

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tahun 2019 *Novel Coronavirus* (2019-nCoV) muncul pertama kali di Wuhan China, mengakibatkan lebih dari 17.388 kasus yang dikonfirmasi dan 361 kematian. *Novel Coronavirus* (2019-nCoV) termasuk kedalam *betacoronavirus* yang cenderung mirip dengan SARS-Cov dan MERS-Cov (Xia et al., 2020). Tanggal 30 Januari 2020 *World Health Organization* (WHO) menyatakan secara darurat mengenai kesehatan global akibat peningkatan kasus di China dan internasional dengan adanya wabah virus corona baru yaitu *Severe Acute Respiratory Syndrome Coronavirus-2* (SARS-CoV-2) merupakan nama baru dari 2019-nCoV yang diberikan oleh WHO (Velavan & Meyer, 2020). SARS-CoV-2 merupakan virus yang menyebabkan penyakit COVID-19 (*Coronavirus Disease*) (Susilo et al., 2020). Penyakit yang disebabkan oleh virus corona baru (SARS-CoV-2) mengakibatkan infeksi pernapasan COVID 19 dengan gejala serius yang berisiko tinggi (Pascarella et al., 2020). WHO menyatakan bahwa COVID-19 dinyatakan sebagai pandemi global pada tanggal 24 Maret 2020 dikarenakan COVID-19 memenuhi tiga kriteria, yaitu : virus jenis baru, dapat menginfeksi banyak orang dengan mudah, dan secara meluas (Lella & Pja, 2021; Novidianto et al., n.d.). Berdasarkan data dari WHO secara global yang telah dikonfirmasi, pada tanggal 12 November 2021 terdapat 251.788.329 kasus COVID-19, termasuk jumlah kematian dengan angka 5.077.907 jiwa.

COVID-19 menyerang saluran pernapasan pada manusia yaitu pada sel yang melapisi alveoli. Seseorang yang terinfeksi akan mengalami beberapa gejala seperti demam, kelelahan, batuk kering, serta gangguan terhadap organ pernapasan seperti batuk, sesak napas, sakit tenggorokan, batuk darah, dan nyeri dada. Dari semua gejala tersebut, terdapat gejala yang sering ditemui yaitu demam dengan angka 83-98%, batuk dengan angka 76-82%, dan sesak napas (*dyspnea*) dengan angka 31-55% (Levani et al., 2021). Menurut Siti Nadia Tarmizi dari Kementerian Kesehatan, mayoritas pasien yang terinfeksi COVID-19 bergejala ringan dan tanpa bergejala

dengan angka 60-80%, sehingga diperlukan untuk memperbanyak jumlah pemeriksaan (*testing*) dan pelacakan (*tracing*).

Pemeriksaan COVID-19 pada tubuh seseorang dapat dilakukan dengan metode *rapid test* dan *Polymerase Chain Reaction* (PCR). *Rapid test* untuk deteksi COVID-19 dapat mendeteksi antibodi dan antigen. *Rapid test* antigen dapat mengetahui seseorang terinfeksi atau tidak dengan cepat dan mudah. Menurut WHO (WHO, 2020) tes ini akan memberikan hasil dalam 15-30 menit, bukan dalam beberapa jam atau hari. Tes antigen menggunakan hasil pemeriksaan *swab* hidung dan tenggorokan atau bisa menggunakan air liur. *Rapid test* antigen baik dilakukan ketika seseorang baru terinfeksi karena antibodi belum muncul untuk melawan virus penyebab COVID-19. Namun, dengan menggunakan tes ini dapat diperoleh hasil negatif palsu dan juga hasil positif palsu karena kemungkinan antigen bukan berasal dari virus penyebab COVID-19 apabila antibodi pada strip uji bereaksi dengan antigen virus selain COVID-19 (Yanti et al., 2020). Sedangkan pada *rapid test* antibodi, mendeteksi antibodi IgM dan IgG yang diproduksi tubuh dalam melawan virus. Jika antibodi terdeteksi dalam tubuh seseorang berarti tubuh orang tersebut pernah terpapar atau terinfeksi oleh virus penyebab penyakit COVID-19. Sampel yang digunakan untuk tes ini adalah darah. Namun, dalam pembentukan antibodi memerlukan waktu yang lama karena ketika seseorang terpapar virus penyebab penyakit COVID-19 tubuh tidak langsung memproduksi antibodi. Hal tersebut menyebabkan keakuratan *rapid test* antibodi sangat rendah dan menyebabkan hasil negatif palsu dan positif palsu (Yanti et al., 2020). Kedua tes tersebut masih kurang efektif dan akurat sehingga perlu dilakukan pengambilan *swab* untuk tes PCR agar memastikan apakah seseorang terinfeksi SARS-CoV-2. Tes PCR merupakan tes yang paling disarankan karena keunggulannya yang sangat tinggi berdasarkan spesifitas, efisiensi dan keakuratannya. Namun, tes PCR biayanya masih tergolong tinggi, memakan waktu dan melanggar jarak sosial saat pelaksanaan tes. Pemeriksaan yang ada saat ini bukan merupakan deteksi dini sehingga tidak dapat langsung dilakukan pemeriksaan ketika dibutuhkan saat itu juga. Pemeriksaan yang ada saat ini memerlukan bantuan petugas medis karena terdapat prosedur pemeriksaan yang hanya dapat diketahui oleh petugas medis.

