Institut Riset dan Publikasi Indonesia (IRPI)
MALCOM: Indonesian Journal of Machine Learning and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/malcom
Vol. 4 Iss. 3 July 2024, pp: 875-884
ISSN(P): 2797-2313 | ISSN(E): 2775-8575

Review of Supervised Reinforcement Learning on Medical Actions for Diabetes Mellitus

Tinjauan Supervised Reinforcement Learning pada Tindakan Medis Penyakit Diabetes Melitus

Indah Pratiwi Putri[1*], Dona Marcelina[2], Evi Yulianti[3]

[1,2,3]Program Studi SIstem Informasi, Universitas Indo Global Mandiri, Indonesia

E-Mail: [1]wiwid@uigm.ac.id, [2]donamarcelina@uigm.ac.id, [3]eviyulianti@uigm.ac.id

Received April 29th 2024; Revised May 11th 2024; Accepted May 22th 2024
Corresponding Author: Indah Pratiwi Putri

Abstract

Diabetes Mellitus (DM) is a chronic disease that requires ongoing medical management. Management of diabetes control depends on blood glucose levels in order to take appropriate action to prevent blood glucose levels from becoming too low or high. In the context of DM medical care, the use of machine learning technology, especially Supervised Reinforcement Learning (SRL) has presented an innovative approach. This research aims to investigate and summarize several scientific works that discuss the application of SRL in the context of medical procedures for DM. Several experiments were carried out by the researchers using data from diabetes patients to determine optimal model parameters, conducting simulations and validation studies in real-time so as to provide further insight into the practical application of reinforcement learning models in clinical settings. Through SRL, learning agents can combine environmental feedback with explicit information from supervisors to produce optimal decisions in DM management. In this paper, the authors analyze the literature review regarding the application of SRL to the medical management of DM, and explore the potential and challenges associated with the use of this approach in clinical practice.

Keywords: AI, Deep Reinforcement Learning, Diabetes Mellitus, Medical Treatment, Supervised Reinforcement Learning

Abstrak

Diabetes Melitus (DM) merupakan penyakit kronis yang memerlukan pengelolaan medis yang berkelanjutan. Pengelolaan pengendalian penyakit diabetes bergantung pada kadar glukosa dalam darah guna mengambil tindakan yang tepat agar dapat mencegah kadar glukosa darah menjadi terlalu rendah atau tinggi. Dalam konteks perawatan medis DM, penggunaan teknologi pembelajaran mesin, khususnya Supervised Reinforcement Learning (SRL) telah mengahadirkan pendekatan yang inovatif. Penelitian ini bertujuan untuk menyelidiki dan merangkum beberapa karya ilmiah yang membahas tentang penerapan SRL dalam konteks tindakan medis untuk penyakit DM. Beberapa percobaan dilakukan oleh para peneliti dengan menggunakan data dari pasien diabetes untuk menentukan parameter model yang optimal, melakukan simulasi dan studi validasi secara real-time sehinga dapat memberikan wawasan lebih lanjut tentang penerapan praktis model pembelajaran penguatan dalam pengaturan klinis. Melalui SRL, agen pembelajaran dapat menggabungkan umpan balik lingkungan dengan informasi eksplisit dari supervisor untuk menghasilkan keputusan yang optimal dalam pengelolaan DM. Dalam makalah ini, penulis menganalisis kajian literatur terkait penerapan SRL pada pengelolaan medis DM, serta mengeksplorasi potensi dan tantangan yang terkait dengan penggunaan pendekatan ini dalam praktik klinis

Kata Kunci: AI, Deep Reinforcement Learning, Diabetis Melitus, Supervised Reinforcement Learning, Tindakan Medis

1. PENDAHULUAN

Dalam layanan kesehatan, proses diagnosa klinis selalu bersifat dinamis karena tingginya prevalensi penyakit kompleks dan perubahan dinamis dalam status klinis pasien [1]. Sistem rekomendasi pengobatan yang diterapkan saat ini menggunakan protokol berbasis aturan yang ditentukan oleh dokter berdasarkan pedoman klinis berbasis data rekam medis. Selain itu, protokol dan rekomendasi tradisional ini mungkin tidak memperhitungkan beberapa aspek seperti gejala penyerta. Dokter biasanya mengandalkan rekomendasi dari beberapa hasil uji medis laboratorium, tinjauan sistematis, dan analisis pada saat mereka.

Diabetes Melitus (DM) adalah penyakit kronis yang memerlukan pengelolaan medis yang berkelanjutan untuk mengurangi komplikasinya dan memastikan hasil optimal bagi pasien [1]. Kompleksitas pengelolaan DM menuntut pendekatan inovatif dalam pengobatan dan perawatan. Dalam beberapa tahun terakhir, teknik pembelajaran mesin telah muncul sebagai alat yang menjanjikan dalam domain kesehatan, menawarkan solusi potensial untuk mengoptimalkan proses pengambilan keputusan medis. Di antara teknik-teknik tersebut, Supervised Reinforcement Learning (SRL) menyajikan perpaduan unik antara prinsip-prinsip pembelajaran penguatan dengan paradigma pembelajaran terawasi, memungkinkan integrasi umpan balik lingkungan dengan bimbingan eksplisit dari supervisor.

