

BAB I

PENDAHULUAN

Dalam bab ini akan dibahas latar belakang dilaksanakannya penelitian, rumusan masalah, batasan masalah, tujuan penelitian, dan sistematika penulisan.

1.1 Latar Belakang Penelitian

Perkembangan peradaban dan teknologi telah membawa perubahan dalam beraktivitas. Kehadiran teknologi yang semakin efektif dan efisien terus dikembangkan pada berbagai bidang yang salah satunya adalah bidang pengenalan biometrik.

Pengenalan biometrik merupakan pengenalan atau identifikasi berdasarkan keunikan karakteristik biologis, yang salah satunya adalah suara. Suara merupakan bentuk berinteraksi paling dasar yang kita digunakan dan saat ini digunakan juga untuk berinteraksi dengan perangkat teknologi. Pengenalan suara merupakan salah satu penerapan pengenalan biometrik tersebut. Pengenalan suara adalah teknik mengenali masukan berupa suara dan menghasilkan respon yang sesuai dengan masukan tersebut (Aswin Juari, 2010). Pengenalan suara terbagi dua, pertama pengenalan ucapan adalah mengenali kata atau kalimat yang diucapkan dan kedua pengenalan pembicara adalah mengenali siapakah pembicara yang mengucapkan kata atau kalimat tersebut (Tandyo, Martono, & Widyatmoko, 2008)

Pengenalan suara pada pengaplikasiannya digunakan antara lain untuk sistem kontrol perintah, sistem keamanan dan sistem absensi. Intonasi, frekuensi dan amplitudo merupakan karekteristik suara yang dapat dijadikan referensi pengenalan suara. Melalui pengenalan suara mempermudah kita melakukan berbagai aktivitas seperti penggunaan aplikasi *Google Voice*, *Siri* dan *Cortana*. Seiring pengaplikasian tersebut, penggunaan suara kini dapat menjadi suatu kebutuhan bagi kalangan pengguna teknologi. Maka berdasarkan hal itu memunculkan ketertarikan penulis untuk melakukan penelitian pengenalan suara

Mengenai penelitian pengenalan suara, terdapat berbagai teknik dan metode untuk melakukannya. Berikut beberapa penelitian pengenalan suara tersebut.

Pada tahun 2010 telah dibuat penelitian dengan judul *Deteksi Out Of Vocabulary (OOV) Menggunakan Hasil Pengenalan Suara Otomatis Untuk Bahasa Indonesia*. OOV dilakukan dengan transduksi fonem ke kata yang telah terklasifikasi melalui perubahan model bahasa dan evaluasi kamus data. Hasilnya adalah 77% untuk deteksi akurasi OOV dan peningkatan 4% dari modifikasi kamus (Aswin Juari, 2010).

Lalu pada tahun 2012 penelitian dengan judul *Pengenalan Ucapan Suku Kata Bahasa Lisan Menggunakan Ciri Linear Prediction Coding (LPC), Mel Frequency Cepstrum Coefficients (MFCC) dan Jaringan Syaraf Tiruan*. Penelitian dilakukan dengan memecah data kata ke bentuk suku kata bahasa lisan yang diharapkan mengurangi jumlah data kata yang akan dikenali, namun tetap mengakomodasi seluruh kata. Hasil penelitian ini yaitu dengan kemampuan pengenalan terbaiknya adalah 0.65%. Hal ini ternyata pendekatan ekstraksi ciri MFCC dan LPC tidak cocok untuk mengenali suara dengan jumlah target yang besar karena adanya permasalahan terhadap kesamaan bunyi pada huruf dan kata-kata tertentu (Arbiyono, 2012).

