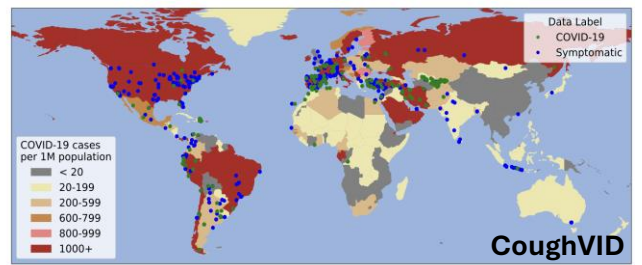


A. Geo-distribution for datasets used for pretraining.

Dataset	Region
COVID-19 Sounds	Global
CoughVID	Global
UK COVID-19	European
ICBHI	European
HF Lung	Asian



B. Demographics for datasets used for pretraining (Figures are included in Appendix).

Dataset	Modality	#Sample(#P articipants)	Age	Gender	Medical conditions
COVID-19 Sounds	Cough	40866 (22162)	0-20: 1413 20-29: 3991 30-39: 5459 40-49: 4928 50-59: 3486 60-69: 1981 70-79: 672 80-89: 89 90+ : 4	Female: 8146 Male: 13733	High Blood Pressure: 2704, Asthma: 1712, Other long-term condition: 1217, Diabetes: 733, Other heart disease: 353, COPD/Emphysema: 234, Other lung disease: 228, Previous heart attack: 217, Valvular heart disease: 162, Previous stroke or Transient ischaemic attack: 144, Cancer: 112, Angina: 107, HIV or impaired immune system: 106, Previous organ transplant: 35, Pulmonary fibrosis: 23, Cystic fibrosis: 20, COVID-19 positive: 534
	Breath	36605 (20635)	0-19: 1238 20-29: 3741 30-39: 5070 40-49: 4585 50-59: 3310 60-69: 1848 70-79: 634 80-89: 93 90+ : 3	Female: 7322 Male: 13074	High Blood Pressure: 2571, Asthma: 1609, Other long-term condition: 1112, Diabetes: 697, Other heart disease: 324, Other lung disease: 223, COPD/Emphysema: 216, Previous heart attack: 212, Valvular heart disease: 156, Previous stroke or Transient ischaemic attack: 141, Cancer: 111, Angina: 104, HIV or impaired immune system: 92, Previous organ transplant: 22, Pulmonary fibrosis: 20, Cystic fibrosis: 19, COVID-19 positive: 532
UK COVID-19	Cough	19533 (NA)	18-44: 5134 45-64: 8767 65+: 5632	Female: 11460 Male: 8068	COVID-19 positive: 7240 Asthma: 2184 Other respiratory conditions: 569
	Exhalation	20719 (NA)	18-44: 5090 45-64: 9440 65+: 6189	Female: 11902 Male: 8815	COVID-19 positive: 7283 Asthma: 2253 Other respiratory conditions: 601
CoughVID	Cough	7179 (NA)	0-20: 405 20-29: 1128 30-39: 1020 40-49: 728 50-59: 343 60+ : 3555	Female: 1342 Male: 2646	Healthy: 3077 Symptomatic: 631 COVID-19: 325 Other respiratory conditions: 729
ICBHI	Lung sound	538 (79)	0-10: 20 10-19: 9 20-29: 1 50-59: 5 60-69: 18 70+ : 26	Female: 32 Male: 47	Sample-level statistics: Has crackle: 310, has wheeze: 203 Participant-level statistics: Healthy: 12, COPD: 39, Pneumonia: 6, URTI: 10, Bronchiectasis: 6, Bronchiolitis: 3, LRTI: 2, Asthma: 1
HF Lung	Lung sound	10554 (299)	>20, Mean = 66.58	Female: 21 Male: 35	Sample-level statistics: Wheeze: 2253, Rhonchi: 944, Stridor: 253 Participant-level statistics: Acute exacerbation of chronic obstructive pulmonary disease: 2, Acute respiratory distress syndrome: 1, Acute respiratory failure: 4, Asthma: 1, Bronchitis: 1, Chronic respiratory failure: 14, Chronic obstructive pulmonary disease: 7, Emphysema: 1, Pleural effusion: 1, Pneumoconiosis: 1, Pneumonia: 13, Pulmonary embolism: 1