

CHAPTER 20

KECERDASAN ARTIFISIAL DALAM *GENOME SEQUENCING*

Telly Kamelia

Kolaborasi Riset dan Inovasi Industri Kecerdasan Artifisial (KORIKA)

ABSTRAK

Materi genetik, baik berupa *deoxyribonucleic acid* (DNA) maupun *ribonucleic acid* (RNA), yang tersimpan dalam sel-sel penyusun tubuh disebut dengan genom. Informasi yang dibawa oleh genom merupakan informasi biologi yang diperlukan untuk membangun jaringan tubuh, mempertahankan hidup, dan mewariskan ke keturunan berikutnya. *Genome sequencing* merupakan suatu pengurutan nukleotida DNA dari suatu genom organisme. Saat ini, generasi terbaru menggunakan teknologi *next-generation sequencing* (NGS) yang dapat diaplikasikan dari level makro hingga mikro. Proses *genome sequencing*, aplikasinya pada berbagai bidang, dan pemanfaatan kecerdasan artifisial dalam prosesnya diuraikan dalam tinjauan literatur ini. Artikel ini adalah tinjauan literatur terhadap dua puluh tujuh artikel yang didapatkan dari basis data daring. Hasil dari berbagai studi menunjukkan terdapat berbagai metode pengurutan genom serta pengaplikasian pengurutan genom pada biologi, manusia, hewan, dan tumbuhan. Pengaplikasian *genome sequencing* atau genom berbasis kecerdasan artifisial juga telah dikembangkan, misalnya dalam penyebutan varian genom, penyuntingan genom, farmakologi molekuler, dan proses pertukaran data. Tinjauan ini menyimpulkan bahwa pengurutan genom berbasis kecerdasan artifisial dapat meningkatkan efektivitas suatu proses *genome sequencing*. Dibutuhkan kolaborasi antar ahli dalam bidang biologi, khususnya ahli genomik dan ahli kecerdasan artifisial, untuk menghasilkan suatu algoritma kecerdasan artifisial yang bermanfaat untuk masyarakat Indonesia.

Kata kunci: pengurutan genom, kecerdasan artifisial, *next-generation sequencing*, *biologi molekuler*

A. PENDAHULUAN

Setiap makhluk hidup memiliki materi genetik, baik berupa *deoxyribonucleic acid* (DNA) maupun *ribonucleic acid* (RNA), yang tersimpan dalam sel-sel penyusun tubuh yang disebut genom [1]. Genom merupakan suatu kesatuan dari materi genetik sebuah organisme yang mengandung pusat data sebuah organisme dalam menjalani

T. Kamelia

Kolaborasi Riset dan Inovasi Industri Kecerdasan Artifisial (KORIKA), e-mail: ilmiah.korika@gmail.com

©2023 Kolaborasi Riset dan Inovasi Industri Kecerdasan Artifisial (KORIKA) & Penerbit BRIN

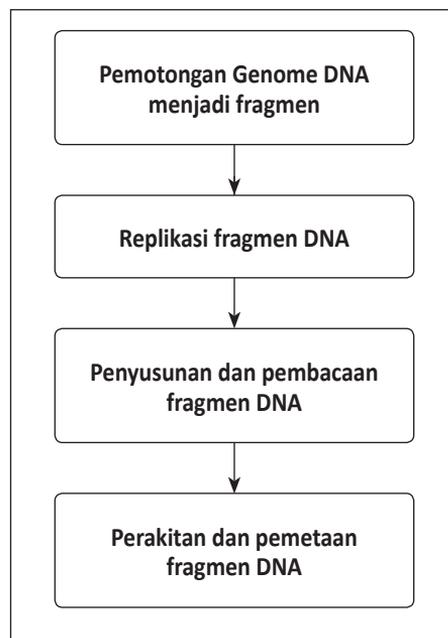
T. Kamelia, "Kecerdasan artifisial dalam *Genome Sequencing*," in *Prosiding Use Cases Artificial Intelligence Indonesia: Embracing Collaboration for Research and Industrial Innovation in Artificial Intelligence*, B. R. Trilaksono, H. Riza, A. Jarin, N. D. S. Darmayanti, and S. Liawatimena, Eds. Jakarta: Penerbit BRIN, Februari 2023, ch. 20, pp. 235-247, doi: 10.55981/brin.668.c556

ISBN: 978-623-8052-49-3, E-ISBN: 978-623-8052-50-9

fungsinya dalam kehidupan. Informasi yang dibawa oleh genom merupakan informasi biologi yang diperlukan untuk membangun jaringan tubuh, mempertahankan hidup, dan mewariskan ke keturunan berikutnya [2]. Pengurutan genom (*genome sequencing*) merupakan pengurutan dan perangkaian nukleotida DNA dari suatu genom organisme menggunakan komputasi dan algoritma yang kompleks.

Genome sequencing pertama kali dikenalkan oleh Frederick Sanger pada tahun 1975. Hal ini terus berkembang hingga pada tahun 1977 Frederick Sanger merumuskan sebuah metode pengurutan genom, yaitu metode *dideoxy*, diikuti oleh Maxam-Gilbert yang merumuskan sebuah metode pemutusan rantai [1][2][3]. Kedua metode ini dianggap menjadi generasi pertama dalam dunia pengurutan genom. Saat ini, generasi terbaru menggunakan teknologi *next-generation sequencing* (NGS) yang dapat diaplikasikan dari level makro hingga mikro, termasuk pengurutan seluruh *genom de novo* (*whole-genome sequencing*), pengurutan ulang genom, pembuatan profil mRNA, menganalisis evolusi molekuler, mendeteksi pola metilasi, memecahkan kasus kriminal, menjelajahi catatan, serta mengevaluasi protein pengikat DNA dan struktur kromatin [4].

