

BAB I

PENDAHULUAN

1.1 Latar Belakang

Coronavirus disease 2019 (COVID-19) ditetapkan menjadi pandemi global oleh Organisasi Kesehatan Dunia (WHO) sejak 11 Maret 2020, dan telah mewabah hampir di seluruh belahan dunia (Lella and Pja, 2021). Perkembangan secara global yang telah dilaporkan kepada WHO per 12 November 2021, terdapat 251.788.329 kasus terkonfirmasi COVID-19 termasuk 5.077.907 kematian. Indonesia sendiri memiliki total kasus yang cukup tinggi yaitu 4.250.157 positif COVID-19 dengan 143.628 kematian (WHO, 2021). Jumlah angka positif dan kematian yang terus meningkat dalam kurun waktu dekat menunjukkan bahwa COVID-19 sangat cepat penyebarannya.

Penyakit COVID-19 menyerang saluran pernapasan bagian bawah, fungsi neuromuskular, indera perasa dan penciuman yang disebabkan oleh virus *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2) baik disertai dengan gejala maupun tanpa gejala. Hal tersebut menyebabkan pengurangan kompleksitas koordinasi gerakan pernapasan dan laring, sehingga pasien kesulitan bernafas secara signifikan (Coppock *et al.*, 2021). Gejala yang dapat dirasakan jika terinfeksi COVID-19, diantaranya adalah demam, batuk berdahak, dan sesak napas. Penularannya dapat menyebar melalui kontak dekat, terpapar batuk, bersin atau partikel *aerosol* dari individu terinfeksi (Shereen *et al.*, 2020). Menyentuh benda terkontaminasi kemudian kontak dengan hidung, mulut atau mata tanpa membersihkan tangan terlebih dahulu dapat meningkatkan risiko penularan (Goel *et al.*, 2021).

Berbagai upaya untuk mencegah penularan COVID-19 telah dilakukan, misalnya menjaga jarak minimal 1 meter, menggunakan masker, sering mencuci tangan dengan sabun atau dengan menggunakan *hand sanitizer*, membatasi

mobilitas seperti kebijakan *lockdown* yang sering dikeluarkan oleh pemerintah. Kebijakan *lockdown* atau Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) di Indonesia dapat menekan persebaran virus secara signifikan dan memudahkan pengujian individu dalam skala besar. Sejauh ini *gold standard* dalam menguji atau deteksi awal COVID-19 yaitu dengan metode *Reverse Transcription Polymerase Chain Reaction* (RT-PCR) (Y. Li *et al.*, 2020). Sampel yang digunakan pada RT-PCR diperoleh dari *swab* nasofaring, berupa sampel lendir dari hidung dan tenggorokan. RT-PCR dapat memberikan hasil uji dengan tingkat akurasi tinggi. Namun metode tersebut memiliki beberapa kelemahan yaitu: membutuhkan mobilisasi sehingga dapat meningkatkan risiko penularan penyakit, membutuhkan biaya cukup mahal, membutuhkan tenaga ahli dalam proses pengujian, dan membutuhkan waktu cukup lama sekitar 1-2 hari untuk mengetahui hasilnya (Sharma *et al.*, 2020). Lama waktu yang dibutuhkan tersebut menjadi sangat krusial karena cepat lambatnya tindakan medis yang diberikan kepada pasien ditentukan dari hasil uji, sehingga keterlambatan penanganan hingga risiko penularan akan semakin meningkat.

Meskipun saat ini sudah berhasil dikembangkan vaksin untuk meminimalkan penularan, namun hal tersebut memunculkan berbagai masalah. Masalah yang muncul misalnya adanya keraguan mengenai vaksin, adanya keterbatasan pasokan, dan distribusi yang kurang merata. Berdasarkan hasil tinjauan dari 126 survei terkait vaksinasi COVID-19 yang mencakup 31 negara menunjukkan adanya penurunan penerimaan vaksin global sekitar 20% dari bulan Maret hingga Oktober 2020 (Lin, Tu and Beitsch, 2021). Hingga kini jumlah penduduk Indonesia yang telah melakukan vaksin menurut Kementerian Kesehatan (Kemenkes) RI per 11 November 2021 sebanyak 81,7 juta jiwa dari total 277 juta lebih penduduk. Setelah melakukan vaksin bukan berarti terbebas dari COVID-19 seutuhnya karena risiko terpapar setelah vaksin tetap ada. Maka protokol kesehatan tetap harus dipatuhi serta deteksi dini masih berperan penting untuk pengendalian pandemi. Oleh karena itu, diperlukan suatu perangkat atau algoritma untuk pendukung deteksi COVID-

