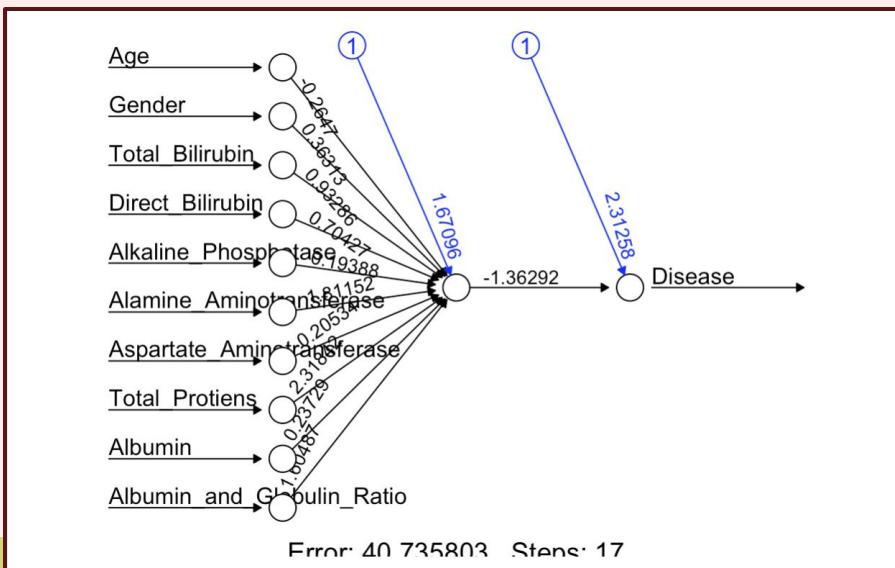


#### Direction :

- Positive impact : Age, Alamine Aminotransferase, Total Protiens
- Negative impact : Albumin

## 4(a) Neural networks with binary output Disease - with all the explanatory variables

### ▶ Hidden Nodes = 1 (Model 1)



Confusion Matrix and Statistics	
Reference	
Prediction	0 1
0	0 0
1	52 122
Accuracy : 0.7011	
95% CI : (0.6272, 0.7681)	
No Information Rate : 0.7011	
P-Value [Acc > NIR] : 0.5374	
Kappa : 0	
McNemar's Test P-Value : 1.522e-12	
Sensitivity : 0.0000	
Specificity : 1.0000	
Pos Pred Value : NaN	
Neg Pred Value : 0.7011	
Prevalence : 0.2989	
Detection Rate : 0.0000	
Detection Prevalence : 0.0000	
Balanced Accuracy : 0.5000	
'Positive' Class : 0	

```
```{r}
set.seed(42)
patient_index <- sample(nrow(patient_clean), 0.7 * nrow(patient_clean),
                        replace = FALSE)
patient_train <- patient_clean[patient_index, ]
patient_test <- patient_clean[-patient_index, ]
...```

```

```
```{r}
set.seed(42)
model_neural <- neuralnet(Disease ~.,
                           data = patient_train,
                           hidden = 1,
                           lfitesign = "minimal",
                           linear.output = FALSE,
                           threshold = 0.01)
...```

```

```
```{r}
result_model_neural1 <- compute(model_neural, patient_test[1:10])
...```

```

```
```{r}
df_result_model_neural1 <- data.frame(actual=patient_test$Disease,
                                         prediction=result_model_neural1$net.result)
...```

```

```
```{r}
df_result_model_neural1$round_predic <- round(df_result_model_neural1$prediction)
...```

```