Pemeriksaan dini yang diharapkan dapat dilakukan dimanapun dan kapanpun saat dibutuhkan serta tidak membutuhkan bantuan petugas medis dalam melakukan pemeriksaannya.

Oleh karena itu, banyak peneliti mencoba mengembangkan cara pemeriksaan untuk mendeteksi virus penyebab COVID-19 yang lebih efektif untuk pemeriksaan dini tanpa menggantikan pemeriksaan yang ada sebelumnya. Pengembangan pemeriksaan ini dapat membantu mengurangi risiko penyebaran COVID-19 karena dapat dilakukan secara mandiri (Rahman et al., 2022). Menurut Lela & Pja (2021), COVID-19 dapat diklasifikasikan berdasarkan tingkat keseriusannya menginfeksi sistem pernapasan yaitu ektrim, sedang, dan ringan. Hal tersebut menjadi perhatian para ilmuwan dan komunitas peneliti dalam masalah klasifikasi suara pernapasan untuk diagnosis penyakit COVID-19. Dalam mendiagnosis COVID-19 salah satunya dapat menggunakan suara pernapasan (*respiratory sound*) yang dihasilkan oleh tubuh seperti vibrasi, suara paru, jantung, pernapasan (breathing), batuk (cough), dan desahan (sighs) (Lella & Pja, 2021).

Semakin berkembangnya zaman yang diikuti oleh perkembangan teknologi, *Artificial Intelligence* hadir untuk mempermudah serta meringankan pekerjaan manusia, tidak terkecuali dalam dunia medis. Khususnya dalam mendiagnosa COVID-19 dapat memanfaatkan perkembangan teknologi seperti *Artificial Intelligence* (AI). *Artificial Intelligence* sendiri merupakan sebuah kecerdasan buatan yang dikembangkan oleh manusia untuk diterapkan pada suatu sistem. Teknologi ini dibuat seperti layaknya manusia yang dapat berfikir dan dapat melakukan pekerjaan dengan lebih efektif dan lebih efisien. *Artificial Intelligence* tidak akan berubah selama sistem komputer dan program tidak mengubahnya sehingga bersifat permanen. *Artificial Intelligence* yang biasa disebut kecerdasan buatan bersifat konsisten karena merupakan bagian dari teknologi komputer, berbeda dengan kecerdasan alami dapat berubah rubah (Lubis, 2021). Pada penelitian yang dilakukan oleh Astuti *Artificial Intelligence* digunakan dalam pemanfaatan penguatan sistem kesehatan sebagai pengendalian dalam kasus COVID-19 (Fitri Andri Astuti, 2021).