19 yang mampu mengidentifikasi kasus bergejala dan tidak bergejala, hemat biaya, cepat, serta tidak memerlukan mobilisasi ke rumah sakit.

Permasalahan yang muncul dalam kehidupan, khususnya dalam bidang medis tidak menutup kemungkinan dapat diselesaikan dengan bantuan teknologi. Salah satu teknologi yang sering digunakan yaitu kecerdasan buatan (AI). AI memberikan dampak yang revolusioner dalam diagnosis medis. Penggunaanya dapat menginterpretasikan data secara komprehensif dengan waktu singkat sehingga lebih efisien dan akurat (Bhattad and Jain, 2020). Banyak model kecerdasan buatan yang telah digunakan untuk menyelesaikan permasalahan medis baik dengan menggunakan *deep learning*, *machine learning*, dan *signal processing* (Lella and Pja, 2021). Khususnya dalam penanganan COVID-19, penggunaan AI dapat mengurangi kompleksitas perawatan dan waktu yang dibutuhkan jika dibandingkan tanpa menggunakan teknologi AI (Vaishya *et al.*, 2020).

Para peneliti telah memanfaatkan teknologi AI untuk mengembangkan metode deteksi dini COVID-19 dari sampel suara napas. Literatur medis menunjukkan bahwa pola pernapasan terkait erat dengan perubahan anatomi dan fisiologi sistem pernapasan (Sharma *et al.*, 2020). Penelitian yang dilakukan oleh Lella & Pja tahun 2021 mengembangkan diagnosis otomatis untuk membedakan antara COVID-19 dan *Non COVID-19* dari beberapa suara, yaitu: suara batuk, sampel suara, dan suara napas. Metode yang diimplementasikan yaitu *Deep Convolutional Neural Network* (DCNN), dimana input untuk DCNN diperoleh dari multi fitur seperti *Gamma-tone Frequency Cepstral Coefficients*, *De-noising Auto Encoder*, dan *Improved Multi-frequency Cepstral Coefficients*. Hasil penelitian menunjukkan COVID-19 dapat dideteksi dengan tingkat akurasi mencapai 95,45 % (Lella and Pja, 2021).

Coppock juga melakukan penelitian dengan menggunakan data suara batuk dan pernapasan untuk dapat mendeteksi COVID-19. Metode yang digunakan yaitu CNN dengan arsitektur ResNet yang disebutkan dalam jurnalnya sebagai COVID-19 *Identification ResNet* (CIdeR) (Coppock *et al.*, 2021). Coppock CIdeR, mampu

menunjukkan bahwa rekaman audio suara napas dapat digunakan untuk mendeteksi kasus COVID-19 bergejala dan tanpa gejala. Hasil penelitian ini menunjukkan kinerja empiris yang kuat, dengan nilai AUC-ROC 0,846 pada klasifikasi COVID-19. Kelemahan dari penelitian ini yaitu jumlah sampel positif COVID-19 yang terbatas sehingga memiliki interval kepercayaan yang relatif lebar.

Penggunaan metode *deep learning* yaitu CNN baik dalam penelitian Lella & Pja yang dijelaskan sebelumnya dan Coppock, memiliki kelemahan dalam proses pelatihan CNN. Pelatihan pada CNN membutuhkan waktu lama dan sulit karena juga membutuhkan dataset berlabel yang sangat besar untuk membangun model sebelum siap digunakan. Selain itu, spesifikasi *hardware* yang dibutuhkan menjadi cukup tinggi (Mohsen *et al.*, 2018).

