2a)
Since the disease is very rare and affects only a very small population, this is an imbalanced classification where the dataset is biased.
In this particular sample dataset, only 1/10th of the sample is infected whereas the rest are healthy, which leads to a very low percentage of the sample population being in category 2(Infected).
Moreover, since there are no symptoms shown for the disease, it becomes very hard for the machine to differentiate between the two especially with an imbalanced dataset.
Hence, even though the machine is able to catch all the people who are infected, it also shows that a lot of healthy people are infected by the disease. It says that 9/10th of the people are infected, when it’s only 1/10th in reality. Hence the accuracy rate drops down a lot and the model becomes useless.

2b)
This is a right approach to reduce the bias, as by adding 100 more data points, you are increasing the sample train dataset, which will help the model train better. Adding 5 more categories will also help the model identify more features instead of the general prejudiced ones thereby producing a balanced dataset with more output categories which will again help reduce the bias.