Banyak penelitian yang telah memanfaatkan *Artificial Intelligence* maupun sub bidang dari *Artificial Intelligence* yaitu *Machine Learning* untuk mendiagnosa COVID-19 dengan menggunakan suara napas. Menurut Sharma *et al* (2021) COVID-19 dapat didiagnosa dengan suara pernapasan menggunakan teknik pembelajaran mesin. Misalnya dalam sebuah penelitian yang dilakukan oleh Meng *et al.*, (2020) menggunakan 3 metode klasifikasi yaitu *Artificial Neural Network* (ANN), *Support Vector Machine* (SVM) dan *k - Nearest Neighbor* (k-NN) untuk mengklasifikasikan sinyal suara pernapasan. Menurut penelitiannya, menggunakan metode ANN akurasi lebih besar dibandingkan menggunakan metode SVM dan k-NN, dengan akurasi ANN yang didapatkan sebesar 85,43%. Penelitian lainnya dilakukan Lela & Pja (2021) dengan data dari coswara (*crowd sourced*) dan didapatkan akurasi mencapai 95,45% dengan menggunakan pengklasifikasian *Deep Convolutional Neural Networks* (DCNN) untuk mendeteksi penyakit COVID-19 dari suara pernapasan dengan fitur yang diterapkan adalah teknik *Denoising Auto Encoder* (DAE), GFCC (*Gamma-tone Frequency Cepstral Coefficients*), and IMFCC (*Improved Multi-frequency Cepstral Coefficients*). Penelitian yang dilakukan oleh Aly *et al* (2021) menggunakan dataset coswara dengan data suara batuk, pernapasan, dan ucapan didapatkan hasil akurasi 96% dengan menggunakan *binary classifiers* sederhana dan ekstraksi ciri MFCC. Dalam penelitiannya bertujuan untuk menarik perhatian masyarakat mengenai pentingnya suara pernapasan untuk diagnosis COVID-19 (Aly *et al.*, 2021). Selanjutnya, penelitian yang dilakukan oleh Palaniappan *et al* (2015) mengenai diagnosis patologi pernapasan, hasil akurasi keseluruhan menggunakan k-NN dengan *k-value* 1 adalah 94,99%. Pengklasifikasi SVM dengan kernel *Radial Basis Function* (RBF) mendapatkan akurasi 96,85% dan untuk pengklasifikasi *Extreme Learning Machine* (ELM) dengan kernel *Radial Basis Function* (RBF) diperoleh nilai akurasi 98,52%. Waktu yang dibutuhkan untuk melatih model juga lebih singkat pada pengklasifikasi ELM dibandingkan classifier SVM dan KNN (Palaniappan *et al.*, 2015). Sehingga dalam penelitiannya lebih unggul dengan menggunakan metode ELM.

Dari beberapa penelitian yang telah dilakukan, penulis melihat bahwa dengan menggunakan *Convolutional Neural Network* (CNN) terdapat kelemahannya yaitu dalam melakukan training memerlukan waktu yang lama dan beban komputasinya besar sehingga memerlukan *Graphic Processing Unit* (GPU) untuk mempercepat prosesnya. Sedangkan dengan menggunakan ANN memiliki permasalahan pada *learning speed* untuk *feedforward*. ANN *feedforward* memiliki *learning speed* yang rendah karena menggunakan *slow gradient based learning algorithm* dalam melakukan *training* dan parameter dalam jaringannya ditentukan secara *iterative* (berulang). Parameter tersebut adalah *input weight* dan *hidden bias*, parameter ini berhubungan antara satu layer dengan layer yang lain sehingga membutuhkan *learning speed* yang lama.

ELM merupakan ANN *feedforward* dengan *single hidden layer* dan biasa disebut *single hidden layer feedforward neural networks* (SLFNs). Dengan menggunakan metode ELM mendapatkan akurasi yg bagus dalam mendiagnosa COVID-19, memiliki *learning speed* yang lebih cepat dan mempunyai tingkat kesalahan yang rendah. Selain itu, diharapkan memperoleh hasil diagnosis yang lebih efektif karena ELM dibuat untuk mengatasi masalah pada ANN *feedforward* dalam hal *learning speed* (Ashar et al., 2018). Dalam pembelajaran *Conventional Gradient Based Learning algorithm* seperti *backpropagation* dan variannya *Lavenberg Marquadt* (LM), semua ANN *feedforward* menentukan parameternya secara manual, sedangkan ELM menentukan parameter *input weight* dan *hidden bias* secara acak sehingga memiliki *learning speed* yang baik.