Secara umum, pengurutan genom diawali dengan pemotongan rantai-rantai DNA yang nantinya akan disusun kembali sesuai dengan urutannya dari fragmen potongan DNA tersebut. Proses pengurutan genom secara singkat dapat dilihat pada Gambar 1. Metode pengurutan DNA yang pertama kali dirumuskan dinamakan Sanger *sequencing* atau metode pemutusan rantai. Metode ini menggunakan sintesis enzim untuk memecah rantai fragmen dari DNA. Sintesis enzim membutuhkan

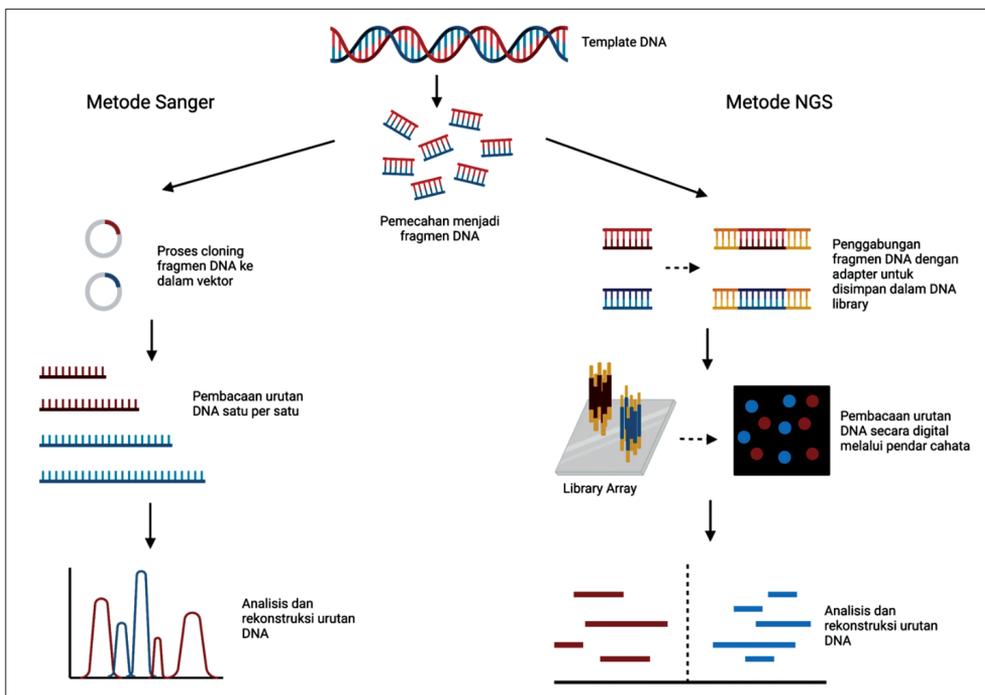


Gambar 1. Skema Singkat Pengurutan Genom

cetakan ssDNA yang akan diurutkan, primer, Taq polimerase untuk memperkuat cetakan DNA, *buffer*, *deoxynucleotida* (dNTPs), dan pewarna nukleotida, yaitu *dideoxynucleotide triphosphates* (ddNTPs). Metode ini cukup mahal dan memakan waktu yang lama untuk mendapatkan hasil yang baik untuk dibaca. Kunci utama dalam metode ini adalah penggunaan ddNTPs sebagai pemutus rantai DNA [3].

Metode pengurutan genom juga dikeluarkan oleh Maxam-Gilbert yang melibatkan bahan kimia radioaktif dalam prosesnya. Reagen kimia akan bekerja secara spesifik pada basa yang ada di dalam molekul DNA. Walaupun metode ini sangat cocok untuk pengurutan DNA yang heterogen, penggunaan bahan kimia radioaktif cukup berbahaya dan sulit untuk digunakan pada pengurutan DNA berskala besar. Metode Sanger *sequencing* masih lebih diminati dikarenakan lebih efisien [1][2][3].

Saat ini, terdapat metode pengurutan genom yang telah dimodifikasi sedemikian rupa sehingga tercipta sebuah metode yang efektif dan efisien dalam pengurutan genom, yaitu NGS. NGS menggantikan sintesis enzim dengan mesin sehingga proses pengurutan genom menjadi lebih cepat. Langkah awal dari metode ini adalah persiapan fragmen dari sampel DNA menjadi sebuah bentuk yang sesuai dengan data pengurutan. Persiapan dilakukan dalam sebuah *library preparation* dengan beberapa tahap, seperti fragmentasi DNA, pengurutan adaptor, seleksi ukuran, dan kuantifikasi, serta kontrol kualitas dari fragmen tersebut. Setelah melewati tahapan ini, fragmen tersebut akan dibawa ke sebuah mesin pengurutan DNA [1][3]. Saat ini, terdapat beberapa aplikasi pengurutan genom yang lazim digunakan, yaitu 454 GS Roche,



Gambar 2. Perbandingan Metode Sanger dan NGS

HiSeIllumina, dan Life Technology Solid 4, dengan kelebihan dan kekurangannya masing-masing. Bidang ilmu genomika saat ini terus berkembang dengan memanfaatkan perkembangan teknologi komputer, salah satunya kecerdasan artifisial, yang diaplikasikan untuk meningkatkan performa dari pengurutan genom ini.

