1. What do you think it occurred during this model development (trainning & evaluation)?

   I think that what happened is that nobody considered the possibility of a bald man being confused as a soccer ball because, in that case, a better training database to avoid that problem would have been used.

2. How would you fix this behavior? Please provide at least 2 options explaining their pros and drawbacks

   To fix this behaviour, I would provide a better database to the system, with better representations of the soccer ball, but maybe this solution would have a big computational cost.
   On the other hand, a good approach would be estimating the area of the ball and using that information as a threshold. If the area detected is larger, it is probable that it is not a ball. The problem here, is the implementation of an accurate system that estimates the area of the ball.

3. Extra: Do you know any tracking algos (Deep learning based) that could be used here?

   An idea that comes to my mind, is implementing a segmentation of the image. What I mean, is that a soccer ball has a round area and is not connected to anything, whereas a person has the head connected to the neck and the rest of the body, so it would be great to detect if there's an unexpected area surrounding the ball, and if so, you know it could be everything but a ball.