Sehingga dengan kelebihan yang dimiliki oleh ELM maka dalam penelitian ini akan digunakan metode klasifikasi ELM. Penyelesaian masalah menggunakan ELM dimulai dengan melakukan normalisasi data, dilanjutkan dengan melakukan proses *training* dan *testing*, kemudian dilakukan perhitungan nilai *error* dengan menggunakan *Mean Square Error* (MSE). Namun ELM memiliki kelemahan karena jumlah dari *hidden nodes*-nya ditentukan secara *trial and error* (coba-coba) sehingga tidak dapat mengetahui jumlah *hidden nodes* yang optimal. Sehingga untuk mengatasi permasalahan tersebut maka akan digabungkan metode ELM

dengan metode optimasi ABC yang bertujuan menghasilkan *hidden notes* yang optimal sehingga hasilnya akurat (Handika et al., 2016).

Sebelum melakukan pengklasifikasian dengan menggunakan metode ELM, terlebih dahulu dilakukan ekstraksi fitur (*feature extraction*), dikarenakan pada metode ELM tidak memiliki seleksi fitur. Ekstraksi fitur mengekstraksi sinyal suara kedalam berbagai domain seperti domain waktu, domain frekuensi, domain morfologi dan statistik (Anggoro et al., 2020). Tujuan dilakukannya ekstraksi fitur adalah untuk memperjelas serta menajamkan pola dan mempermudah untuk memisahkannya ke dalam kategori kelas dalam proses klasifikasi (Umar et al., 2019). Ada beberapa metode untuk melakukan ekstraksi fitur seperti *Linear Predictive Coding* (LPC), *Melf-Frequency Cepstral Coefficients* (MFCC). Namun yang paling banyak digunakan dalam pemrosesan suara adalah MFCC karena metode ini dapat mempresentasikan sebuah sinyal dengan baik (Umar et al., 2019). MFCC dianggap sebagai fitur terpenting dalam membedakan suara napas COVID-19 dan non-COVID-19 (Alsabek et al., 2020). Metode MFCC dapat mengkonversi sebuah signal suara kedalam beberapa parameter dan sangat efektif untuk mengidentifikasi audio dalam pemodelan nada dan frekuensi dari sebuah *signal audio* (Idananta et al., 2018). Penelitian yang dilakukan oleh Alkhodari (2021) untuk mendeteksi COVID-19 dengan data suara napas dari coswara menggunakan ekstraksi fitur MFCC. Fitur ini dapat merepresentasikan sinyal suara dengan baik karena MFCC menyediakan satu set koefisien logaritma *descreate cosine transform* (DCT) dari spectrum sinyal *mel frequency cepstrum* (MFC) yang dianggap sebagai representasi dari keseluruhan informasi dalam sinyal dengan spektrum band yang berbeda, didapatkan akurasi pada data suara napas shallow sebesar 94.58% dan deep sebesar 92.08%. Selanjutnya dalam penelitian Idananta et al (2018) untuk mengenali siapa lawan bicara dengan menggunakan metode MFCC dihasilkan akurasi yang tinggi yaitu 92%. Penelitian lainnya yang telah disebutkan sebelumnya yaitu Aly et al (2021) menggunakan ekstraksi fitur MFCC didapatkan akurasi 96%. MFCC banyak digunakan untuk *speech technology* seperti *speaker recognition* ataupun *speech recognition*, dimana metode ini dapat menangkap informasi yang ada dalam sinyal suara, data seminimal mungkin tanpa menghilangkan informasi

penting, serta metode ini merupakan metode yang mengadaptasi organ pendengaran manusia dalam melakukan persepsi sinyal suara (Nasution, 2012). Sehingga dari pemaparan di atas akan digunakan ekstraksi ciri (*feature extraction*) dengan menggunakan metode MFCC untuk mengekstraksi suara napas.

