When I ran evaluate\_model.py on #29 (fold 0) I got 0.82/0.58, which is significantly lower than the values in the paper. I am confusion

F1 F23 F1 F23

0.49 0.75 0.62 0.79

0.67 0.84 0.74 0.94

0.53 0.77 0.61 0.81

0.68 0.82 0.76 0.86

0.71 0.91 0.83 0.94

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

0.61 0.82 0.71 0.87

0.73