```
```{r}
df_result_model_neural1

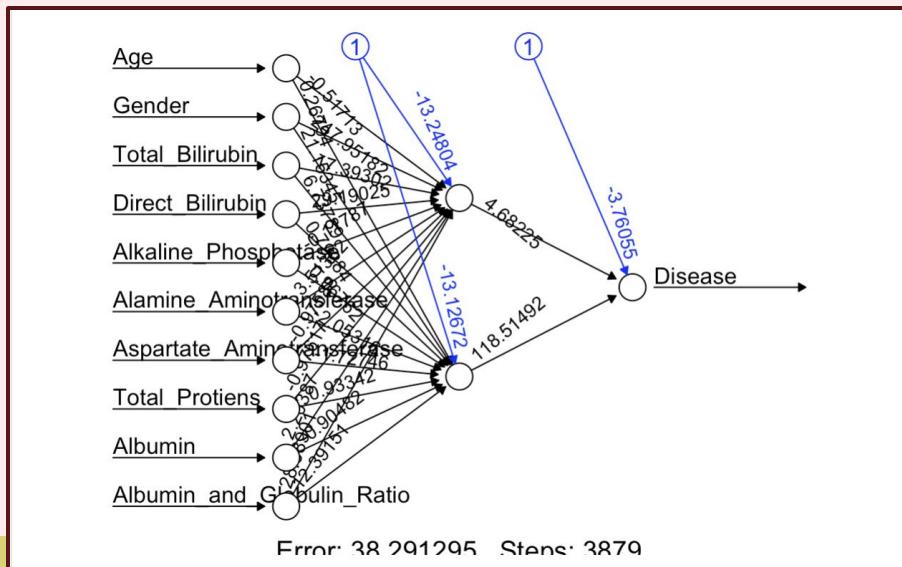
```

Description: df [174 × 3]

actual <dbl>	prediction <dbl>	round_predic <dbl>
1	0.7210467	1
1	0.7210467	1
1	0.7210467	1
1	0.7210467	1
0	0.7210467	1
1	0.7210467	1
0	0.7210467	1
1	0.7210467	1
0	0.7210467	1
1	0.7210467	1

## 4(a) Neural networks with binary output Disease - with all the explanatory variables

### ▶ Hidden Nodes = 2 (Model 2)



Confusion Matrix and Statistics

		Reference	
		0	1
Prediction	0	2	2
	1	50	120

Accuracy : 0.7011

95% CI : (0.6272, 0.7681)