Dalam penelitian ini fokus pada mendiagnosa atau mendeteksi virus penyebab penyakit COVID-19 dengan menggunakan data suara napas yang dibedakan menjadi dua kelas yaitu kelas pertama orang yang terpapar virus (positif COVID-19) dan kelas kedua suara napas non COVID-19 (*healthy*). Data yang digunakan dalam penelitian ini merupakan dataset dari *Coswara* menggunakan suara napas, dimana sampel ini dikumpulkan melalui *Crowdsourcing* diseluruh dunia dengan menggunakan aplikasi situs *web* (N. Sharma et al., 2020). Hanya dengan menggunakan *smartphone* kita dapat merekam sampel napas yang nantinya akan digunakan sebagai klasifikasi audio pernapasan (Meister et al., 2021). Dalam penelitian Alkhodari et al., (2021) menyarankan menggunakan suara pernapasan berbasis *smartphone* untuk mendeteksi COVID-19. Selain itu dapat dianggap sebagai diagnosis dini, cepat, mudah didistribusikan, hemat waktu, dan hampir tanpa biaya, serta tidak melanggar pembatasan jarak selama pandemi (Alkhodari & Khandoker, 2021). Data suara napas akan di ekstraksi dengan menggunakan metode MFCC untuk mendapatkan ekstraksi cirinya, hasil ekstraksi akan diklasifikasi menggunakan metode ELM dimana metode ELM akan digabungkan dengan metode optimasi *Artificial Bee Colony* (ABC) untuk menghasilkan jumlah *hidden nodes* yang optimal sehingga didapatkan akurasi yang terbaik dalam mendiagnosis suara napas orang non COVID-19 (*healthy*) dan orang yang terpapar virus (positif COVID-19). Dengan disusunnya penelitian ini diharapkan menjadi deteksi dini untuk mendeteksi virus penyebab penyakit COVID-19 dengan menggunakan suara napas.

1.2 Rumusan Masalah

1. Berapa jumlah *cepstral coefficient* fitur MFCC yang terbaik dalam deteksi dini COVID-19?

2. Berapa jumlah population size (N) pada metode ABC dan jumlah neuron di *hidden layer* “h” pada metode ELM yang menghasilkan kinerja terbaik?

1.3 Batasan Masalah

1. Data diperoleh dari database *Coswara*, yang terdiri dari breathing, cough dan voice sound yang dikumpulkan dari seluruh dunia melalui *Crowdsourcing* menggunakan situs web arXiv:2005.10548v2.
2. Data yang digunakan adalah data suara napas *deep* untuk sampel dengan dua kelas yaitu COVID-19 dan non COVID-19 (*healthy*)
3. Keterangan pada dataset berupa usia, jenis kelamin, lokasi (negara dan provinsi), status kesehatan saat ini (sehat/terpapar/sembuh/terinfeksi) serta penyakit bawaan diabaikan.
4. Jumlah *cepstral coefficient* fitur MFCC di variasikan antara 10, 13, dan 15.
5. Penelitian hanya mengklasifikasi data suara napas menjadi dua kelas yaitu COVID-19 dan non COVID-19 (*healthy*)

1.4 Tujuan

1. Mendapatkan jumlah *cepstral coefficient* fitur MFCC yang optimal dalam deteksi dini COVID-19 berbasis sinyal suara napas.
2. Mendapatkan jumlah population size (N) pada metode ABC dan jumlah neuron di hidden layer “h” pada metode ELM yang menghasilkan kinerja terbaik.

1.5 Manfaat

Manfaat dari penelitian ini diharapkan dapat menjadi deteksi dini dari virus penyebab COVID-19 dengan memberikan pemahaman adanya algoritma yang digunakan yaitu ekstraksi fitur MFCC, metode klasifikasi ELM, dan metode optimasi ABC. Sehingga dapat dikembangkan lebih lanjut dan dapat mempermudah tenaga medis dalam pendeteksian dini virus penyebab COVID-19.