Machine learning merupakan salah satu aplikasi dari *Artificial Intelligence* (AI) yang memungkinkan sistem belajar secara otomatis dari data yang diberikan, mengidentifikasi pola dan menentukan keputusan tanpa diprogram secara eksplisit. Perbedaan mendasar dengan *deep learning* yaitu pada *machine learning* tidak membutuhkan waktu lama dalam proses pelatihan dan memerlukan ekstraksi fitur data terlebih dahulu (Kulyukin, Mukherjee and Amlathe, 2018). Ekstraksi fitur menjadi bagian terpenting dalam sistem pengenalan berbasis *machine learning* untuk mengubah *raw signal* menjadi beberapa jenis representasi parametrik (Ahammad, 2016). Dengan ekstraksi fitur maka akan diperoleh informasi spesifik, sehingga dapat dihasilkan model prediksi yang baik dan komputasi yang intensif (Ramadevi and Rani, 2015). Dalam deteksi COVID-19 berbasis audio, terdapat beberapa fitur audio yang biasanya digunakan seperti *speech time duration*, onset, tempo, periode, RMS *energy*, dan MFCC (Dash *et al.*, 2021). Metode *Mel Frequency Cepstral Coefficients* (MFCC) sangat populer dan efektif karena dapat mempresentasikan sinyal dengan baik (Umar, Riadi and Hanif, 2018). Penggunaan fitur tersebut dapat mencapai konvergensi yang lebih cepat dan meningkatkan akurasi. Oleh sebab itu, seleksi fitur menjadi langkah penting dalam membangun model klasifikasi yang dibutuhkan.

Terdapat berbagai macam algoritma klasifikasi yang biasanya digunakan, salah satunya adalah *k-Nearest Neighbor* (k-NN), algoritma untuk *machine learning* yang paling sederhana dengan teknik *supervised learning*. Kelebihan k-NN dibandingkan pengklasifikasi lainnya yaitu tidak memerlukan fase pelatihan yang lama dan mahal, serta memungkinkan penggunaan pembelajaran adaptif (Hoyos-barceló *et al.*, 2017). Metode k-NN mampu mengklasifikasikan suara napas normal dan abnormal dengan tingkat akurasi mencapai 95% (Neili, Fezari and Redjati, 2020). R. Palaniappan mengklasifikasikan patologi pernapasan dari database suara paru-paru RALE menggunakan *Autoregressive Coefficients* dan k-NN sebagai pengklasifikasi dengan hasil akurasi 98,26% (Palaniappan, Sundaraj and Sundaraj, 2014). Pada dasarnya k-NN melakukan klasifikasi dengan mempertimbangkan semua titik sampel secara berurutan untuk mengetahui jarak antara titik-titik data. Performa dari k-NN bergantung pada jumlah data tetangga terdekat atau disebut *k*. Nilai *k* digunakan untuk mencari jumlah tetangga terdekat yang menentukan label kelas untuk sampel yang tidak diketahui. Nilai *k* tersebut nantinya akan berpengaruh terhadap hasil akurasi yang diperoleh. Namun penentuan jumlah nilai *k* menjadi kekurangan dalam penerapan k-NN karena masih dilakukan manual secara berulang-ulang, sehingga cukup sulit untuk memperoleh hasil klasifikasi dengan akurasi yang optimal.

Para peneliti menambahkan metode optimasi pada proses klasifikasi untuk memperoleh hasil akurasi yang optimal. Sejumlah metode optimasi telah dikembangkan yang terinspirasi dari proses fisik atau biologis di alam, seperti menggunakan pendekatan terhadap perilaku sekelompok hewan. Salah satu contoh algoritma nya yaitu *Artificial Bee Colony* (ABC), algoritma *swarm intelligence* yang terinspirasi dari perilaku kawanan lebah madu dalam mencari makan (Karaboga, Akay and Karaboga, 2020). ABC telah banyak digunakan sebagai teknik optimasi karena memiliki kelebihan yaitu sederhana, fleksibel dan *robust* (Hussain, Pervez and Hussain, 2020). Karaboga *et al*, membandingkan kinerja algoritma ABC pada sejumlah set data dari repositori UCI dengan beberapa algoritma *clustering*. Hasilnya menunjukkan bahwa ABC memiliki kinerja yang

lebih baik (Karaboga and Akay, 2009). Penelitian yang dilakukan oleh Kauser dan Acharjya tentang deteksi dini penyakit hepatitis dengan melakukan hibridisasi ABC dan pendekatan matematis *rough set* memberikan hasil akurasi 96.2% (Kauser and Acharjya, 2021).

