# Developing a multi-variate prediction model for COVID-19 from crowd-sourced respiratory voice data

Yuyang Yan[1*], Wafaa Aljbawi[1], Sami O. Simons[2], Visara Urovi[1]

[1]Institute of Data Science, Maastricht University, 6229 EN Maastricht, The Netherlands

[2]Department of Respiratory Medicine, Maastricht University Medical Center, Maastricht University, 6229 HX Maastricht, The Netherlands

**Table S1.** Patient characteristics in the Coswara dataset

| Attribute | Variable | Value | Count | Percentage |
|---|---|---|---|---|
| Demograpics | Gender | Female | 814 | 31.0% |
| | | Male | 1808 | 68.9% |
| | | Other | 2 | 0.1% |
| | Age | Max age | 99 | |
| | | Min age | 1 | |
| | | Avg age | 35 | |
| Symptoms | | Cold | 452 | 17.2% |
| | | Cough | 605 | 23.0% |
| | | Diarrhoea | 52 | 2.0% |
| | | Fever | 379 | 14.4% |
| | | Loss of smell | 162 | 6.2% |
| | | Muscle pain | 309 | 11.8% |
| | | Breathing difficulty | 206 | 7.9% |
| | | Fatigue | 367 | 14.0% |
| | | Sore throat | 278 | 10.6% |
| | | Other existing conditions | 96 | 3.7% |
| Prior Medical History | | Hypertension | 224 | 8.5% |
| | | Diabetes | 223 | 8.5% |
| | | Ischemic Heart Disease | 37 | 1.4% |
| | | Asthma | 132 | 5.0% |
| | | Chronic lung disease | 42 | 1.6% |
| | | Pneumonia | 44 | 1.7% |
| | | Other respiratory illnesses | 69 | 2.6% |
| | | No listed medical history | 2093 | 79.8% |
| Smoking Status | | Non-smoker | 2418 | 92.1% |
| | | Smoker | 206 | 7.9% |
| COVID-19 test result | | Healthy | 1405 | 53.5% |
| | | Positive-mild | 424 | 16.2% |
| | | No-resp-illness-exposed | 247 | 9.4% |
| | | Positive-moderate | 164 | 6.3% |
| | | Resp-illness-not-identified | 152 | 5.8% |
| | | Recovered-full | 145 | 5.5% |
| | | Positive-asymp | 87 | 3.3% |

**Table S2.** Results from 10 folds of Logistic Regression (LR) in Table 2

| Logistic Regression | Accuracy | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|---|
| Fold 1 | 0.84 | 0.29 | 0.97 | 0.67 | 0.85 |
| Fold 2 | 0.85 | 0.5 | 0.93 | 0.64 | 0.89 |
| Fold 3 | 0.81 | 0.29 | 0.93 | 0.5 | 0.85 |
| Fold 4 | 0.86 | 0.36 | 0.98 | 0.83 | 0.87 |
| Fold 5 | 0.86 | 0.29 | 1 | 1 | 0.86 |
| Fold 6 | 0.82 | 0.33 | 0.95 | 0.62 | 0.85 |
| Fold 7 | 0.89 | 0.5 | 0.98 | 0.88 | 0.89 |
| Fold 8 | 0.82 | 0.36 | 0.93 | 0.56 | 0.86 |
| Fold 9 | 0.90 | 0.57 | 0.98 | 0.89 | 0.9 |
| Fold 10 | 0.83 | 0.29 | 0.97 | 0.67 | 0.85 |

**Table S3.** Results from 10 folds of Support Vector Machine (SVM) in Table 2

| Support Vector Machine | Accuracy | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|---|
| Fold 1 | 0.85 | 0.21 | 1 | 1 | 0.84 |
| Fold 2 | 0.88 | 0.43 | 0.98 | 1 | 0.88 |
| Fold 3 | 0.84 | 0.14 | 1 | 0.86 | 0.83 |
| Fold 4 | 0.86 | 0.29 | 1 | 1 | 0.86 |
| Fold 5 | 0.85 | 0.43 | 1 | 1 | 0.84 |
| Fold 6 | 0.78 | 0.21 | 0.98 | 0 | 0.79 |
| Fold 7 | 0.89 | 0 | 1 | 1 | 0.88 |
| Fold 8 | 0.85 | 0.43 | 1 | 1 | 0.84 |
| Fold 9 | 0.86 | 0.21 | 1 | 1 | 0.85 |
| Fold 10 | 0.88 | 0.29 | 0.98 | 0.86 | 0.88 |

