

INTISARI

METODE PEMBELAJARAN HIBRID BERDASARKAN KLASTERISASI FITUR DAN NILAI PENTING UNTUK MENINGKATKAN KINERJA GENOSE C19 DALAM MENDETEKSI COVID-19

SHIDIQ NUR HIDAYAT
17/422646/SPA/00633

Analisis pola napas menggunakan hidung elektronik (GeNose C19) memiliki kelebihan, yaitu tidak invasif, cepat, dan biaya relatif murah, telah lama digunakan secara berkelanjutan untuk mendeteksi penyakit, termasuk deteksi virus corona 2019 (COVID-19). Dalam penelitian ini, kami mengembangkan GeNose C19 berbasis 10 larik sensor gas *metal oxide semiconductor* untuk mendeteksi COVID-19 melalui sampel napas. Namun, hasil menunjukkan data positif dan negatif COVID-19 memiliki variasi tinggi dan saling tumpang tindih. Jumlah sensor yang banyak dan saling berkorelasi tidak selalu menghasilkan performa yang tinggi. Oleh karena itu, kami mengembangkan algoritme hibrid, yaitu kombinasi metode *filter (hierarchical agglomerative clustering)* dan *wrapper* (permutasi fitur penting), untuk optimasi kinerja GeNose C19. Melalui algoritme ini, kombinasi fitur yang efektif dan optimal dapat diperoleh dengan mudah dan sederhana, memungkinkan pengurangan jumlah sensor tanpa menurunkan performa model klasifikasi. Berdasarkan hasil validasi silang terhadap data latih, model optimasi menghasilkan akurasi ($85,0 \pm 0,8$) %, sensitivitas ($87,0 \pm 1,2$) %, dan spesifisitas ($83,0 \pm 1,2$) % dengan CI 95%. Sedangkan, performa terhadap data uji menghasilkan akurasi, sensitivitas, dan spesifisitas sebesar 85,0% dengan kenaikan performa akurasi sebesar 10,4% (terhadap model klasifikasi tanpa optimasi) dan dengan jumlah sensor 50% lebih sedikit. Hasil ini telah menunjukkan kelayakan penggunaan algoritme hibrid untuk mengoptimalkan kinerja GeNose C19.

Kata kunci: analisis napas, GeNose C19, hidung elektronik, *hierarchical agglomerative clustering*, *machine learning*, permutasi fitur penting.

ABSTRACT

A HYBRID LEARNING METHOD BASED ON FEATURE CLUSTERING AND SCORING FOR IMPROVING PERFORMANCE OF GENOSE C19 IN COVID-19 DETECTION

SHIDIQ NUR HIDAYAT
17/422646/SPA/00633

Breath pattern analysis based on electronic nose (GeNose C19) that is non-invasive, fast, and low-cost has been continuously used for detecting human diseases, including the coronavirus disease 2019 (COVID-19). Here, we develop GeNose C19 based on 10 metal oxide semiconductor (MOS) gas sensor array for the detection of COVID-19 through breath samples. However, the results show positive and negative COVID-19 data have high variation and overlapping. Nevertheless, having many high correlated gas sensors is not always beneficial getting high performance. Thus, we develop a hybrid algorithm, a combination of filter (hierarchical agglomerative clustering) and wrapper (permutation feature importance) methods to optimize the performance of GeNose C19. Utilizing this learning approach, effective and optimum feature combination can be obtained, enabling reduction by half the number of employed sensors without downgrading the classification model performance. Based on the cross-validation test results on the training data, the hybrid algorithm can result in accuracy, sensitivity, and specificity values of $(85.0 \pm 0.8) \%$, $(87.0 \pm 1.2) \%$, and $(83.0 \pm 1.2) \%$, respectively with CI 95%. Meanwhile, for the testing data, a value of 85.0% is obtained for all the three metrics that increasing in accuracy performance of 10.4% (against the classification model without optimization) with 50% fewer sensors. These results have exhibited the feasibility of using this hybrid filter-wrapper feature-selection method to pave the way for optimizing the GeNose C19 performance.

Keywords: breath analysis, electronic nose, feature permutation importance, GeNose C19, hierarchical agglomerative clustering, machine learning.