Tujuan utama dari penelitian ini adalah untuk mengeksplorasi penggunaan SRL dalam prosedur medis untuk mengelola DM. Penelitian ini dimotivasi oleh kebutuhan untuk mengatasi sifat kronis dan kompleks dari DM, yang menyoroti kebutuhan akan strategi penanganan penyakit yang inovatif. Penelitian ini bertujuan untuk memberikan wawasan tentang penerapan dari model pembelajaran reinforcement dalam pengaturan klinis. Tujuan utama dari penelitian ini adalah untuk menyelidiki dan merangkum aplikasi SRL dalam konteks prosedur medis untuk DM. Motivasi di balik penelitian ini adalah untuk mengatasi sifat kronis dan kompleks dari DM, dengan menekankan kebutuhan akan pendekatan inovatif dalam manajemen medis. Penelitian bertujuan untuk memberikan wawasan lebih lanjut tentang aplikasi praktis dari model pembelajaran reinforcement dalam pengaturan klinis, khususnya dalam manajemen DM. Dampak dari penelitian ini terletak pada potensinya untuk membandingkan beberapa solusi komprehensif dan efektif untuk mengontrol dan menekan kadar glukosa darah pada pasien DM, yang pada akhirnya akan mengarah pada peningkatan kepatuhan pasien terhadap perawatan diabetes dan saran pengobatan yang lebih baik untuk pasien. Penggunaan SRL sebagai model dalam penelitian ini sangat penting karena kombinasi uniknya dari prinsip-prinsip dari reinforcement learning dan supervised learning, yang memungkinkan integrasi umpan balik lingkungan artifisial dengan informasi eksplisit dari supervisor untuk menghasilkan keputusan optimal dalam manajemen DM [2],[3]. Pendekatan ini menawarkan solusi yang menjanjikan dan inovatif untuk mengoptimalkan proses pengambilan keputusan dalam manajemen medis DM.

Dalam konteks pengelolaan DM dimana keputusan pengobatan bersifat multifaset dan dinamis, penerapan SRL dapat memberikan dampak yang signifikan. Dengan memanfaatkan SRL, praktisi kesehatan dapat memanfaatkan sejumlah data historis pasien dan beragam keahlian praktisi klinis untuk menavigasi kompleksitas perawatan DM secara lebih efektif [4]. Penelitian ini bertujuan untuk memberikan gambaran tentang penerapan SRL dalam ranah intervensi medis untuk DM. Melalui tinjauan literatur yang sistematis, kami bertujuan untuk menjelaskan lanskap saat ini dari implementasi SRL dalam pengelolaan DM, menyoroti manfaat potensial dan tantangan dalam praktik klinis. Dengan mengeksplorasi perkembangan keilmuan pembelajaran mesin dan pengambilan keputusan medis dalam konteks pengelolaan DM, penelitian ini berusaha memberikan tinjauan dari beberapa metode spesifik pembelajaran mesin pada kemajuan pendekatan inovatif guna meningkatkan personalisasi perawatan dan pengobatan pada pasien DM.

2. TINJAUAN LITERATUR

2.2 Diabetes Melitus

Menurut [5] Diabetes Mellitus (DM) adalah gangguan metabolisme yang ditandai oleh hiperglikemia atau tingkat glukosa darah yang tinggi, yang disebabkan oleh ketidakmampuan tubuh untuk memproduksi atau menggunakan insulin dengan baik. DM adalah penyakit kompleks dan kronis yang membutuhkan manajemen medis yang berkelanjutan, dan dapat menyebabkan komplikasi serius jika tidak diobati. DM diklasifikasikan menjadi empat jenis berdasarkan penyebab dan perkembangannya [6], yakni:

1. Diabetes tipe 1, juga dikenal sebagai diabetes juvenil atau diabetes tergantung insulin, disebabkan oleh penghancuran sel beta di pankreas, yang memproduksi insulin. Jenis DM ini biasanya terjadi pada anak-anak dan dewasa muda, dan memerlukan terapi insulin seumur hidup.
2. Diabetes tipe 2, juga dikenal sebagai diabetes non-insulin-dependent atau diabetes pada orang dewasa. Ini terjadi ketika tubuh menjadi resisten terhadap insulin atau ketika pankreas tidak dapat menghasilkan cukup insulin untuk mempertahankan tingkat glukosa normal. Diabetes tipe 2 biasanya berkembang pada orang dewasa, tetapi semakin sering didiagnosis pada anak-anak dan remaja karena meningkatnya tingkat obesitas. Diabetes tipe 2 biasanya dipengaruhi pola gaya hidup, seperti diet dan olahraga, tetapi beberapa orang mungkin memerlukan obat atau terapi insulin.
3. Diabetes mellitus gestasional (DMG) adalah bentuk DM yang terjadi selama kehamilan. Ini disebabkan oleh resistensi insulin, yang merupakan perubahan fisiologis normal selama kehamilan. DMG biasanya hilang setelah persalinan, tetapi wanita yang pernah mengalami DMG berisiko lebih tinggi untuk mengembangkan diabetes tipe 2 di kemudian hari.
4. Jenis DM lainnya meliputi bentuk-bentuk monogenik diabetes, seperti diabetes neonatal dan diabetes muda (MODY), dan diabetes karena penyebab lain, seperti operasi, obat-obatan, infeksi, dan sindrom genetik.

