assess the agreement between raters on coughvid dataset

ratets' diagnoses

```
 subj_status         diagnosis_1              diagnosis_2                 diagnosis_3              diagnosis_4
─────────────   ───────────────────   ───────────────────────   ───────────────────────   ─────────────────

"symptomatic"   {'healthy_cough'  }   {'COVID-19'            }   {'upper_infection'    }   {'lower_infection
"COVID-19"      {'lower_infection'}   {'lower_infection'     }   {'lower_infection'    }   {'lower_infection
"healthy"       {'lower_infection'}   {'COVID-19'            }   {'healthy_cough'      }   {'healthy_cough'
"COVID-19"      {'COVID-19'       }   {'obstructive_disease'}   {'obstructive_disease'}   {'upper_infection
"healthy"       {'COVID-19'       }   {'upper_infection'     }   {'obstructive_disease'}   {'upper_infection
"symptomatic"   {'COVID-19'       }   {'upper_infection'     }   {'upper_infection'    }   {'COVID-19'
"COVID-19"      {'upper_infection'}   {'COVID-19'            }   {'upper_infection'    }   {'upper_infection
"COVID-19"      {'lower_infection'}   {'COVID-19'            }   {'healthy_cough'      }   {'lower_infection
"symptomatic"   {'lower_infection'}   {'obstructive_disease'}   {'upper_infection'    }   {'lower_infection
"healthy"       {'COVID-19'       }   {'COVID-19'            }   {'upper_infection'    }   {'healthy_cough'
```

agreement matrix

No. votes (1 per rater) for each category.

```
   status       COVID-19    healthy_cough    lower_infection    obstructive_disease    upper_infection
─────────────   ────────    ─────────────    ───────────────    ───────────────────    ───────────────

"symptomatic"      1              1                 1                    0                    1
"COVID-19"         0              0                 4                    0                    0
"healthy"          1              2                 1                    0                    0
"COVID-19"         1              0                 0                    2                    1
"healthy"          1              0                 0                    1                    2
"symptomatic"      2              0                 0                    0                    2
"COVID-19"         1              0                 0                    0                    3
"COVID-19"         1              1                 2                    0                    0
"symptomatic"      0              0                 2                    1                    1
"healthy"          2              1                 0                    0                    1
```

fleiss'es kappa

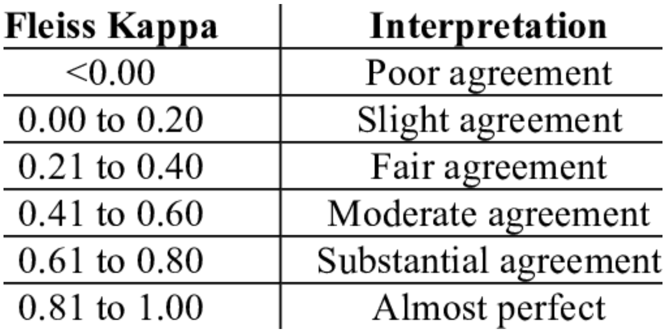
Fleiss' kappa is a statistical measure for assessing the reliability of agreement between a fixed number of raters when assigning categorical ratings to a number of items or classifying items. This contrasts with other kappas such as Cohen's kappa, which only work when assessing the agreement between not more than two raters or the intra-rater reliability (for one appraiser versus themself). The measure calculates the degree of agreement in classification over that which would be expected by chance.

The kappa K can be defined as:

$$k = \frac{p_o - p_e}{1 - p_e}$$

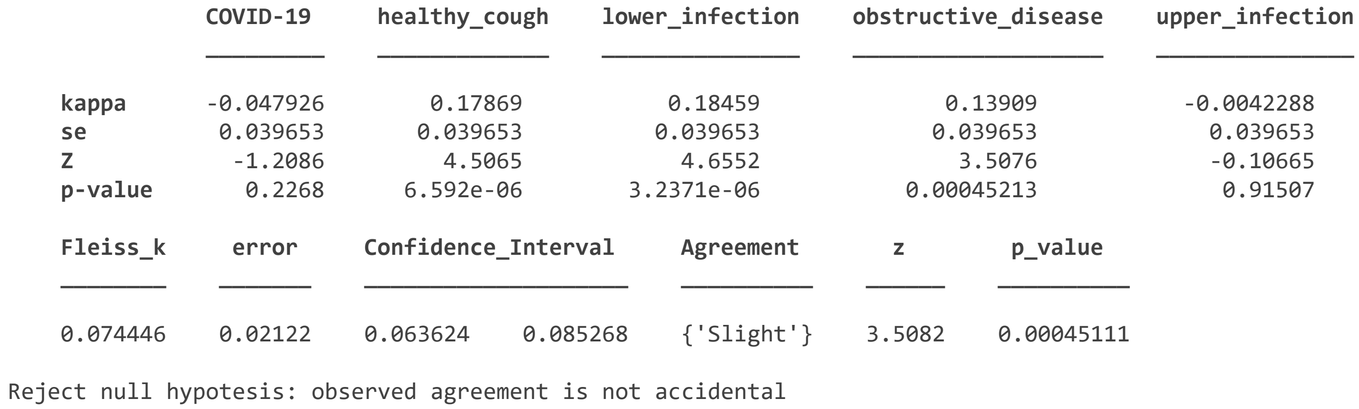
The factor $1 - P_e$ gives the degree of agreement that is attainable above chance, and $P_0 - P_e$ gives the degree of agreement actually achieved above chance. If the raters are in complete agreement then $k = 1$. If there is no agreement among the raters (other than what would be expected by chance) then $k \leq 0$.

table for interpreting k values:



Fleiss Kappa	Interpretation
<0.00	Poor agreement
0.00 to 0.20	Slight agreement
0.21 to 0.40	Fair agreement
0.41 to 0.60	Moderate agreement
0.61 to 0.80	Substantial agreement
0.81 to 1.00	Almost perfect

fleiss'es kappa results:



```
               COVID-19     healthy_cough     lower_infection     obstructive_disease     upper_infection
               _________    _____________     _______________     ___________________     _______________

    kappa      -0.047926       0.17869            0.18459                0.13909             -0.0042288
    se          0.039653       0.039653           0.039653               0.039653             0.039653
    Z            -1.2086         4.5065             4.6552                 3.5076             -0.10665
    p-value       0.2268      6.592e-06         3.2371e-06             0.00045213              0.91507

    Fleiss_k     error      Confidence_Interval      Agreement        z        p_value
    ________    _______    ____________________     __________     ______    __________

    0.074446    0.02122    0.063624    0.085268     {'Slight'}     3.5082    0.00045111

  Reject null hypotesis: observed agreement is not accidental
```