B. PEMBAHASAN

1. Aplikasi Pengurutan Genom pada Biologi

Pengurutan genom dapat bermanfaat dalam perkembangan dunia obat-obatan, salah satunya dalam menganalisis interaksi obat dengan genom inang (simulasi molekular). Saat ini, tidak hanya identifikasi *strain* patogenik dan faktor virulensi, tetapi dapat melawan kedua hal tersebut menggunakan rekayasa genetik dan protein yang sesuai.

Terdapat proyek yang dikhususkan dalam menganalisis evolusi virus influenza melalui penyediaan data lengkap dari genom virus tersebut. Data ini dapat digunakan untuk pengembangan vaksin baru, pengobatan, dan diagnostik dari penyakit tersebut. National Center for Biotechnology Information Viral Genomes Project juga membentuk sebuah proyek yang menyediakan data molekuler lengkap mengenai diagnosis virus untuk digunakan sebagai bahan penelitian [6]. Selain proyek yang melibatkan virus, terdapat juga proyek yang menganalisis urutan genom spesies mikroba yang dinamakan *microbial genome sequencing*. Proyek ini mengumpulkan informasi mengenai fisiologis dan evolusi spesies mikroba yang nantinya akan digunakan sebagai pendekatan untuk diagnostik dan pengobatan dari penyakit infeksius yang disebabkan oleh mikroba tersebut.

Keanekaragaman hayati yang ada mencetuskan sebuah proyek Earth BioGenome Project, di mana membuat sebuah daftar urutan genom dan mengategorikan urutan genom tersebut. Hasil dari proyek ini dapat memberikan informasi dari berbagai masalah yang akan dihadapi, seperti dampak perubahan iklim terhadap keanekaragaman hayati, konservasi spesies, pelestarian ekosistem yang terancam punah, dan peningkatan ekosistem yang sudah ada [7].

2. Aplikasi Pengurutan Genom pada Manusia

Aplikasi dari pengurutan genom ini, antara lain pengurutan keseluruhan genom (*whole genome sequencing*), pengurutan target (*target sequencing*), pembuatan profil ekspresi gen, pengurutan presipitasi imun kromatin, dan pengurutan RNA. Pengaplikasian pengurutan genom dapat diterapkan dalam hal personalisasi terapi, penyakit keturunan, diagnosis beberapa penyakit yang disebabkan oleh virus maupun penyakit kanker. Beberapa contoh aplikasi klinis dari pengurutan genom ini adalah penemuan mutasi gen pembawa penyakit kanker, seperti BRCA1, BRCA2, dan TP53; penemuan *mismatch* dari gen MLH1, MSH2, MSH6, dan PMS2; penerapan *karyotyping* untuk mendiagnosis keganasan dalam bidang hematologi; analisis gen ERBB2 dalam kanker payudara; mutasi gen KRAS pada kanker *colorectal*; dan masih banyak lagi [8].

Salah satu penerapan pengurutan genom keseluruhan adalah Human Genome Project yang dimulai sejak tahun 1990 dan pertama kali terselesaikan pada tahun 2003. Program ini memberikan pengetahuan dengan pembacaan cetak biru (*blue-print*) genetik lengkap. Tujuan proyek ini adalah mengetahui lokasi gen pada semua kromosom dan memetakan gen-gen pewaris sifat dari tetua keturunannya. Hal ini berfungsi untuk mengetahui dan melacak berbagai penyakit yang diwariskan dari generasi ke generasi secara genetik [9].

Program lain yang mengaplikasikan pengurutan genom adalah The Cancer Genome Atlas Programme (TCGA), sebuah program genomik kanker yang akan mengarakterisasi 20.000 kanker utama dan mencocokkan dengan sampel normal secara molekular. Tujuan utama dibentuknya program ini adalah untuk meningkatkan kemampuan dalam diagnostik, pengobatan, dan pencegahan kanker.

Selain dalam bidang diagnostik, penerapan ilmu genomik dan kecerdasan artifisial juga diterapkan dalam bidang terapi. Salah satunya adalah penelitian yang dilakukan oleh Abedi dkk. yang mengembangkan sebuah inovasi dalam terapi kanker menggunakan bakteri *Escherichia coli* yang telah diprogram untuk membunuh hanya sel kanker [10]. Peneliti merekayasa bakteri *E. coli* menjadi dua kelompok gen, yaitu pertama ditugaskan untuk memproduksi *nanobodies* yang berfungsi sebagai protein untuk mematikan sinyal yang digunakan oleh sel tumor untuk mencegah datangnya respons dari sistem imun tubuh. Hal ini akan mengakibatkan sistem imun tubuh dapat menyerang sel tumor tersebut. Sementara itu, kelompok gen yang kedua adalah bakteri *E. coli* yang akan aktif mengeluarkan obat antikanker saat mengenai suhu 42–43°C, yang pada penelitian ini diaktifkan oleh gelombang *ultrasound* [10].

Melihat banyaknya manfaat yang didapatkan dari penerapan bidang ilmu genomik terhadap kehidupan manusia, hal ini dapat dijadikan ladang penelitian, khususnya di Indonesia, untuk mengetahui lebih dalam mengenai persebaran genomik dan kecenderungan ras tertentu yang lebih rentan terkena suatu penyakit sehingga dapat dilakukan suatu pencegahan sebelum menjadi sakit. Selain itu, pengurutan genom juga dimanfaatkan untuk mendiagnosis suatu penyakit turunan langka yang nantinya dapat digunakan sebagai alat prediksi kecenderungan dari hasil perkawinan dua subjek untuk memiliki turunan dengan kelainan genetik.