Gejala DM meliputi haus dan buang air kecil yang meningkat, nafsu makan yang meningkat, kelelahan, penglihatan kabur, dan penyembuhan luka yang lambat. Namun, banyak orang dengan diabetes tipe 2 mungkin tidak memiliki gejala apa pun selama bertahun-tahun, dan penyakit ini mungkin terdeteksi hanya selama pemeriksaan medis rutin. DM dapat menyebabkan komplikasi serius jika tidak diobati, termasuk penyakit kardiovaskular, penyakit ginjal, kerusakan saraf, kerusakan mata, dan masalah kaki. Oleh karena itu, penting untuk mendiagnosis dan mengelola DM secara dini untuk mencegah atau menunda komplikasi tersebut.

Mengacu pada [7], Diagnosis DM didasarkan pada pengukuran tingkat glukosa darah. ADA merekomendasikan penggunaan tes hemoglobin A1c (HbA1c), yang mengukur rata-rata tingkat glukosa darah selama dua hingga tiga bulan terakhir. Tingkat HbA1c 6,5% atau lebih tinggi menunjukkan adanya DM. Tes lainnya yang dapat digunakan untuk mendiagnosis DM meliputi tes glukosa plasma puasa (FPG), yang mengukur tingkat glukosa darah setelah puasa semalam, dan tes toleransi glukosa oral (OGTT), yang mengukur tingkat glukosa darah setelah minum minuman manis.

Manajemen DM [8] melibatkan kombinasi perubahan gaya hidup dan pengobatan. Perubahan gaya hidup meliputi mengikuti diet sehat, meningkatkan aktivitas fisik, menurunkan berat badan jika overweight, dan berhenti merokok. Obat-obatan dapat meliputi terapi insulin, obat oral yang meningkatkan sekresi atau sensitivitas insulin, atau obat lain yang membantu menurunkan tingkat glukosa darah.

2.2 Supervised Reinforcement Learning

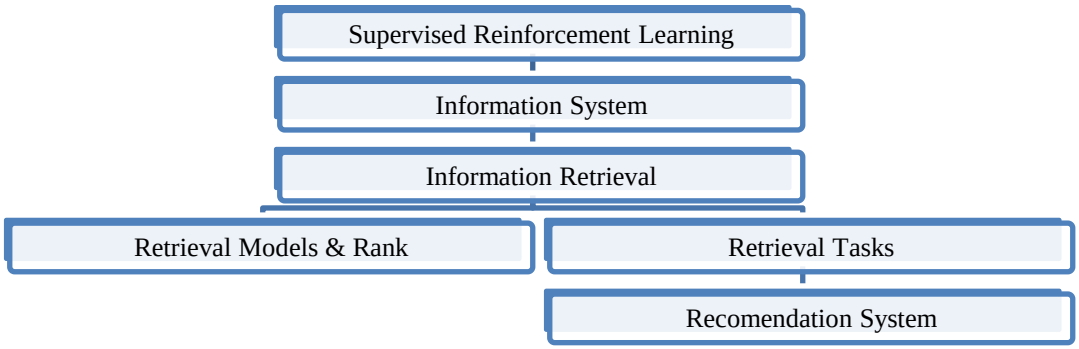
Supervised Reinforcement Learning (SRL) merupakan pendekatan yang menggabungkan elemen-elemen dari dua paradigma utama dalam pembelajaran mesin, yaitu Reinforcement Learning (RL) dan Supervised Learning (SL) [9]. Dalam SRL, agen pembelajaran bertindak dalam lingkungan yang dinamis, seperti dalam RL, namun juga menerima umpan balik yang terarah seperti dalam SL.

Dalam RL konvensional yang tertuang dalam [10], agen belajar RL berinteraksi dengan lingkungan untuk mencapai tujuan tertentu. Agen ini membuat keputusan di setiap langkahnya dan menerima umpan balik dalam bentuk hadiah atau hukuman dari lingkungan sebagai tanggapan terhadap tindakannya. Tujuan RL adalah untuk menemukan strategi yang optimal untuk bertindak dalam lingkungan tertentu guna memaksimalkan total hadiah yang diterima.

Di sisi lain, dalam SL, agen belajar dari contoh-contoh yang diberikan secara eksplisit oleh supervisor [11], [12]. Setiap contoh terdiri dari pasangan input dan output yang diinginkan, dan agen bertugas untuk menemukan hubungan atau pola yang mengaitkan input dengan output. Tujuan SL adalah untuk membuat prediksi atau klasifikasi yang tepat untuk input baru yang diberikan berdasarkan pembelajaran dari contoh-contoh sebelumnya.

Dalam SRL, paradigma RL digabungkan dengan elemen supervisi dari SL. Ini berarti bahwa agen pembelajaran tidak hanya belajar dari umpan balik hadiah atau hukuman yang diberikan oleh lingkungan, tetapi juga dari contoh-contoh yang diberikan oleh supervisor [9]. Dengan kata lain, agen belajar untuk membuat keputusan yang optimal dalam lingkungan yang dinamis berdasarkan informasi langsung dari lingkungan serta informasi tambahan yang diberikan oleh supervisor.

Kelebihan dari SRL [12],[13],[14] adalah kemampuannya untuk memanfaatkan pengetahuan yang tersedia secara eksplisit, yang dapat membantu dalam mempercepat proses pembelajaran dan meningkatkan kinerja agen pembelajaran. Namun, tantangan utama dalam SRL adalah bagaimana mengintegrasikan informasi dari kedua sumber (yaitu umpan balik lingkungan dan contoh-contoh supervisor) dengan baik untuk memastikan bahwa agen pembelajaran dapat belajar dengan efisien dan menghasilkan keputusan yang optimal.



