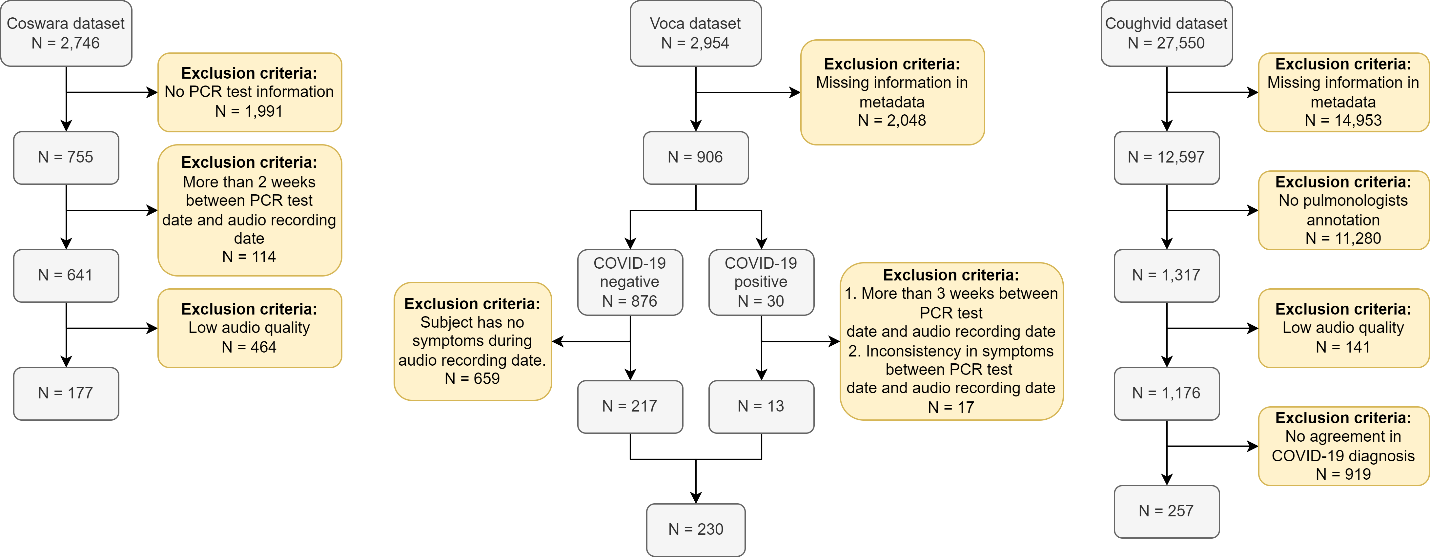
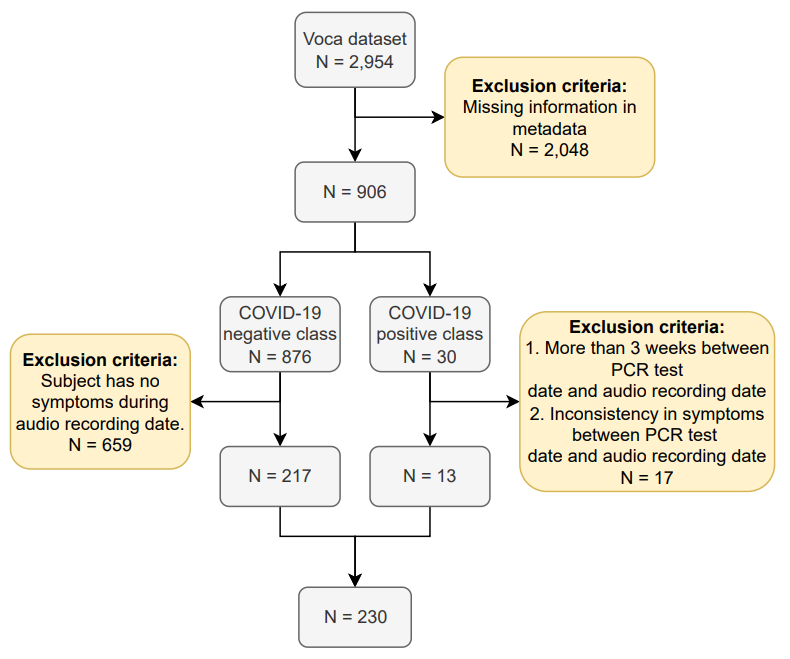
**Subjects inclusion exclusion criterion:**

****

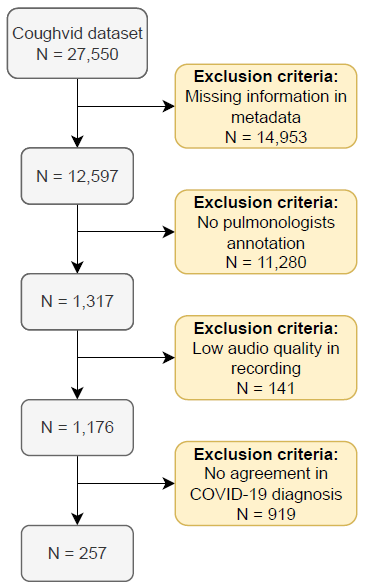
**Voca:** 2,954 נבדקים לפני סינון.