**Table S4.** Results from 10 folds of CNN model (MFCC images) in Table 2

| CNN (MFCC images) | Accuracy | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|---|
| Fold 1 | 0.85 | 0.21 | 1 | 1 | 0.84 |
| Fold 2 | 0.77 | 0.36 | 0.86 | 0.38 | 0.85 |
| Fold 3 | 0.74 | 0.5 | 0.8 | 0.37 | 0.87 |
| Fold 4 | 0.62 | 0.29 | 0.69 | 0.18 | 0.8 |
| Fold 5 | 0.41 | 0.93 | 0.29 | 0.24 | 0.94 |
| Fold 6 | 0.74 | 0.2 | 0.88 | 0.3 | 0.81 |
| Fold 7 | 0.68 | 0.43 | 0.74 | 0.29 | 0.84 |
| Fold 8 | 0.79 | 0 | 0.98 | 0 | 0.8 |
| Fold 9 | 0.76 | 0 | 0.95 | 0 | 0.8 |
| Fold 10 | 0.81 | 0 | 1 | 0 | 0.81 |

**Table S5.** Results from 10 folds of LSTM in Table 2

| LSTM | Accuracy | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|---|
| Fold 1 | 0.86 | 0.43 | 0.97 | 0.75 | 0.88 |
| Fold 2 | 0.85 | 0.57 | 0.92 | 0.62 | 0.9 |
| Fold 3 | 0.86 | 0.36 | 0.98 | 0.83 | 0.87 |
| Fold 4 | 0.89 | 0.57 | 0.97 | 0.8 | 0.9 |
| Fold 5 | 0.84 | 0.43 | 0.93 | 0.6 | 0.87 |
| Fold 6 | 0.84 | 0.33 | 0.97 | 0.71 | 0.85 |
| Fold 7 | 0.89 | 0.57 | 0.97 | 0.8 | 0.9 |
| Fold 8 | 0.85 | 0.29 | 0.98 | 0.8 | 0.85 |
| Fold 9 | 0.9 | 0.57 | 0.98 | 0.89 | 0.9 |
| Fold 10 | 0.81 | 0.57 | 0.9 | 0.5 | 0.87 |

**Table S6.** Results from 10 folds of CNN model (Mel-spectrogram) in Table 2

| CNN (Mel-spectrogram) | Accuracy | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|---|
| Fold 1 | 0.85 | 0.36 | 0.97 | 0.71 | 0.86 |
| Fold 2 | 0.85 | 0.5 | 0.93 | 0.64 | 0.89 |
| Fold 3 | 0.82 | 0.71 | 0.85 | 0.53 | 0.93 |
| Fold 4 | 0.84 | 0.36 | 0.95 | 0.62 | 0.86 |
| Fold 5 | 0.79 | 0.21 | 0.93 | 0.43 | 0.83 |
| Fold 6 | 0.90 | 0.53 | 1 | 1 | 0.89 |
| Fold 7 | 0.78 | 0.5 | 0.84 | 0.44 | 0.88 |
| Fold 8 | 0.90 | 1 | 0.88 | 0.67 | 1 |
| Fold 9 | 0.88 | 1 | 0.84 | 0.61 | 1 |
| Fold 10 | 0.90 | 1 | 0.88 | 0.67 | 1 |

**Table S7.** Results from 10 folds of HuBERT model in Table 2

| HuBERT | Accuracy | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|---|
| Fold 1 | 0.82 | 0.90 | 0.78 | 0.68 | 0.94 |
| Fold 2 | 0.82 | 0.94 | 0.76 | 0.67 | 0.96 |
| Fold 3 | 0.89 | 0.81 | 0.93 | 0.86 | 0.90 |
| Fold 4 | 0.88 | 0.73 | 0.95 | 0.88 | 0.88 |
| Fold 5 | 0.90 | 0.80 | 0.95 | 0.89 | 0.90 |
| Fold 6 | 0.84 | 0.77 | 0.88 | 0.77 | 0.89 |
| Fold 7 | 0.85 | 0.71 | 0.93 | 0.85 | 0.85 |
| Fold 8 | 0.92 | 0.94 | 0.91 | 0.85 | 0.96 |
| Fold 9 | 0.87 | 0.71 | 0.95 | 0.88 | 0.86 |
| Fold 10 | 0.85 | 0.71 | 0.93 | 0.85 | 0.86 |