Gambar 1. Skema dasar Supervised Reinforcement Learning [15]

3. METODOLOGI PENELITIAN

3.1. Expert-Supervised Reinforcement Learning (ESRL)

Menurut [9], Expert-Supervised Reinforcement Learning (ESRL) menggunakan kuantifikasi ketidakpastian pada pembelajaran kebijakan secara offline. ESRL merupakan pendekatan yang menjanjikan untuk mempelajari kebijakan yang optimal di lingkungan di mana eksplorasi langsung tidak mungkin dilakukan. ESRL memberikan tiga kontribusi yaitu metode yang dapat mempelajari kebijakan yang aman dan optimal melalui pengujian hipotesis, ESRL memungkinkan penerapan penghindaran risiko pada tingkat yang berbeda-beda yang disesuaikan dengan konteks aplikasi, dan terakhir, ESRL menafsirkan kebijakan ESRL di setiap negara bagian melalui distribusi posterior, dan menggunakan kerangka kerja ini untuk menghitung posterior fungsi nilai di luar kebijakan.

Pendekatan ESRL merupakan pembelajaran offline berbasis Bayesian RL [9]. Metode ini menghasilkan kebijakan yang aman dan optimal karena dapat mempelajari kapan harus mengadopsi perilaku pakar dan kapan harus mengambil tindakan alternatif. ESRL dapat mempelajari kebijakan yang aman dari kumpulan data yang diamati dengan memperhitungkan ketidakpastian menggunakan Markov Decision Process (MDP) dan proses pencatatan data untuk menilai kapan tindakan tersebut aman dan bermanfaat ataupun menyimpang dari kebijakan perilaku. Distribusi MDP pada [9] dimana f (·|DT) secara implisit menentukan distribusi posterior untuk setiap fungsi Q: Qμ̃,t(s,a) ∼ fQ(·|s,a,t,DT). Dimana M∗ adalah stokastik, kita ingin mendekati nilai rata-rata bersyarat Q: EQM∗(s,a)|s,a,t,D. Dimana μ̃,t T pada sampel model MDP K, hitung Q(k)(s, a), k = 1,....,K and use Qˆμ̃,t(s, a) ≡ μ̃,t 1 PK Q(k)(s,a). Pemodelan tersebut dapat dimodelkan menjadi pseudo-algorithm berikut:

Algoritma 1. Pseudo ESRL

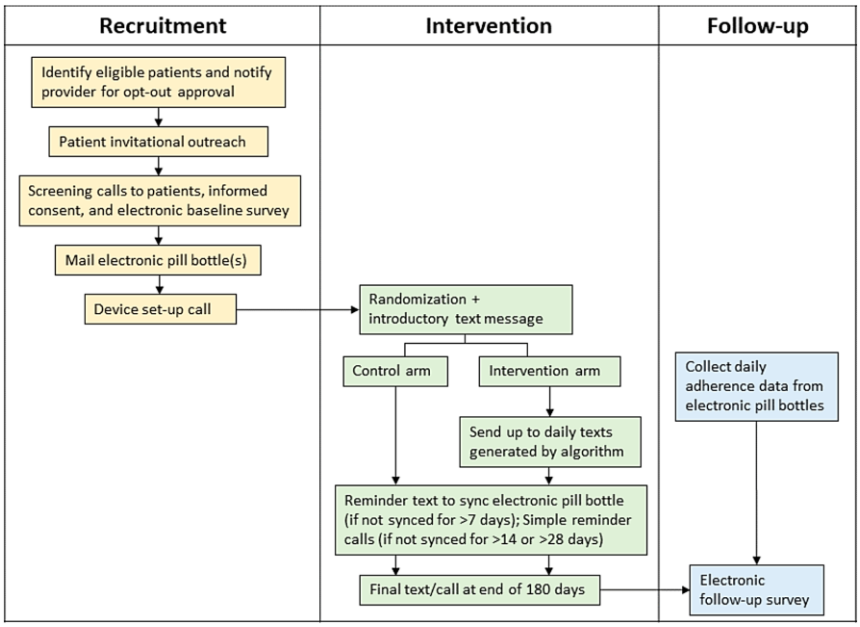
Sample $M_k \sim f(\cdot|\boldsymbol{D}_T)\ k = 1, \ldots, K$, set $\mathcal{I}_1 = \{1, \ldots, \lceil \frac{K}{2} \rceil\}, \mathcal{I}_2 = \{\lceil \frac{K}{2} \rceil + 1, \ldots, K\}$;
Set $\hat{V}_{\tau+1}^{(k)}(s) \leftarrow 0\ \forall s \in \mathcal{S}, k = 1, \ldots, K$;
Compute behavior distribution $\pi(a|s,t)$ from $\boldsymbol{D}_T$, set $\pi(s,t) = \arg\max_a \pi(a|s,t)$;
for $t = \tau, \ldots, 1$ **do**
  for $s \in \mathcal{S}$ **do**
    for $k = 1, \ldots, K$ **do**
      $\mu_k(s,t) \leftarrow \arg\max_a Q_{\mu^\alpha,t}^{(k)}(s,a)$;
    end
    $\hat{\mu}(s,t) \leftarrow$ maj. vote$\{\mu_k(s,t), k \in \mathcal{I}_1\}$;
    Compute $\hat{\mathbb{P}}(H_0|s,t,\boldsymbol{D}_T) = \frac{1}{|\mathcal{I}_2|} \sum_{k \in \mathcal{I}_2} I\left(Q_{\mu^\alpha,t}^{(k)}(s,\hat{\mu}(s,t)) < Q_{\mu^\alpha,t}^{(k)}(s,\pi(s,t))\right)$;
    for $k = 1, \ldots, K$ **do**
      $\mu_k^\alpha(s,t) \leftarrow I\left(\hat{\mathbb{P}}(H_0|s,t,\boldsymbol{D}_T) < \alpha\right)\mu_k(s,t) + I\left(\hat{\mathbb{P}}(H_0|s,t,\boldsymbol{D}_T) \geq \alpha\right)\pi(s,t)$;
      $\hat{V}_t^{(k)}(s) \leftarrow Q_{\mu^\alpha,t}^{(k)}(s,\mu_k^\alpha(s,t))$;
    end
    $\hat{\mu}^\alpha(s,t) \leftarrow$ maj. vote$\{\mu_k^\alpha(s,t), k \in \mathcal{I}_1\}$;
    $\mathcal{M}^\alpha(s,t) \leftarrow \{k | k \in \mathcal{I}_1, \mu_k^\alpha(s,t) = \hat{\mu}^\alpha(s,t)\}$;
  end
end
Define majority voting set: $\text{MV}^\alpha = \cap_{(s,t)} \mathcal{M}^\alpha(s,t)$;
if $\exists k \in MV^\alpha$ **then**
  choose $k \in \text{MV}^\alpha$ at random, set $k^{\text{MV}} \leftarrow k$
else
  Set k^{MV} to most common $k \in \mathcal{M}^\alpha(s,t), \forall(s,t)$
end
Set $\mu^\alpha = \mu_{k^{MV}}$