|  |  |
| --- | --- |
| חיוביים: | שליליים: |
| קיום מידע: מצב נבדק, קובץ שמע, תאריך בדיקת PCR, סימפטומים בזמן הקלטה  (סימון גיל, תאריך הקלטה וסימפטומים בזמן בדיקת PCR הופיעו אצל מי שנשאר) | לאף נבדק אין סימון תאריך בדיקת PCR => לא חלק משיטת הסינון. |
| סימון נבדק על בדיקת PCR עם תוצאה חיובית כולל תאריך הבדיקה. | קיום מידע: מצב נבדק, קובץ שמע, סימון סימפטומים בזמן הקלטה, סימון מחלות רקע (בדיקת אמינות), סימון גיל, טווח גילאים בין 10 ל-90 (לא אמין מחוץ לטווח זה => לא סומך על הנבדק). |
| 30 נבדקים לאחר סינון. | סימון נבדק על בדיקת PCR עם תוצאה שלילית. |
| הפרש זמנים של לא יותר מ-3 שבועות בין בדיקת PCR לבין ההקלטה (בדיקה ידנית). | 876 נבדקים לאחר סינון. |
| סימון נבדק על סימפטומים בזמן בדיקת PCR ובזמן ההקלטה (בדיקה ידנית). | סימון נבדק שהוא ללא סימפטומים במועד הקלטה. |
| עקביות בסימפטומים בין בדיקה לבין הקלטה (בדיקה ידנית). | 217 נבדקים לאחר סינון. |
| 13 נבדקים לאחר סינון. |  |

**Voca – סוגי מחלות רקע:**

|  |  |
| --- | --- |
| "Allergies" | "Ischemic heart disease" |
| "Hypertension" | "Chronic lung disease" |
| "Diabetes" | "I don’t have any of the above" |
| "Asthma" | "Autoimmune disease" |
| "Chronic kidney disease" | "Active cancer" |
| "Seasonal allergies" |  |

**Coughvid:** 27,550 נבדקים לפני סינון

אפשרויות מצבים:

* סימון של נבדק על מצב COVID-19: "symptomatic" | "healthy" | "COVID-19".
* סימון של מומחה על איכות אודיו: "no cough" | "poor" | "ok" | "good".
* סימון של מומחה על מצב COVID-19: "upper infection" | "obstructive disease" | "lower infection" | "healthy cough" | "COVID-19".

\*ללא סימון מומחים:

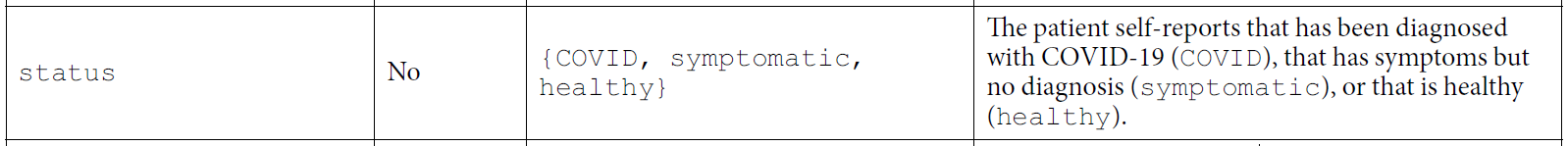
* הסתברות לקיום שיעול בהקלטה >= 0.7 (אלגוריתם אוטומטי של מפרסמי המאגר).
* חיוביים: סימון של נבדק על מצב COVID-19: "COVID-19".
* שליליים: סימון של נבדק על מצב COVID-19: "healthy".

|  |  |  |
| --- | --- | --- |
| כללי: | חיוביים: | שליליים: |
| סימון של לפחות 1 מתוך 4 מומחים (איכות אודיו + מצב COVID-19). | סימון של נבדק על מצב COVID-19: "COVID-19". | סימון של נבדק על מצב COVID-19: "healthy". |
| קיום מידע: מצב נבדק, סימון גיל, טווח גילאים בין 10 ל-90 (לא אמין מחוץ לטווח זה => לא סומך על הנבדק). | סימון של מומחים על מצב COVID-19: לפחות 1 סימן "COVID-19", ואף אחד לא סימן: "healthy cough" | "upper infection" | "obstructive disease" | "lower infection".  כלומר, רק "COVID-19". | סימון של מומחים על מצב COVID-19: לפחות 1 סימן "healthy", ואף אחד לא סימן:  "upper infection" | "obstructive disease" | "lower infection" | "COVID-19".  כלומר, רק "healthy". |
| סימון איכות הקלטה: "ok" | "good". | 97 נבדקים לאחר סינון. | 160 נבדקים לאחר סינון (ver1). |
|  |  | **אופציה נוספת:**  סימון של מומחים על מצב COVID-19: לפחות 1 סימן "healthy", ואף אחד לא סימן: "COVID-19".  כלומר, ניתן לסמן: " healthy " | " upper infection " | " obstructive disease " | " lower infection ". |
|  |  | 525 נבדקים לאחר סינון (ver2). |