Kemudian pada tahun 2014 dengan judul *Convolution Neural Network (CNN) for Speech Recognition*. Penelitian ini menguji cobakan penerapan metode *deep learning* model CNN untuk pengenalan suara. Model CNN diterapkan dalam kerangka model *Hybrid CNN-HMM* dan digunakan penyaringan lokal dan *max-pooling* dalam domain frekuensi untuk menormalkan variasi pembicara supaya mencapai pengenalan yang baik. Hasilnya adalah model CNN dapat mencapai lebih dari 10% pengurangan kesalahan dan akurasi pengenalannya menjadi lebih tinggi (Abel-Hami, et al., 2014).

Selanjutnya pada tahun 2015 penelitian dengan judul *Implementasi Mel Frequency Cepstral Coefficient dan Vector Quantization Pada Pengenalan Suara Untuk Permainan Pesawat Arcade Berbasis Android*. Pengenalan suara pada

penelitian ini diimplemtasikan untuk mengontrol permainan dalam menghindari rintangan dan mendapatkan poin dengan diharuskan mengucapkan kata “kanan”, “kiri” atau “diam”. Proses pengenalan suara yang dilakukan adalah dengan membuat *codebook* terlebih dahulu dan kemudian suara yang akan dikenali dibandingkan dengan *codebook* yang ada. Hasilnya, tingkat akurasi pengenalan mencapai 96,67% dengan *codebook* dan dengan persentase keberhasilan kata “kanan”, “kiri” dan “diam” adalah 100%, 100% dan 90%. Sedangkan pada pengujian oleh responden adalah 80,3% tanpa *codebook* dengan persentase keberhasilan kata “kanan”, “kiri” dan “diam” adalah 88,5%, 79% dan 73,5%. Penurunan akurasi ini dikarenakan belum diterapkannya *noise cancelling* dan kalibrasi suara dengan suara pemain (Firdaus, 2015).

Dan pada tahun 2016 penelitian dengan judul *Pengenalan Suara Pada Sistem Notulen Rapat Menggunakan Convolution Neural Network*. Penelitian ini adalah CNN diimplemtasikan untuk mampu mengenali identitas seseorang dari suaranya. Pengenalan dan perhitungan akurasinya dilakukan dengan setiap data suara disamakan durasinya terlebih dahulu dan dikonversi kedalam bentuk *spectrogram image*, *scaling* serta *thresholding*. Hasilnya, rata-rata akurasi pengenalan adalah 80,33% dan untuk kekurangannya adalah disebabkan data suara yang kurang jelas, adanya *noise* dan juga masih terbilang kurangnya data untuk proses pelatihan (Mulyani, 2016).

Dari hasil kajian penelitian sebelumnya tersebut, permasalahan suara mempunyai tingkat variabilitas sangat tinggi, yaitu suara dan objek yang sama dapat menghasilkan data berbeda-beda, terlebih lagi untuk suara dan objek yang berbeda. Permasalahan ini menjadi tantangan dalam hal pengenalan suara. Pendekatan pengenalan pola (*pattern recognition*), pembelajaran mendalam (*deep learning*) dan kecerdasan buatan (*artificial intelligence*) dianggap sebagai cara paling kompetitif untuk digunakan pada penelitian tentang pengenalan suara.

Deep learning banyak digunakan dalam permasalahan pengenalan karena hasil dari akurasi yang cukup tinggi. *Convolution Neural Network* (CNN) merupakan salah satu model *deep learning* tersebut dan menjadi model yang akhir-akhir ini paling bagus dalam hal pengenalan.

CNN adalah *neuromorphic* sebuah model komputasi yang terinspirasi dari otak mamalia (Anwar, Hwang, & Sung, 2015). Model ini mempelajari pola dengan konsep dasar mengimitasi jaringan saraf makhluk hidup menggunakan sel besar dan sel kecil. Keunggulan model ini adalah ketelitiannya dan kemampuannya belajar dalam memecahkan hubungan rumit yang sulit dideskripsikan antara data masukan dan data keluaran. Hal ini dimungkinkan karena pengetahuan pada CNN merupakan jaringan saraf buatan tidak diprogram, tetapi dilatih berdasarkan informasi yang diterima.