3. Aplikasi Pengurutan Genom pada Hewan

Penerapan bidang genomik juga banyak dilakukan pada proyek non-manusia, seperti The Mouse Genome Project yang memiliki tujuan membuat daftar variasi molekuler dari *strain* tikus yang digunakan laboratorium dan *strain* dari tipe liar.

Penerapan teknologi pengurutan genom juga diaplikasikan pada bidang agrikultur, yaitu hewan ternak, seperti ayam, babi, sapi, domba, dan kuda [11]. Salah satu hewan ternak yang pertama kali dilakukan pengurutan genom adalah ayam (*Gallus gallus*) yang merupakan hewan dengan laju rekombinasi tinggi sehingga ideal untuk

dilakukan penelitian [12]. Pengurutan genom yang dilakukan menjadikan dasar untuk perkembangbiakan keturunan ayam yang baik.

Selain ayam, pengurutan genom juga dilakukan pada babi (*Sus scrofa*). Pengurutan genom ini dapat dijadikan dasar identifikasi gen yang dapat digunakan untuk perkembangbiakan babi berkualitas dan menjadi dasar untuk melakukan penelitian terkait bidang medis manusia yang melibatkan hewan ini [11].

Perkembangan teknologi pengurutan genom yang semakin maju membuat analisis genom untuk hewan ternak ikut berkembang dengan tujuan untuk memilih turunan yang unggul, memiliki perkembangbiakan yang tinggi, dan mengetahui nutrisi terbaik yang dibutuhkan oleh hewan-hewan tersebut [11]. Hal ini dapat sangat bermanfaat mengingat Indonesia merupakan negara agrikultural yang sangat dekat dunia hewani. Adanya penelitian yang dapat membantu pemilihan turunan hewan yang unggul, terutama hewan dengan hasil produksi yang dapat dimanfaatkan, seperti ayam (daging dan telur) dan sapi (susu dapat diolah menjadi berbagai macam sumber makanan). Industri peternakan Indonesia dapat semakin berkembang dan maju yang tentunya juga akan meningkatkan kesejahteraan tidak hanya bagi para peternak, tetapi juga masyarakat karena mendapat pasokan protein hewani yang mencukupi.

4. Aplikasi Pengurutan Genom pada Tumbuhan

Selain dimanfaatkan untuk manusia dan hewan, pengurutan genom juga dilakukan pada tumbuhan. Pengurutan genom pertama kali dilakukan pada tumbuhan padi (*Oryza sativa*) [13][14] yang dibandingkan tumbuhan model dalam bidang genom, yaitu *Arabidopsis thaliana*. Pengurutan genom pada tumbuhan terus dilakukan seiring dengan kemajuan teknik yang ada, seperti pengurutan genom pada *Theobroma cacao* yang menggunakan teknologi Roche 454, genom tumbuhan apel yang menggunakan kombinasi dari metode Sanger dan Roche 454, serta pengurutan genom pada *Brassica rapa* yang juga menggunakan teknologi gabungan antara Sanger dan Illumina Solexa [15].

Pengurutan genom menjadi dasar dalam mengembangkan ilmu biologi dalam bidang tumbuhan yang bermanfaat bagi kehidupan manusia, salah satunya teknik rekombinan. Huang dkk. melakukan penelitian dengan menggabungkan padi *Oryza sativa* subsp. *indica* dengan *Oryza sativa* subsp. *japonica*. Penelitian ini menghasilkan satu bibit unggul hasil rekombinasi dari kedua tanaman tersebut [16]. Selain untuk menghasilkan bibit unggul, pengurutan genom juga dimanfaatkan untuk menghindari adanya tumbuhan dengan fenotipe yang buruk. Sebagai contoh, penelitian yang dilakukan Schneeberger dkk. yang melakukan *re-sequencing* pada tanaman *Arabidopsis thaliana* dengan fenotipe pertumbuhan lambat dan warna daun yang pucat. Hasil *re-sequencing* ini menemukan sebuah gen mutan yang ada pada tanaman tersebut [17].

Aplikasi ilmu genomik pada bidang tumbuhan juga bisa menjadi ladang penelitian di Indonesia yang kaya akan kekayaan floranya. Dalam bidang pertanian misalnya, jika ilmu genomik diterapkan, Indonesia dapat menghasilkan beras dari tumbuhan padi berbibit unggul yang telah ditemukan melalui identifikasi pengurutan genom sebelumnya. Selain dalam bidang pertanian, kekayaan flora yang eksotis yang ada di Indonesia dapat diidentifikasi dan diketahui kecocokan dengan habitat, bagaimana cara tumbuh dan kembangnya, dan sebagainya.

5. Kecerdasan Artifisial dan Pengurutan Genom

Kecerdasan artifisial adalah perkembangan dari sebuah sistem komputer yang dapat menjalankan tugas yang biasanya membutuhkan kecerdasan manusia. Saat ini perkembangan perangkat keras dan perangkat lunak berbasis kecerdasan artifisial semakin berkembang dan menarik untuk di aplikasikan ke dalam dunia biologi termasuk bidang genomik [18].