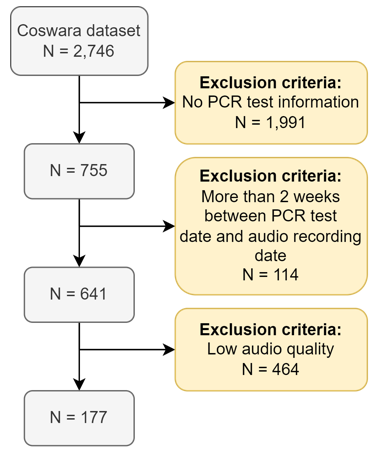
אפשרויות סימונים:

* סימון של נבדק על מצב COVID-19: "symptomatic" | "healthy" | "COVID-19".
* סימון של מומחה על מצב COVID-19: "upper infection" | "obstructive disease" | "lower infection" | "healthy cough" | "COVID-19".

הסבר מהמאמר של Coughvid על סימון עצמי של נבדקים:



נראה ש-symptomatic לא מספק מידע לגבי האם לנבדק יש קורונה או לא => לקחת סימונים רק של "COVID" | "healthy".

**Coswara:** 2,746 נבדקים לפני סינון.

* סימון של נבדק על בדיקת PCR כולל תאריך.

755 נבדקים לאחר סינון.

* הפרש זמנים של מקסימום 2 שבועות בין בדיקת PCR לבין ההקלטה.

614 נבדקים לאחר סינון.

* סימון ידני של מפרסמי המאמר על איכות הקלטה.

אפשרויות: 2 (excellent) | 1 (good) | 0 (bad) | -1 (not labeled)

נשארתי רק עם 2 (excellent), 1 (good).

121 נבדקים לאחר סינון.

* בדיקת איכות אודיו ידנית שלי (ניר) עבור הקלטות שהמפרסמים לא ביצעו להם בדיקה (סימון 1-).

במקום 121 => 179 נבדקים לאחר סינון.

* סימון ידני של נדב על איכות ההקלטה.

אפשרויות: "NoGood" | "Good".

נשארתי רק עם "Good".

177 נבדקים לאחר סינון.

* הפרדה לחיוביים/שליליים לפי תוצאת בדיקת PCR.

2 נבדקים עם הקלטה אחת פר נבדק (cough-heavy NoGood). 175 נבדקים עם 2 הקלטות פר נבדק (cough-heavy & cough-shallow).

**names of symptoms**:

symptomsYes = ["positive\_mild" "positive\_moderate" "resp\_illness\_not\_identified"];

symptomsNo = ["positive\_asymp" "no\_resp\_illness\_exposed" "healthy"];

אפשרויות סימון של נבדקים במאגר coswara:

“healthy” | “no\_resp\_illness\_exposed” | “positive\_asymp” | “positive\_mild” | “positive\_moderate” | “recovered\_full” | “resp\_illness\_not\_identified”

N – negative PCR test result

P – positive PCR test result

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | positive\_mild  positive\_moderate  resp\_illness\_not\_identified  P | positive\_asymp  no\_resp\_illness\_exposed  P | resp\_illness\_not\_identified  N | Healthy  no\_resp\_illness\_exposed  N |
|  | All classes | Covid-19 pos Symptoms Yes | Covid-19 pos Symptoms No | Covid-19 neg Symptoms Yes | Covid-19 neg Symptoms No |
| Coswara | 119 | 80 | 43 | 10 | 41 |

|  |  |  |
| --- | --- | --- |
| **No. subjects**  **True labels** | Previous method | After annotations and subjects exclusion |
| All subjects | 119 | 177 |
| COVID-19 **Positive** | 93 | 125 |
| COVID-19 **Negative** | 26 | 52 |
| COVID-19 **Positive**  Symptoms **Yes** | 80 | 80 |
| COVID-19 **Positive**  Symptoms **No** | 11 | 43 |
| COVID-19 **Negative**  Symptoms **Yes** | 7 | 10 |
| COVID-19 **Negative**  Symptoms **No** | 18 | 41 |

**כמות נבדקים בהתאם לחלוקה למחלקות:**

אפשרויות: Covid-19 positive | Covid-19 negative

|  |  |  |  |
| --- | --- | --- | --- |
|  | All classes | Covid-19 positive | Covid-19 negative |
| All data sets | 664 | 235 | 429 |
| Voca | 230 | 13 | 217 |
| Coughvid | 257 | 97 | 160 |
| Coswara | 177 | 125 | 52 |

**כמות נבדקים בהתאם לחלוקה למחלקות (בדיקה יותר מעמיקה):**

אפשרויות: CovidPosSymptomsYes | CovidPosSymptomsNo | CovidNegSymptomsYes | CovidNegSymptomsNo

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | All classes | Covid-19 pos Symptoms Yes | Covid-19 pos Symptoms No | Covid-19 neg Symptoms Yes | Covid-19 neg Symptoms No |
| All data sets | 650 |  |  |  |  |
| Voca |  |  |  |  |  |
| Coughvid |  |  |  |  |  |
| Coswara |  |  |  |  |  |