No Information Rate : 0.7011

P-Value [Acc > NIR] : 0.5374

Kappa : 0.03

Mcnemar's Test P-Value : 7.138e-11

Sensitivity : 0.03846

Specificity : 0.98361

Pos Pred Value : 0.50000

Neg Pred Value : 0.70588

Prevalence : 0.29885

Detection Rate : 0.01149

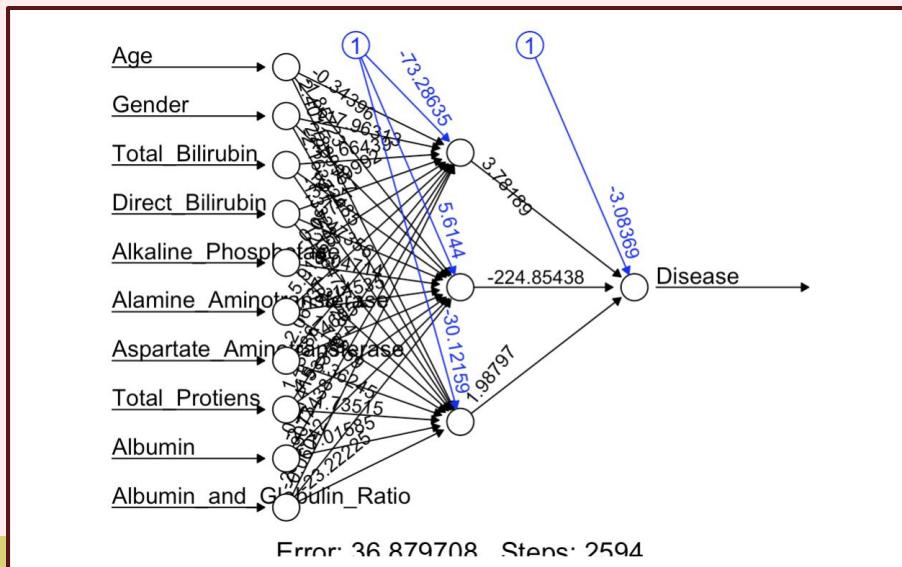
Detection Prevalence : 0.02299

Balanced Accuracy : 0.51103

'Positive' Class : 0

## 4(a) Neural networks with binary output Disease - with all the explanatory variables

### ▶ Hidden Nodes = 3 (Model 3)



Confusion Matrix and Statistics

		Reference	
		Prediction	0 1
Prediction	0	2 1	
	1	50 121	

Accuracy : 0.7069  
95% CI : (0.6333, 0.7733)

No Information Rate : 0.7011  
P-Value [Acc > NIR] : 0.4714

Kappa : 0.0415

Mcnemar's Test P-Value : 1.801e-11

Sensitivity : 0.03846  
Specificity : 0.99180

Pos Pred Value : 0.66667  
Neg Pred Value : 0.70760

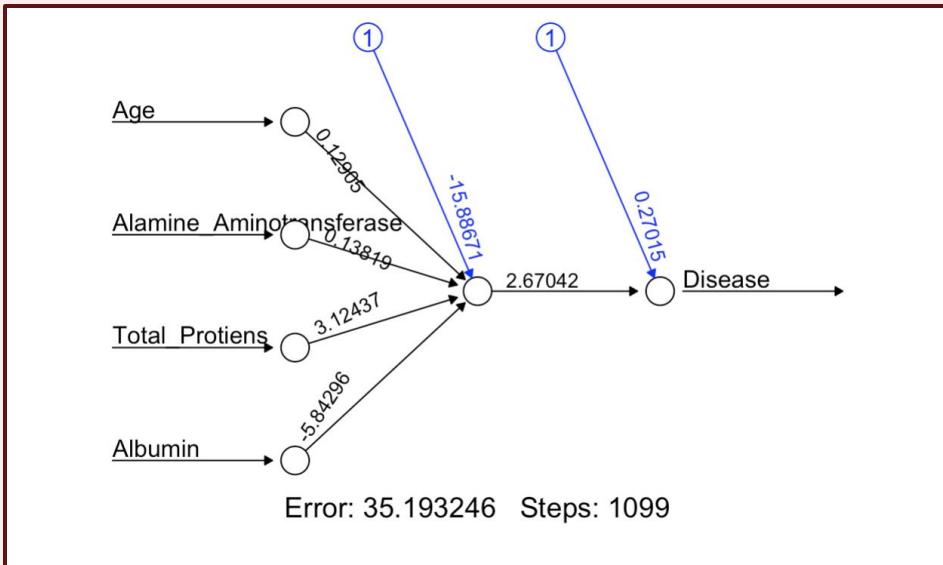
Prevalence : 0.29885  
Detection Rate : 0.01149

Detection Prevalence : 0.01724  
Balanced Accuracy : 0.51513

'Positive' Class : 0

## 4(b) Neural networks with binary output Disease - with the significant variables

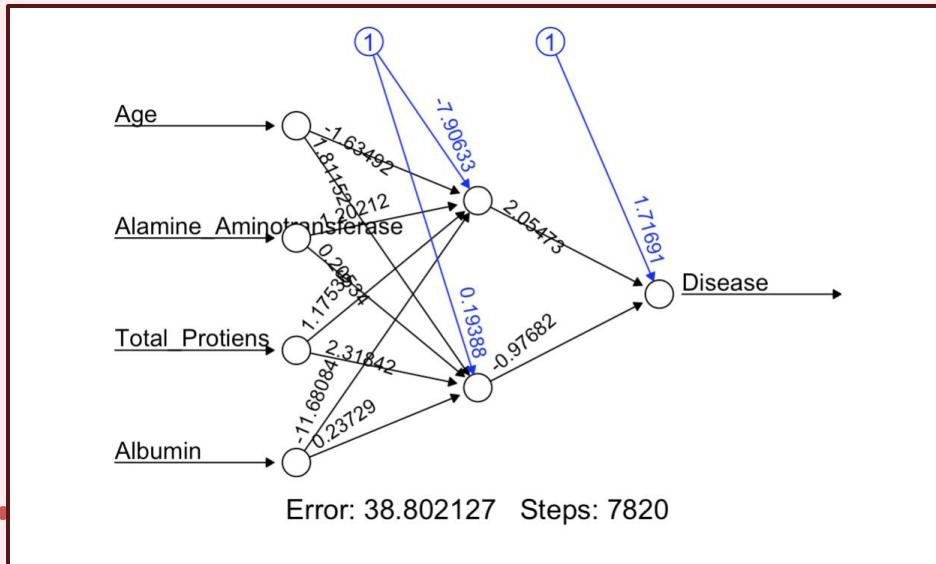
### Hidden Nodes = 1 (Model 4)