Asumsi: dimana P∗(H0|s,t,DT) didefinisikan dalam baris pertama untuk M∗*yang bernilai benar*. Parameter α yang dipilih adalah risiko-penolakan dan memenuhi α ∈ [0,1] *satisfies* P∗(H0|s,t,DT)≠ α ∀(s,t) ∈ S × {1,...,τ}. Karena α ditetapkan oleh pengguna, Pseudo alhorithm dengan mudah terpenuhi selama α diseleksi dengan hati-hati. M∗ dan Vμα∗,1(s) menjadi nilai di bawah MDP dengan M bernilai benar dan μ menjadi kebijakan ESRL yang menggunakan hipotesis nol yang ditentukan di bawah M∗. Maka untuk i, kita dapat mendefinisikan penolakan μα P M∗ M∗ i sebagai E[Regret(T)] = E.

Berdasarkan algoritma diatas, ESRL dapat mempelajari kebijakan yang paling aman dari beberapa simulasi. ESRL memperhitungkan ketidakpastian MDP dan menrekam berbagai data untuk memberikan keputusan yang paling sesuai dan aman bagi pasien. ESRL memungkinkan untuk menguraingi berbagai risiko yang dipilih dalam penerapan pengobatan medis. ESRL juga dapat digunakan untuk memperoleh distribusi posterior yang dapat diinterpretasikan untuk fungsi Q. Posterior ini fleksibel untuk memperhitungkan fungsi kebijakan apa pun yang mungkin dan dapat diinterpretasikan dalam konteks aplikasi [9].

3.2. REINFORCE

Guna mencapai pengendalian diabetes yang optimal maka diperlukan beberapa perilaku manajemen diri sehari-hari, terutama kepatuhan pada pengobatan. Penggunaan pesan teks dapat menunjang kepatuhan seorang pasien untuk melaksanakan manajemen diri [16], [17], namun sesungguhnya masih banyak peluang yang dapat dilakukan untuk meningkatkan efektivitas manajemen diri melalui metode REINFORCE.



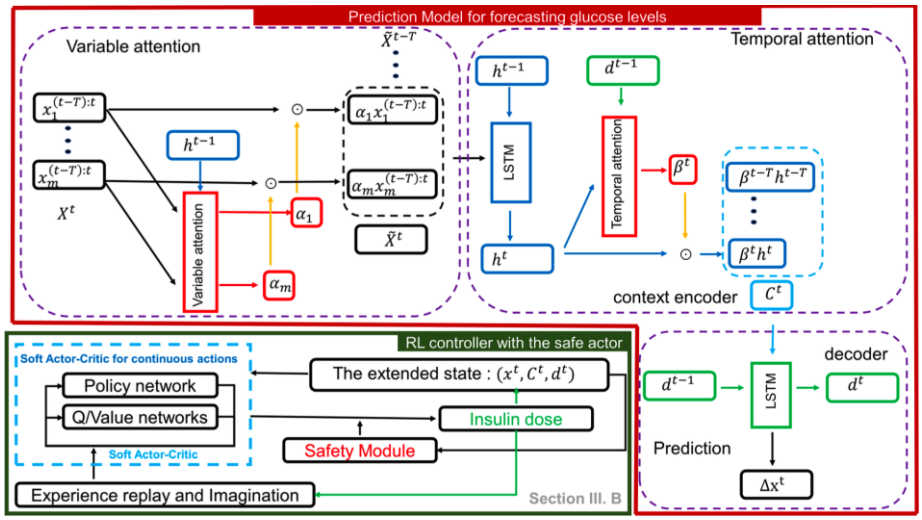
Gambar 2. Diagram prosedur REINFORCE [16]