Kecerdasan artifisial diyakini dapat menyederhanakan dan mempercepat proses interpretasi pengurutan genom. Saat ini sistem pengurutan genom berbasis kecerdasan artifisial sedang dikembangkan. Masalah utama yang masih menjadi hambatan utama dalam pengaplikasian pengurutan genom dalam bidang diagnostik penyakit langka adalah adanya keterhambatan interpretasi penyakit di mana kecepatan dalam diagnosis sangat dibutuhkan untuk kasus-kasus penyakit langka pada anak baru lahir. Diagnosis yang cepat dan tepat akan mengubah prognosis dari anak tersebut [19].

Masalah yang timbul dalam interpretasi sistem berbasis kecerdasan artifisial dalam pengurutan genom, antara lain tipe kecerdasan artifisial apa yang paling sesuai untuk program ini, seperti Bayesian *networks*, *decision trees*, dan *neural nets*; perbandingan penggunaan sistem berbasis kecerdasan artifisial dengan yang belum apakah terdapat kenaikan performa seperti yang diharapkan; apa yang ditawarkan dari sistem berbasis kecerdasan artifisial ini baik untuk suatu hal yang baru maupun meningkatkan performa dari sistem yang sudah ada; serta bagaimana ketepatan hasil interpretasi sistem berbasis kecerdasan artifisial dalam mendiagnosis dan membuat sebuah keputusan [19].

Saat ini, model peranti lunak dan sistem kecerdasan artifisial yang mulai diaplikasikan adalah penerapannya sebagai perangkat analitis dalam ranah klinikal dan rekayasa genom, yaitu *computer vision*. Sistem ini dapat digunakan untuk penegakan diagnostik pada hasil pemeriksaan penunjang medis berbasis citra, salah satu contoh di Indonesia adalah Neurabot, alat diagnostik yang telah mengadopsi teknologi kecerdasan. Neurabot adalah alat citra diagnostik dari hasil olah data histopatologi dari sampel jaringan tubuh yang memiliki akurasi tinggi berdasarkan hasil pembelajaran pada *dataset* patologi anatomi yang cukup komprehensif.

Algoritma *deep learning* banyak digunakan dalam ranah genomik klinis yang berorientasi pada prediksi faktor risiko terjadinya berbagai kondisi patologis terkait dengan kondisi gen dan kompleksitas pemicu proses ekspresinya. Algoritma tersebut dapat memproses *dataset* genom dalam jumlah besar dengan kompleksitas varian yang amat beragam.

Contoh pengaplikasian sistem berbasis kecerdasan artifisial dalam proses analisis genom adalah proses penyebutan varian genom yang rentan terjadi kesalahan karena sangat tergantung pada persiapan sampel, teknologi yang digunakan, dan beberapa pengaruh biologis yang tidak dapat terprediksi. Sebuah algoritma berbasis kecerdasan artifisial dapat mempelajari kemungkinan bias yang akan terjadi sehingga dapat mengurangi kesalahan dalam proses penyebutan varian genom. Sebuah sistem kecerdasan artifisial berdasarkan konsep *convolutional neural networks* (CNN), DeepVariant, dapat melakukan pembacaan tanpa memiliki pengetahuan mengenai genomik dan hasilnya terbukti mengungguli alat standar yang biasa digunakan untuk mengidentifikasi varian genom [20]. Pola genomik hasil proses pengurutan dari alat berbasis CNN dapat dipetakan dan diolah sebagai bahan pertimbangan untuk mekanisme prediksi keluaran fenotipe ataupun kondisi kelainan patologi berbasis genetika. Pengetahuan ini akan menjadi langkah awal untuk menentukan langkah preventif dalam skala komunal dan sebagai strategi optimalisasi kualitas kesehatan masyarakat berbasis data biologis.

Setelah penyebutan varian, biasanya dilakukan proses interpretasi data genom dengan bergantung pada identifikasi varian genetik yang relevan dengan pengetahuan yang sudah ada sebelumnya. Algoritma berbasis kecerdasan artifisial dapat meningkatkan manfaat dari data pengetahuan sebelumnya dengan menginformasikan sebuah pemetaan dari fenotipe-genotipe.

Hubungan mutasi atau perubahan segmental secara genomik dapat dianalisis dengan *recurrent neural network* (RNN) dalam sebuah *natural language processing* (NLP) yang dapat mengenali potongan teks dan korelasinya dengan struktur genetik acuan melalui *interconnected hidden states*. Adanya konsep memori *long-short term memory network* pada RNN dapat menjadikan algoritma kecerdasan artifisial dalam konteks *time series*. Konsep tersebut efektif dalam memprediksi berbagai kemungkinan dari hasil olahan data yang bersifat berkelanjutan dan berkesinambungan, contohnya seperti pengurutan data genomik. Selain itu, analisis berbasis *time series* juga dapat digunakan untuk mengolah data yang berkesinambungan dari moda pemeriksaan elektrofisiologi, seperti elektrokardiografi (EKG) dan elektroensefalografi (EEG).