Confusion Matrix and Statistics		
	Reference	
Prediction	0	1
0	0	0
1	52	122
Accuracy : 0.7011		
95% CI : (0.6272, 0.7681)		
No Information Rate : 0.7011		
P-Value [Acc > NIR] : 0.5374		
Kappa : 0		
McNemar's Test P-Value : 1.522e-12		
Sensitivity : 0.0000		
Specificity : 1.0000		
Pos Pred Value : NaN		
Neg Pred Value : 0.7011		
Prevalence : 0.2989		
Detection Rate : 0.0000		
Detection Prevalence : 0.0000		
Balanced Accuracy : 0.5000		
'Positive' Class : 0		

## 4(b) Neural networks with binary output Disease - with the significant variables

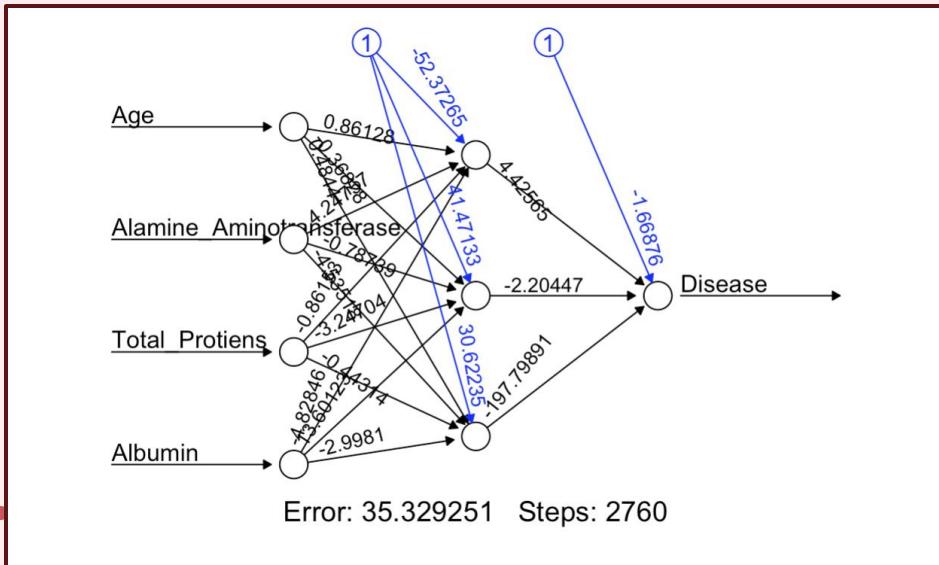
### Hidden Nodes = 2 (Model 5)



Confusion Matrix and Statistics		
Reference		
Prediction	0	1
0	0	0
1	52	122
Accuracy : 0.7011		
95% CI : (0.6272, 0.7681)		
No Information Rate : 0.7011		
P-Value [Acc > NIR] : 0.5374		
Kappa : 0		
McNemar's Test P-Value : 1.522e-12		
Sensitivity : 0.0000		
Specificity : 1.0000		
Pos Pred Value : NaN		
Neg Pred Value : 0.7011		
Prevalence : 0.2989		
Detection Rate : 0.0000		
Detection Prevalence : 0.0000		
Balanced Accuracy : 0.5000		
'Positive' Class : 0		