Metode REINFORCE [16] pada diagram diatas dibangun untuk meningkatkan kepatuhan terhadap pengobatan diabetes dengan mengoptimalkan respons dan menyesuaikan keterlibatan diri, yang merupakan uji coba acak pragmatis terhadap pasien diabetes terkontrol suboptimal untuk menguji strategi personalisasi komunikasi menggunakan konsep RL guna meningkatkan kepatuhan terhadap pengobatan dan kontrol diri pasien diabetes. Peneliti melakukan randomisasi terhadap 60 pasien dengan kontrol diabetes suboptimal yang diobati dengan obat diabetes oral untuk menerima intervensi pembelajaran penguatan atau kontrol. Subjek di kedua kelompok akan menerima botol pil elektronik untuk digunakan, dan mereka di kelompok intervensi akan menerima pesan teks hingga setiap hari. Pesan-pesan akan disesuaikan secara individual menggunakan algoritma prediksi pembelajaran penguatan berdasarkan pengukuran kepatuhan harian dari botol pil.

Secara keseluruhan, uji klinis REINFORCE akan mengevaluasi efek personalisasi penyajian pesan teks untuk pasien untuk mendukung kepatuhan pengobatan dan memberikan wawasan tentang bagaimana hal ini dapat disesuaikan secara luas untuk meningkatkan intervensi manajemen diri lainnya dengan mengoptimalkan respons dan keterlibatan pasien.

3.3. Reinforcement Learning with Safety and Interpretability (RLSI)

Berdasarkan [18], RLSI merupakan kerangka kerja untuk memperkirakan dan mengendalikan glukosa darah, yang dapat diadopsi dengan aman di lingkungan klinis, dan memberikan interpretasi perilaku model intervensi sebelumnya. RLSI mengimplementasikan algoritma SAC, random forest regressor dan dual attention network diterapkan untuk prediksi kadar glukosa dan perluasan variabel keadaan pasien. Jaringan aktor-kritik dapat menentukan dosis insulin berdasarkan kontrol Proporsional-Integral-Derivatif (PID) [19]. Kemudian simulator menggunakan FDA untuk memvalidasi algoritma, dan kinerja algoritma SAC untuk pengaturan kadar glukosa darah agar sebanding dengan kontrol PID. RLSI dapat mengeksploitasi data sebelumnya tentang dinamika fisiologis internal karena fleksibilitas dalam mencerminkan dinamika perubahan waktu dan meminimalkan kesenjangan kinerja antara simulasi dan lingkungan sebenarnya selama pengujian. Untuk mengimbangi minimnya data, kontrol PID memandu pelatihan SAC, dan aktor adaptif dalam memodulasi dosis insulin. Perluasan pendekatan ini efektif dalam menangkap hubungan fisiologis antar variabel berdasarkan data, dan korelasi dengan keadaan dasar yang memberikan informasi dinamika internal, ditunjukkan pada gambar 3.



Gambar 3. Model prediksi glukosa RLSI [18]

Gambar diatas merupakan sttruktur rinci dari model prediksi dan kontrol ramalan dan regulasi glukosa darah. Skor perhatian untuk variabel dan hubungan temporal diperoleh melalui encoder dan decoder. Keadaan model prediksi diperluas dengan variabel tersembunyi dan vektor konteks dari jaringan dual-attensi untuk memprediksi perubahan dalam keadaan. SAC dengan aktor aman adaptif dalam menghitung dosis insulin menggunakan keadaan yang diperluas untuk pembelajaran dan tindakan [18].

3.4. Double Deep Q Network (Double DQN)

Double DQN pada [19], [20] merupakan turunan berbasis nilai sederhana RL yang digunakan untuk mengatasi beberapa kelemahan dalam suatu jaringan, seperti masalah perkiraan yang berlebihan. Double DQN memiliki dua artsitektur jaringan saraf identik, yaitu jaringan evaluasi dan jaringan target. Jaringan evaluasi merupakan jaringan utama yang akan digunakan untuk memperoleh pengobatan yang optimal setelah pelatihan. Jaringan target digunakan untuk memperkirakan nilai keluaran target dalam menghitung fungsi kerugian. Jaringan saraf menggunakan vektor keadaan/kondisi pasien sebagai masukan dan vektor Q menghasilkan tindakan sebagai keluarannya, dimana tindakan dapat meliputi terapi, penggunaan obat, dll. Vektor Q juga dipakai dalam mengevaluasi keseluruhan efek tindakan pengobatan terhadap status kondisi pasien [19]. Sebagai hasil, Double DQN tidak hanya dapat mengendalikan kadar HbA1c, namun juga berhasil mengendalikan glukosa jangka panjang dengan lebih baik setelah satu tahun penerapan. Hasil percobaan ini juga menunjukkan bahwa metode Double DQN mampu mempelajari pola pengobatan yang baik dari dokter sehingga mampu menunjang pengendalian glikemik jangka panjang secara efektif.