Salah satu contoh adalah *combined annotation dependent depletion approach* (CADD) yang menggabungkan beberapa fitur prediktif dalam *machine-learning* yang digunakan guna menghasilkan sebuah algoritma untuk memprediksi varian mutasi genetik. Sistem berbasis kecerdasan yang sudah menggunakan metode ini adalah DANN (*a deep learning approach for annotating the pathogenicity of genetic variants*)

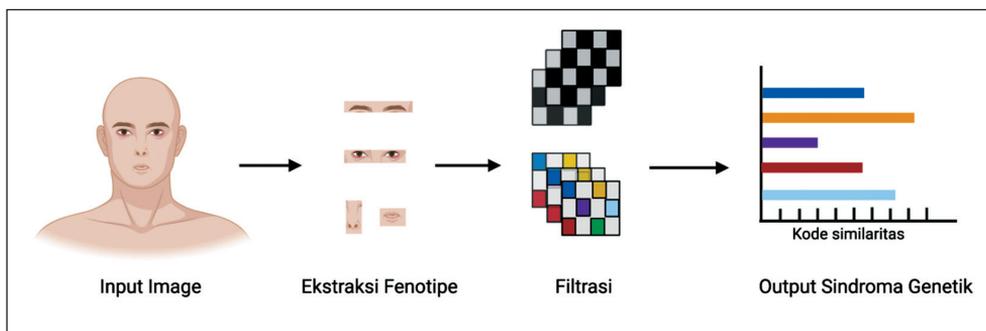
yang dapat menunjukkan hasil dalam meningkatkan performa dalam melakukan identifikasi varian dengan mutasi genetic [21].

Dalam proses identifikasi, PrimateAI, sebuah sistem berbasis kecerdasan artifisial dengan CNN yang dilatih pada varian patogen yang sudah diketahui, melakukan prediksi langsung dari data DNA atau urutan protein dengan fitur buatan tangan yang minimal [22]. Hasil dari prediksi oleh sistem PrimateAI terbukti mengungguli metode sebelumnya. Salah satu keunggulannya adalah alat ini dapat membedakan mutasi *de-novo* jinak dan patogen pada manusia yang berpotensi mengalami keturunan gangguan perkembangan [23].

Contoh lain adalah SpliceAI, sebuah sistem yang tersusun dari tiga puluh dua lapis *deep neural networks* yang mampu mengidentifikasi susunan genom pada gangguan perkembangan syaraf manusia dengan *input range* DNA yang lebih panjang sehingga menghasilkan prediksi dan akurasi sebesar 95% [24]. Pendekatan berbasis *deep-learning* juga digunakan dalam DeepSEA, sebuah sistem CNN yang dilatih dengan data genomik berskala besar sehingga mampu untuk melakukan pengurutan data secara multiple [25]. Salah satu pengaplikasiannya adalah penelitian Zhou dkk. yang mengidentifikasi mutasi genetik *non-coding* pada 1.790 manusia dengan *autism spectrum disorders* [26].

Kolaborasi antara kecerdasan artifisial proses diagnosis penyakit genetik juga diterapkan pada alat DeepGestalt, sebuah alat *computer vision* berbasis CNN untuk mendiagnosis sebuah penyakit kelainan genetik dari analisis gambar wajah [27]. Input dari alat ini adalah gambar wajah dengan fenotipe dari berbagai sindrom gangguan genetik yang ada dan selanjutnya dijadikan data dasar dari alat ini. Skema sederhana cara kerja alat ini digambarkan pada Gambar 3.

Di Indonesia, salah satu lembaga swasta penyelenggara jasa layanan pemeriksaan penunjang medik yang telah memperkenalkan layanan genomik berbasis teknologi adalah Prodia Wellness Genomic (PWG). PWG merupakan pemeriksaan yang dilakukan untuk menentukan risiko dari berbagai jenis penyakit berdasarkan profil genomiknya, seperti risiko penyakit kanker, diabetes, hipertensi, kardiovaskular, autoimun, dan muskuloskeletal. Apabila sistem pengambilan data genomik ini



Gambar 3. Alur Kerja Sederhana Pengaplikasian Kecerdasan Artifisial

dapat diintegrasikan dengan penerapan teknologi berbasis kecerdasan artifisial dalam pengelolaan dan pengolahannya, dapat dipastikan adanya hasil prediksi patologi dengan tingkat akurasi dan validasi yang sangat tinggi.

Dalam bidang mikrobiologi, penerapan sistem berbasis kecerdasan artifisial dapat dilihat pada dikembangkannya teknologi *genomic editing*. Sesuai dengan hasil riset Emanuelle Charpentier dan Jennifer Doudna, penerima Nobel Kimia 2020, yang menemukan sebuah “gunting genetika”, yaitu CRISPR Cas9. Implikasi *genomic editing* sangatlah besar dan mengundang beberapa pendapat pro dan kontra. Salah satunya adalah kasus He Jiankui pada tahun 2018 yang mengaku telah melakukan *genomic editing* pada embrio yang diimplantasikan ke rahim manusia. Dari proses ini lahir seorang bayi yang menghentak dunia karena *genomic editing* yang telah dilakukan. Hal ini memunculkan kesadaran bahwa modifikasi DNA adalah hal yang nyata dan dapat dilakukan.

