**RESULTS**

*BENAMMAR Ismail – CHERGOU Amine Ismail – HAMIDI Lylia*

**SVM: Training and Prediction**

**polynomial kernel:**

On the one hand, we have the confusion regarding the detection of infection:

|  |  |  |
| --- | --- | --- |
| T  P | Blood Donor | Infected |
| Blood Donor | 107 | 1 |
| Infected | 3 | 11 |

Table 1: confusion matrix ( P: Predicted, T: True )

On the other hand, we have the confusion matrix of the advance of the disease:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| T  P | Donor | Hepatitis | Fibrosis | Cirrhosis |
| Donor | 107 | 1 | 0 | 0 |
| Hepatitis | 1 | 3 | 0 | 0 |
| Fibrosis | 2 | 1 | 1 | 0 |
| Cirrhosis | 0 | 0 | 1 | 5 |

Table 2: confusion matrix ( P: Predicted, T: True )

Precision:We are going to calculate the detection and hit precision that the model has in each category and the *accuracy* value, which is the average of the hit percentage for each category

|  |  |  |  |
| --- | --- | --- | --- |
|  | Detection | Success | Confusion |
| Blood Donor | 99.07407 | 97.27273 | 2.727273 |
| Infected | 78.57143 | 91.66667 | 8.333333 |
| Hepatitis | 75.00000 | 60.00000 | 40.000000 |
| Fibrosis | 25.00000 | 50.00000 | 50.000000 |
| Cirrhosis | 83.33333 | 100.00000 | 0.000000 |

Table 3: detection, hit precision and accuracy of each category

### The accuracy obtained is 95%

**Result with other kernels:**

#### Radial basis:

The accuracy obtained with a radial Basis is 88.52% and It does not improve the result of the polynomial kernel.

#### Sigmoid:

The accuracy obtained with a sigmoid kernel is 88.52% and It also does not improve the result of the polynomial kernel

## **Random Forest: Training and Prediction**

The same training and validation data will be used as for the SVM model.

First, we have confusion about the detection of infections:

|  |  |  |
| --- | --- | --- |
| T  P | Blood Donor | Infected |
| Blood Donor | 107 | 1 |
| Infected | 6 | 8 |

Table 4: confusion matrix ( P: Predicted, T: True )

Then, the confusion matrix of the advance of the disease:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| T  P | Donor | Hepatitis | Fibrosis | Cirrhosis |
| Donor | 107 | 0 | 0 | 1 |
| Hepatitis | 2 | 2 | 0 | 0 |
| Fibrosis | 4 | 0 | 0 | 0 |
| Cirrhosis | 0 | 0 | 1 | 5 |

Table 5: confusion matrix ( P: Predicted, T: True )

Precision:Using the precisionMatrix function we wrote, the percentage of detection and hitting of the model is calculated. The value of accuracy is also calculated.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Detection | Success | Confusion |
| Blood Donor | 99.07407 | 94.69027 | 5.309735 |
| Infected | 57.14286 | 88.88889 | 11.111111 |
| Hepatitis | 50.000000 | 100.00000 | 0.000000 |
| Fibrosis | 0.00000 | 0.00000 | 100.000000 |
| Cirrhosis | 83.33333 | 83.33333 | 16.666667 |

Table 6: detection, hit precision and accuracy of each category

### The accuracy obtained is 93.44%

## **Neural Network:**

With the same method we’ve done with SVM and Random Forest, we obtain an accuracy (score) of 91.13%

## **Conclusions:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | SVM | R.Forest | N.Network |
| Accuracy | 95% | 93.44% | 91.13% |

Table 7: Accuracy of each model

All models have a very high prediction of negative cases of infection (107 out of 108) and it is logic because in our sample, most of the patients are not infected with hepatitis.

The false negatives in the case of SVM are 3 while for the Random Forest model they are 6. In this sense, we can say that the SVM model better predicts negative cases of HCV.

There are 14 cases of infection of which the SVM model has detected 11 and has been confused in 1, while the Random Forest model has detected 8 and has been wrong in 1. From this we can deduce that the SVM model is also better to detect cases of HCV infection.

With the above data we can better consider the SVM model than the Random Forest or Neural Network.

We have seen that the SVM model has, for non-infection cases, a detection of 99% and a confusion of 2.8% and for cases of infection a detection of 78.57% and a confusion of 8.33%.

We can say that the SVM model is good for determining virus infection, although it is not good for classifying the degree of disease. In particular for the cases of Hepatitis (which of 5 predictions has been confused in 2) and Fibrosis (which has only detected 1 of 4).

This model could never be decisive for diagnosis, but it can be very useful for those patients who undergo a liver blood test and have unknowingly had contact with the HCV virus, thus, if the model predicts that the patient has infection, you should be screened.