Pada penerapannya CNN sudah teruji dalam berbagai pengenalan, seperti pada *Face Detection* di *Facebook* (Taigman, Yang, Ranzato, & Wolf, 2014), *Speech Recognition* di *Microsoft* (Hamid, et al., 2014), kompetisi *ImageNet Large Scale Visual Recognition Challenge* 2010-2013 (Krizhevsky, Sutskever, & Hinton, 2012). Serta juga diterapkan pada penelitian *speaker identification* dan *clustering* (Lukic, Vogt, Durr, & Stadelmann, 2016) dan *Pengenalan Suara Pada Sistem Notulen Rapat Menggunakan Convolution Neural Network* (Mulyani, 2016).

Berdasarkan uraian diatas, maka dari itu pada penelitian ini metode *deep learning* dengan model CNN adalah metode yang akan digunakan untuk membuat sistem yang dapat mengenali apa yang dibicarakan oleh pembicara. Penelitian dititik beratkan pada tingkat akurasi pengenalan data uji untuk mengetahui tingkat akurasi klasifikasi metode CNN sehingga diharapkan dapat menghasilkan pengenalan suara yang bagus untuk membantu dalam mengenali suara.

1.2 Rumusan Masalah Penelitian

Berdasarkan latar belakang masalah diatas, maka dapat diidentifikasi masalah sebagai berikut:

1. Bagaimana mengimplementasikan metode *deep learning* untuk pengenalan suara?
2. Bagaimana hasil implementasi metode tersebut?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dituliskan sebelumnya, maka tujuan dari penelitian ini adalah sebagai berikut :

1. Menghasilkan sistem pengenalan suara dengan mengimplementasikan metode *deep learning*.
2. Mengkaji hasil dari implementasi metode tersebut.

1.4 Batasan Masalah

Agar pembahasan penelitian ini lebih terarah maka diberikan batasan-batasan masalah sebagai berikut:

1. Metode *deep learning* yang diimplementasikan adalah *Convolution Neural Network (CNN)*.
2. Data suara yang di teliti hanya data rekaman pengucapan kata satu sampai dengan sepuluh.
3. Pengujian pengenalan suara hanya akan mengenali suara apa yang diucapkan, tidak sampai kepada siapa pemilik ucapan dan tidak dilakukan secara *real time*.

1.5 Sistematika Penulisan

Sistematika penulisan penelitian ini terdiri atas beberapa bab, diantaranya:

BAB I PENDAHULUAN

Bab ini berisi latar belakang penelitian, mengapa penelitian ini dilakukan. Selain itu dijelaskan juga rumusan masalah, batasan masalah, tujuan penelitian yang akan dilakukan, dan sistematikan penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini berisi beberapa hal mengenai terori-teori yang mendukung dan berhubungan dengan penelitian ini. Adapaun teori-teorinya adalah Pengertian Suara, Pengenalan Ucapan, Kecerdasan Buatan, Jaringan Syaraf Tiruan, *Deep Learning*, *Convolutional Neural Network* dan tentang penelitian yang relevan.

BAB III METODOLOGI PENELITIAN

Bab ini berisi penjelasan langkah-langkah yang akan dilakukan dalam penelitian. Seperti Desain Penelitian, Alat dan Data Penelitian.

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

Bab ini berisi uraian tentang hasil penelitian dan pembahasan terhadap hasil penelitian yang dilakukan yang meliputi hasil dari data penelitian yang dikumpulkan, implementasi metode, sistem yang dikembangkan serta hasil dari uji coba sistem berdasarkan skenarionya. Pada bagian akhir terdapat pembahasan dari setiap hasil penelitian.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan penelitian berdasarkan rumusan masalah yang telah disebutkan. Kemudian terdapat juga saran dan rekomendasi untuk penelitian selanjutnya berdasarkan kekurangan dari penelitian ini