Selain dalam bidang molekuler, pendekatan sistem berbasis kecerdasan artifisial bidang genomik dapat diterapkan dalam ranah farmakologi molekuler. Biasanya dalam mendesain suatu obat memerlukan waktu dan proses yang berliku, mulai dari pencarian dan penemuan zat aktif yang tepat dan sesuai dengan kondisi patologi dari suatu penyakit, proses ekstraksi atau pemisahan zat aktif berpotensi obat, uji toksisitas, uji efikasi, hingga ke berbagai jenjang uji klinis dan *post-market*, penerapan sistem berbasis kecerdasan akan memotong penemuan sebuah obat sehingga dapat tercipta suatu obat dalam waktu yang pendek. Proses yang panjang dan berliku, selain memakan waktu, juga membutuhkan sumber daya yang besar. Pendekatan berbasis kecerdasan artifisial dapat dimulai dengan analisis potensi zat aktif dari berbagai *dataset* sumber natural, seperti tumbuhan, jamur, dan mikroba. Uji kesesuaian dengan struktur molekul reseptor atau pun target organ dapat dilakukan secara *in silico* melalui mekanisme *molecular docking*, serta prediksi atau perkiraan teoretis terkait toksisitas dan efektivitas sifat terapeutiknya dapat disimulasikan oleh sistem bioinformatika berbasis kecerdasan artifisial.

Selain dalam bidang diagnostik dan terapi, penerapan kecerdasan artifisial juga dapat diterapkan dalam proses pertukaran data. Saat ini, penerapan protokol standar pertukaran data kesehatan di Indonesia juga mulai diinisiasi oleh Digital Transformation Office (DTO) Kemenkes RI. DTO menggunakan format *fast healthcare interoperability resource* (FHIR) yang dikeluarkan oleh Health Level Seven International atau HL-7. Pengelolaan data akan sangat teroptimalkan jika mengadopsi teknologi berbasis kecerdasan artifisial ini. Pertukaran data dengan format terstandarisasi dan penerapan *application programming interface* (API) akan menjembatani transfer data antaraplikasi serta memberikan peluang terciptanya berbagai *dataset* kesehatan dalam bentuk modular yang terstandar.

Pendekatan melalui NLP dapat mengekstrak data teks atau narasi deskriptif dari *electronic health record*, sementara algoritma *deep neural network*, dengan *hidden layer* dan sifat *interconnected*, dapat menjadi peranti analitis untuk mengolah data gabungan yang berasal dari pengolahan citra dengan *computer vision*, data serial dari pengurutan genom, dan data teks dari rekam medis.

Penerapan teknologi berbasis kecerdasan artifisial dalam ranah genomik juga dapat dimanfaatkan sebagai bagian dari perumusan kebijakan kesehatan dari tingkat individual, komunal, lokal, dan regional, hingga tingkat nasional. Hal ini dapat diterapkan melalui optimalisasi pengelolaan dan pengolahan data genomik, khususnya yang terkait dengan faktor risiko penyakit dan dinamika perilaku. Hasil dari analisis data tersebut dapat digunakan sebagai asupan dalam mekanisme perumusan strategi kebijakan nasional yang tepat sasaran, khususnya dalam pengelolaan status epidemiologi dan optimalisasi pemanfaatan sumber daya.

Kemajuan bidang teknologi, terutama dalam bidang manufaktur peranti keras, dapat memengaruhi kinerja sistem dalam menghadirkan fungsi dan utilitas kecerdasan artifisial. Hal ini akan mempercepat kemajuan riset dan pelayanan di bidang kesehatan dan bioteknologi. Contoh nyata adalah keberhasilan ilmuwan di India Institute of Science (IISc) dalam mengembangkan algoritma *machine learning berbasis graphics processing unit* (GPU) model terbaru dan berkekuatan tinggi serta mengintegrasikannya ke dalam aplikasi pengolah data yang dinamai Regularized, Accelerated, Linear Fascicle Evaluation, atau dapat disebut ReAL LiFe. Aplikasi ini dapat mengolah unit data dari *diffusion magnetic resonance* atau dMRI otak yang nantinya akan menghasilkan gambaran konektivitas antarunit neuron di otak saat menjalankan fungsi neurofisiologis tertentu. Koneksi fungsional antarunit neuron tersebut dikenal sebagai *connectome*. Berbekal kapasitas GPU yang telah diaugmentasi tersebut, tim IISc berhasil memetakan *connectome* otak terkait dengan berbagai fungsi neurofisiologis tertentu.

C. KESIMPULAN

Pengaplikasian sistem berbasis kecerdasan artifisial dapat menjadi nilai tambah tersendiri dalam perkembangan ilmu genomik di Indonesia. Alat berbasis kecerdasan artifisial yang diterapkan dapat meningkatkan keefektifan pengurutan genom. Kolaborasi antarahli dalam bidang biologi, khususnya genomik dan ahli dalam kecerdasan artifisial, dibutuhkan untuk menghasilkan sebuah karya yang akan memberikan banyak manfaat terutama untuk masyarakat Indonesia. Jika berbagai potensi pemrosesan data berbasis kecerdasan artifisial dapat diintegrasikan dalam satu platform terpadu, seperti layanan Cloud Life Science atau CLS yang disediakan oleh Google, maka besar harapan akselerasi pemanfaatan kecerdasan artifisial dalam ranah genomik dan kesehatan akan meningkat secara eksponensial.