Beberapa penelitian diatas menjadi pertimbangan penulis dalam melakukan penelitian untuk deteksi COVID-19 dari sinyal suara napas. Dalam penelitian ini, data sinyal suara napas yang digunakan berasal dari Coswara *database*, yaitu *database open access* dimana sampel suara dikumpulkan melalui *crowd-sourcing* di seluruh dunia menggunakan aplikasi situs *web*. Data sinyal suara napas akan di ekstraksi dengan menggunakan MFCC untuk memperoleh fitur yang dibutuhkan dalam proses pengelompokan. Fitur hasil ekstraksi akan dikelompokkan menjadi dua kelas yaitu COVID-19 dan non COVID-19 (*healthy*). Metode yang digunakan adalah *k-Nearest Neighbor* (k-NN) dan dioptimasi oleh algoritma *Artificial Bee Colony* (ABC) untuk memperoleh nilai akurasi optimal berdasarkan kombinasi fitur ekstraksi. Hasil akhir yang diharapkan penulis yaitu dapat mendeteksi sinyal suara napas apakah tergolong dalam kelas COVID-19 atau non COVID-19 (*healthy*).

1.2 Rumusan Masalah

1. Berapa jumlah koefisien cepstral fitur MFCC yang terbaik dalam deteksi dini COVID-19 ?
2. Berapa jumlah populasi (N) pada metode ABC dan nilai *k* pada metode k-NN yang menghasilkan kinerja sistem terbaik?

1.3 Batasan Masalah

1. Menggunakan data sekunder yang diperoleh dari *database* Coswara yang merupakan sampel suara terdiri dari *breathing*, *cough* dan *voice sound*. Dikumpulkan melalui *crowd-sourcing* di seluruh dunia menggunakan situs *web* arXiv:2005.10548v2 pada penelitian Sharma *et.al* tahun 2020.
2. Data yang digunakan hanya sampel suara napas *deep* yang berasal dari suara napas naracoba COVID-19 dan non COVID-19 (*healthy*).

3. Keterangan data seperti jenis kelamin, usia, lokasi (negara, provinsi), status kesehatan terbaru (sehat/terpapar/sembuh/terinfeksi) dan penyakit bawaan tidak dijadikan pertimbangan dalam penentuan label sampel.
4. Jumlah koefisien cepstral fitur MFCC di variasikan menjadi 10, 13, dan 15.
5. Penelitian hanya mengklasifikasikan data suara napas menjadi dua kelas yaitu COVID-19 dan non COVID-19 (*healthy*) sebagai data latih dan uji.

1.4 Tujuan Penelitian

1. Mendapatkan jumlah koefisien cepstral fitur MFCC yang terbaik dalam deteksi dini COVID-19.
2. Mendapatkan jumlah populasi (N) pada metode ABC dan nilai k pada metode k-NN yang menghasilkan tingkat akurasi, sensitivitas dan spesifitas terbaik dari program untuk deteksi dini COVID-19.

1.5 Manfaat Penelitian

1. Menambah wawasan dalam penggunaan metode k-NN untuk deteksi dini COVID-19.
2. Sebagai bentuk pengembangan metode deteksi dini COVID-19 dengan memanfaatkan suara napas untuk membantu tenaga medis pada khususnya dalam menangani kasus COVID-19.
3. Menambah referensi mengenai algoritma yang dapat digunakan dalam deteksi dini COVID-19 bagi mahasiswa Teknik Biomedis khususnya dan mahasiswa Universitas Airlangga pada umumnya.