

ABSTRACT

Ferinzhy Tristan Halik (01082170035)

COVID-19 COUGH SOUND DETECTION WITH CONVOLUTION NEURAL NETWORK

(xvi + 37 pages: 28 figures; 4 tables; 6 attachment)

COVID-19 is the initial disease that started its outbreak in the Wuhan area, China in 2019 in December. Over time, the transmission caused by COVID-19 reaches the world due to transmission through small droplets that occur when a person speaks, breathes, especially coughs. There have been several studies that have been carried out previously to predict the status of COVID-19 through cough sound predicted using CNN, and the cough sound dataset used previously had many factors that were not included in the research conducted in this thesis, such as age, location, gender, and also some other symptoms, the dataset that has not been cleaned have a total of 27,551 data.

After cleaning, deleting several columns and rows that are not needed as well as do oversampling to the data, the total dataset used is 24,199 data of which 12,479 have a negative COVID-19 status and 11,550 data have a positive COVID-19 status. This study aims to classify coughing sounds that come from people with COVID-19 and those who do not, by using a dataset collected from Kaggle and also the COUGHVID Crowdsourcing dataset. The model is designed using several activation methods such as Dense, ReLU, LeakyReLU, and also Dropout, and classification using Sigmoid Function.

The results of the evaluation using a combined dataset where the number of testing data is 4840 data resulting in 90% accuracy, 98% precision, 2125 True Positive data from 4840 data (43.9%), True Negative data as many as 2272 data from 4840 data (46.9%), False Positive data is 41 data from 4840 (0.8%) data and False Negative data is 402 data from 4840 data (8.3%). The results are generated using a model designed provides satisfactory results, because the training is carried out using as many datasets as possible.

Reference: 36 (1987 – 2021)

ABSTRAK

Ferinzhy Tristan Halik (01082170035)

PENDETEKSI SUARA BATUK COVID-19 DENGAN *CONVOLUTION NEURAL NETWORK*

(xvi + 37 halaman: 28 gambar; 4 tabel; 6 lampiran)

COVID-19 merupakan penyakit pernapasan yang memulai wabahnya di daerah Wuhan, Tiongkok pada tahun 2019 bulan Desember. Seiring waktu berjalan, penularan yang disebabkan oleh COVID-19 mencapai ke seluruh dunia dikarenakan penularan yang melewati tetesan kecil yang terjadi ketika seseorang berbicara, bernapas, terutama batuk. Sudah ada beberapa penelitian yang telah dilakukan sebelumnya untuk memprediksi status COVID-19 melalui suara batuk yang diprediksi menggunakan CNN, dan data suara batuk yang digunakan sebelumnya memiliki banyak sekali faktor-faktor yang tidak dimasukkan ke dalam pengerjaan penelitian yang dilakukan dalam skripsi ini, seperti umur, lokasi, jenis kelamin, dan juga beberapa gejala-gejala lainnya, dataset yang belum dilakukan pembersihan memiliki total data sebanyak 27.551 data.

Setelah melakukan pembersihan, penghapusan beberapa kolom dan baris yang tidak diperlukan juga melakukan oversampling data, data yang digunakan adalah sebanyak 24.199 data dimana yang memiliki status negatif COVID-19 adalah sebanyak 12.479 dan yang memiliki status positif COVID-19 adalah sebanyak 11.550 data. Penelitian ini bertujuan untuk mengklasifikasi suara batuk yang berasal dari pengidap COVID-19 dan yang bukan merupakan pengidap COVID-19 dengan menggunakan dataset yang dikumpulkan dari Kaggle dan juga COUGHVID Crowdsourcing dataset. Model yang dirancang menggunakan beberapa metode aktivasi seperti Dense, ReLU, LeakyReLU, dan juga Dropout, dan klasifikasi menggunakan Sigmoid Function.

Hasil evaluasi yang menggunakan dataset yang telah digabung dimana jumlah testing data sebanyak 4840 data menghasilkan output akurasi sebesar 90.8%, presisi sebesar 98.1%, data True Positive sebanyak 2125 data dari 4840 data (43.9%), data True Negative sebanyak 2272 data dari 4840 data (46.9%) , data False Positive sebanyak 41 data dari 4840 (0.8%) data dan data False Negative sebanyak 402 data dari 4840 data (8.3%). Hasil yang dihasilkan menggunakan model yang telah dirancang memberikan hasil yang memuaskan, dikarenakan pelatihan yang dilakukan menggunakan dataset yang berjumlah sebanyak.

Referensi: 36 (1987 – 2021)