DAFTAR PUSTAKA

- [1] Thankachan A dan M. B. Thomas, "A study of next generation sequencing data, workflow, application and platform comparison," *IOP Conference Series: Materials Science and Engineering*, vol. 396, 2018. doi:10.1088/1757-899X/396/1/012031
- [2] J. M. Keith, *Bioinformatics: Data, Sequence Analysis and Evolution*. New York: Springer, 2017. doi:10.1007/978-1-4939-6622-6
- [3] C. S. Pareek, R. Smoczynski, dan A. Tretyn, "Sequencing technologies and genome sequencing," *Journal of Applied Genetics*, vol. 52, no. 4, pp. 413–435, 2011. doi:10.1007/s13353-011-0057-x
- [4] J. W. Prokop dkk., "Genome sequencing in the clinic: The past, present, and future of genomic medicine," *Translational Physiology Physiol Genomics*, vol. 50, no. 8, pp. 563–579, 2018. doi: 10.1152/physiolgenomics.00046.2018
- [5] N. -W. Zhao, "Recent progress in the methods of genome sequencing," *Brazilian Archives of Biology and Technology*, vol. 53, no. 2, pp. 319–325, 2010. doi:10.1590/S1516-89132010000200010
- [6] Y. Bao dkk., "National center for biotechnology information viral genomes project," *Journal of Virology*, vol. 78, no. 14, pp. 7291–7298, 2004. doi:10.1128/JVI.78.14.7291-7298.2004
- [7] H. A. Lewin dkk., "Earth BioGenome Project: Sequencing life for the future of life," dalam *Proceedings of the National Academy of Sciences*, vol. 115, no. 17, 2018, pp. 4325–4333. doi:10.1073/pnas.1720115115
- [8] B. Pasche dan D. Absher, "Whole-genome sequencing: A step closer to personalized medicine," *JAMA*, vol. 305, no. 15, pp. 1596–1597, 2011. doi:10.1001/jama.2011.484
- [9] E. S. Lander dkk., "Initial sequencing and analysis of the human genome international human genome sequencing consortium," *Nature*, vol. 409, pp. 860–921, 2001. <https://doi.org/10.1038/35057062>
- [10] M. H. Abedi dkk., "Ultrasound-controllable engineered bacteria for cancer immunotherapy," *Nature Communications*, vol. 13, 2022, Art. No. 1585. doi:10.1038/s41467-022-29065-2
- [11] Y. Bai, M. Sartor, dan J. Cavalcoli, "Current status and future perspectives for sequencing livestock genomes," *Journal of Animal Science and Biotechnology*, vol. 3, no. 1, 2012. doi:10.1186/2049-1891-3-8
- [12] D. W. Burt, "Chicken genome: Current status and future opportunities," *Genome Research*, vol. 15, no. 12, pp. 1692–1698, 2005. doi:10.1101/gr.4141805
- [13] S. A. Goff dkk., "A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*)," *Science*, vol. 296, no. 5565, pp. 92–100, 2002. doi:10.1126/science.1068275

- [14] J. Yu dkk., "A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*)," *Science*, vol. 296, no. 5565, pp. 79–92, 2002. doi:10.1126/science.1068037
- [15] D. Edwards dan J. Batley, "Plant genome sequencing: Applications for crop improvement," *Plant Biotechnology Journal*, vol. 8, no. 1, pp. 2–9, 2010. doi:10.1111/j.1467-7652.2009.00459.x
- [16] X. Huang dkk., "High-throughput genotyping by whole-genome resequencing," *Genome Research*, vol. 19, no. 6, pp. 1068–1076, 2009. Doi:10.1101/gr.089516.108
- [17] K. Schneeberger dkk., "SHOREmap: simultaneous mapping and mutation identification by deep sequencing," *Nature Methods*, vol. 6, no. 8, pp. 550–551, 2009. doi:10.1038/nmeth0809-550
- [18] R. Dias dan A. Torkamani, "Artificial intelligence in clinical and genomic diagnostics," *Genome Medicine*, vol. 11, no. 1, 2019. doi:10.1186/s13073-019-0689-8
- [19] F. M. de La Vega dkk., "Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases," *Genome Medicine*, vol. 13, no. 1, 2021. doi:10.1186/s13073-021-00965-0
- [20] R. Poplin dkk., "A universal SNP and small-indel variant caller using deep neural networks," *Nature Biotechnology*, vol. 36, no. 10, pp. 983–987, 2018. doi:10.1038/nbt.4235
- [21] D. Quang, Y. Chen, dan X. Xie, "DANN: A deep learning approach for annotating the pathogenicity of genetic variants," *Bioinformatics*, vol. 31, no. 5, pp. 761–763, 2015. doi:10.1093/bioinformatics/btu703
- [22] L. Sundaram L dkk., "Predicting the clinical impact of human mutation with deep neural networks," *Nature Genetics*, vol. 50, no. 8, pp. 1161–1170, 2018. doi:10.1038/s41588-018-0167-z
- [23] M. J. Landrum dkk., "ClinVar: Improving access to variant interpretations and supporting evidence," *Nucleic Acids Research*, vol. 46, no. D1, pp. D1062–D1067, 2018. doi:10.1093/nar/gkx1153
- [24] K. Jaganathan dkk., "Predicting splicing from primary sequence with deep learning," *Cell*, vol. 176, no. 3, pp. 535–548, 2019. Doi:10.1016/j.cell.2018.12.015
- [25] B. E. Bernstein dkk., "The NIH roadmap epigenomics mapping consortium," *Nature Biotechnology*, vol. 28, no. 10, pp. 1045–1048, 2010. doi:10.1038/nbt1010-1045
- [26] J. Zhou dkk., "Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk," *Nature Genetics*, vol. 51, no. 6, pp. 973–980, 2019. Doi:10.1038/s41588-019-0420-0
- [27] Y. Gurovich dkk., "Identifying facial phenotypes of genetic disorders using deep learning," *Nature Medicine*, vol. 25, no. 1, pp. 60–64, 2019. doi:10.1038/s41591-018-0279-0