## 4(b) Neural networks with binary output Disease - with the significant variables

### Hidden Nodes = 3 (Model 6)



Confusion Matrix and Statistics	
Reference	
Prediction	0 1
0	7 0
1	45 122
Accuracy : 0.7414	
95% CI : (0.6697, 0.8047)	
No Information Rate : 0.7011	
P-Value [Acc > NIR] : 0.1404	
Kappa : 0.1791	
McNemar's Test P-Value : 5.412e-11	
Sensitivity : 0.13462	
Specificity : 1.00000	
Pos Pred Value : 1.00000	
Neg Pred Value : 0.73054	
Prevalence : 0.29885	
Detection Rate : 0.04023	
Detection Prevalence : 0.04023	
Balanced Accuracy : 0.56731	
'Positive' Class : 0	

# Summary the accuracy of all models

	Variable	Hidden Node	Accuracy
Model 1	All	1	0.7011
Model 2	All	2	0.7011
Model 3	All	3	0.7069
Model 4	Significant variables	1	0.7011
Model 5	Significant variables	2	0.7011
Model 6	Significant variables	3	0.7414



# 5. List of Lessons Learned



(a) Did the correlation analysis give insight into the result found in logit?

→ NO

	Age	Gender	Total_Bilirubin	Bilirubin_Phosph	Aminotransf	Aminotransf	Total_Proteins	Albumin	Albumin_and_Globulin_Ratio	Disease
Age	1									
Gender	0.0559	1								
Total_Bilirubin	0.011	0.0881	1							
Direct_Bilirubin	0.0068	0.0992	0.8745	1						
Alkaline_Phosphatase	0.0789	-0.0294	0.2057	0.234	1					
Alanine_Aminotransferase	-0.0878	0.0813	0.2134	0.2332	0.1248	1				
Aspartate_Aminotransferase	-0.0205	0.0794	0.2373	0.257	0.1666	0.7919	1			
Total_Proteins	-0.1862	-0.0951	-0.0079	3E-05	-0.0271	-0.0424	-0.0258	1		
Albumin	-0.2642	-0.0956	-0.2221	-0.2284	-0.1634	-0.0287	-0.0849	0.7831	1	
Albumin_and_Globulin_Ratio	-0.2164	-0.0034	-0.2063	-0.2001	-0.2342	-0.0024	-0.07	0.2349	0.6896	1
Disease	0.1332	0.0813	0.2202	0.2463	0.1834	0.1631	0.1518	-0.0336	-0.1598	-0.1631

```
Call:  
glm(formula = Disease ~ ., family = "binomial", data = indian_liver_patient)  
  
Coefficients:  
Estimate Std. Error z value Pr(>|z|)  
(Intercept) -3.6046430 1.3296884 -2.711 0.00671 **  
Age 0.0182922 0.0063783 2.868 0.00413 **  
Gender 0.0309928 0.2328445 0.133 0.89411  
Total_Bilirubin 0.0112271 0.0869355 0.129 0.89724  
Direct_Bilirubin 0.4496843 0.2414926 1.862 0.06259 .  
Alkaline_Phosphatase 0.0012701 0.0008079 1.572 0.11594  
Alanine_Aminotransferase 0.0119298 0.0051001 2.339 0.01933 *  
Aspartate_Aminotransferase 0.0024729 0.0031198 0.793 0.42798  
Total_Proteins 0.9273824 0.3814252 2.431 0.01504 *  
Albumin -1.6951386 0.7480766 -2.266 0.02345 *  
Albumin_and_Globulin_Ratio 1.8080515 1.1408815 1.585 0.11302  
---  
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



(b) Did the logit model help in identifying the best neural network?

→ YES

# Thanks

---

Group 6 :  
Chih Hao Yuan  
Ching Yu Hsu