Masalah kontrol loop tertutup glukosa basal dalam diabetes dapat dirumuskan sebagai proses keputusan infinite-state Markov dengan noise [20], yang didefinisikan oleh sebuah tuple ⟨S,P,A,R,γ⟩ yang terdiri dari kondisi S (keadaan fisiologis), fungsi transisi keadaan P (model fisiologis), tindakan A (tindakan pengendalian insulin dan glukagon), fungsi reward R (glikemik), dan faktor diskon $\gamma \in [0,1]$ (yaitu, pentingnya hasil prediksi glikemik masa depan). Agen di lingkungan mengambil tindakan $a \in A$ pada setiap langkah waktu (yaitu, setiap pengukuran CGM), dan kemudian keadaannya $s \in S$ berubah menjadi keadaan pengganti s′ sesuai dengan P. Pada [18], kebijakan untuk memilih tindakan untuk keadaan yang diberikan ditunjukkan oleh π. Memaksimalkan akumulasi reward yang diharapkan rt = R(st,at) pada setiap langkah waktu t adalah target dari RL. Sebuah fungsi nilai tindakan (fungsi Q) Qπ (s, a) dapat didefinisikan untuk menghitung reward ditunjukkan pada persamaan 1 dan 2.

$$Q^{\pi}(s,a) = E[\sum_{t'=t}^{\infty} \gamma^{t'-t}\, r_{t'} \,|s_t = s, a_t = a, \pi\,] \tag{1}$$

Selanjutnya fungsi nilai tindakan optimal Q^* (s, a) = max$_\pi$ Q^{π} (s, a) memberikan nilai maksimal, yang dapat ditentukan dengan:

$$Q^*(s,a) = E_{s'}[R(s,a) + \gamma\, max_{a'}\, Q^*(s',a')] \tag{2}$$

Nilai tindakan optimal pada keadaan saat ini s diperoleh dengan memilih tindakan yang memaksimalkan pengembalian yang diharapkan dengan Q* (s', a') pada s′. Meskipun persamaan rekursif ini dapat diestimasi dengan pembaruan iteratif, aproksimator linear dan non-linear umumnya digunakan dalam

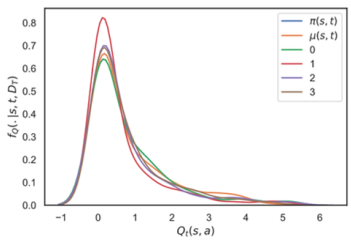
RL untuk generalisasi yang lebih baik. Dalam makalah ini, DQNs digunakan untuk mengaproksimasi nilai tindakan Q(s, a; θ) ≈ Q* (s, a) dimana θ mewakili parameter dari jaringan saraf [19].

4. HASIL PENELITIAN

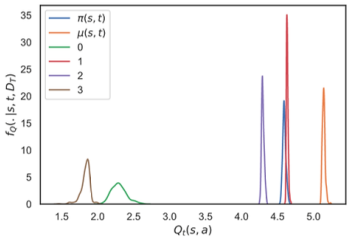
Terapi dan tindakan klinis untuk pasien Diabetes Melitus mencakup serangkaian pendekatan yang bertujuan untuk mengelola kadar glukosa darah, mencegah atau mengurangi komplikasi jangka panjang, dan meningkatkan kualitas hidup pasien. Oleh karena itu penting untuk mempertimbangkan faktor-faktor seperti tujuan spesifik dari pengendalian glukosa darah, kompleksitas dinamika, kebutuhan akan interpretasi, dan lain sebagainya sebelum memilih metodologi mana yang paling sesuai untuk diterapkan dalam lingkungan klinis tertentu.

Semua metodologi diatas bertujuan untuk memberikan pendekatan baru dan praktis terhadap pengendalian glukosa darah dalam pengaturan klinis, menekankan pentingnya interpretasi dan integrasi pengetahuan fisiologis ke dalam algoritma pengendalian. Seluruh metodologi berpotensi memberikan dampak signifikan terhadap pengelolaan kadar glukosa darah pada pasien diabetes dan memiliki kekuatan serta kelemahannya masing-masing.

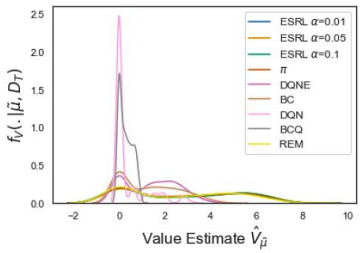
ESRL berfokus pada pembelajaran offline berdasarkan Bayesian RL dengan distribusi posterior yang dapat ditafsirkan dan ditunjukan pada hasil penelitian ditunjukkan pada gambar 4-6 [9].



Gambar 4.
Distribusi posteriur fungsi Q untuk (s, t) = (90, 7), a ∈ {0,1,2,3,π(90,7),μ(90,7)}.



Gamber 5.
Distribusi posteriur fungsi Q untuk (s, t) = (5,8), a ∈ {0,1,2,3,π(5,8),μ(5,8)}.



Gambar 6.
Distribusi posteriur fungsi Q untuk ESRL, BC, BCQ, DQN, DQNE, REM.

Berdasarkan hasil penelitian, maka didapatkan bahwa kelebihan dari ESRL terletak pada kemampuannya untuk memanfaatkan pengetahuan yang tersedia secara eksplisit, yang dapat membantu dalam mempercepat proses pembelajaran dan meningkatkan kinerja agen pembelajaran. Namun, tantangan utama dalam ESRL yakni pada cara mengintegrasikan informasi dari kedua sumber (yaitu umpan balik lingkungan dan contoh-contoh supervisor) dengan baik untuk memastikan bahwa agen pembelajaran dapat belajar dengan efisien dan menghasilkan keputusan yang optimal [9].

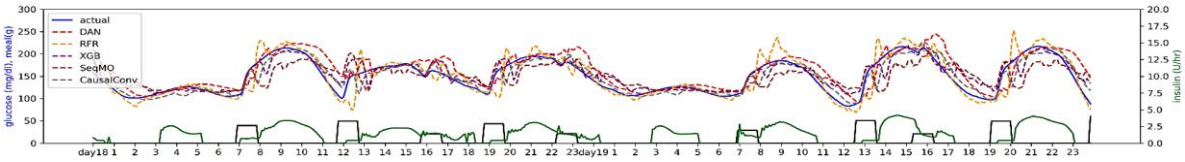
Lain halnya dengan REINFORCE yang merupakan uji coba pragmatis menggunakan pesan teks berbasis pembelajaran penguatan untuk meningkatkan kepatuhan pengobatan. Pengujian REINFORCE [16] dirancang untuk memaksimalkan validitas internal dan generalisasi serta menggunakan data yang dikumpulkan secara rutin untuk mengevaluasi hasil. Dengan menggunakan tindak lanjut selama 6 bulan untuk mengevaluasi hasil kepatuhan, uji coba telah memeriksa penggunaan obat jangka panjang dan hasil klinis pasien DM (misalnya, kontrol glikemik).

Tabel 1. Hasil penerapan REINFORCE [16]

Hasil	Pengukuran	Penilaian
Primer	Kepatuhan pada pengobatan: rata-rata proporsi harian dari masa penyembuhan	Proporsi hari mulai dari pembukaan obat elektronik dalam periode tindak lanjut 6 bulan, dirata-ratakan untuk seluruh pengobatan diabetes yang diteliti
Sekunder	Pengawalan glikemik: Tindak lanjut HbA1c	Nilai yang terdekat dengan periode pengobatan selama 6 bulan, ditambah dengan nilai laboratorium rekam medis elektronik
Sekunder	Pelaporan kepatuhan mandiri	Pemantauan selama 12 bulan, menggunakan skor kepatuhan yang divalidasi oleh Wilson et al [20]
Sekunder	Pengawalan glikemik: Perubahan tindak lanjut HbA1c dari basis tindak lanjut pertama	Perubahan antara basis awal dan kunjungan 6 bulan, dalam nilai yang dikumpulkan oleh alat rekam medis elektronik

Meskipun metode obat elektronik pada REINFORCE sangat akurat dalam mengukur konsumsi pil yang sebenarnya, pemantauan secara teoritis ini dapat mempengaruhi kepatuhan, namun efek pengamat ini biasanya dapat menurun seiring berjalannya waktu dan serupa pada kelompok kontrol dan intervensi.

RLSI efektif mengatur kadar glukosa darah dengan meningkatkan manajemen diri pasien. RLSI dapat meramalkan dan mengendalikan glukosa darah, yang dapat diadopsi dengan aman di lingkungan klinis, dan memberikan interpretasi perilaku model untuk intervensi sebelumnya. Sebuah simulator yang disetujui oleh FDA digunakan untuk memvalidasi algoritma, dan kinerja algoritma SAC untuk regulasi kadar glukosa darah sebanding dengan kontrol PID [18]. Model-model tersebut memanfaatkan data sebelumnya tentang dinamika fisiologis internal sekecil mungkin karena fleksibilitas dalam mencerminkan dinamika yang bervariasi seiring waktu dan meminimalkan kesenjangan kinerja antara simulasi dan lingkungan aktual selama pengujian glukosa. Untuk mengkompensasi data yang minimal, kontrol PID memandu pelatihan SAC, dan aktor aman adaptif mengatur dosis insulin. Berikut ini kurva prediksi glukosa dan insulin pada pasien, ditunjukkan pada gambar 7.



Gambar 7. Kurva Prediksi Glukosa dan Insulin Pada Pasien

Kurva diatas meramalkan kadar glukosa darah subjek remaja selama dua hari terakhir. Setiap model dimodifikasi untuk dilatih dalam pembelajaran online tanpa pengetahuan sebelumnya tentang dinamika glukosa. Nilai-nilai yang diukur dari glukosa, insulin, dan asupan makanan direpresentasikan dalam kurva biru, hijau, dan hitam. Garis putus-putus mewakili nilai prediksi kadar glukosa setelah 30 menit, yang dibandingkan dengan nilai-nilai yang diukur [18].

Double DQN menggunakan dua buah jaringan syaraf identik agar dapat mempelajari pola pengobatan pakar kesehatan dengan aman dan dapat mengendalikan glikemik jangka panjang. Hasil pengujian disajikan dalam dua buah tabel 1 dan 2 [19].

Tabel 1. Pengujian performa penekanan kadar gula darah pada orang dewasa

Metode	TIR (%)	Hypo (%)	Hyper (%)	Rata – Rata (mg/dL)	RI
LGS	77.55±6.78	2.87±1.38	19.58 ± 5.79	140.78±8.23	2.52±0.89
DRL-SH	80.94 ± 7.00*	2.06±1.33*	17.00 ± 5.82	140.36±5.98	2.28±0.72
DRL-DH	85.55±7.33**,†	1.92±1.90*	13.81 ± 6.67**,†	140.12±8.13	2.16±0.65†

Tabel 2. Pengujian performa penekanan kadar gula darah pada remaja

Metode	TIR (%)	Hypo (%)	Hyper (%)	Rata – Rata (mg/dL)	RI
LGS	55.50±14.68	6.93±4.69	37.57±11.64	162.15±20.46	4.76±2.70
DRL-SH	65.85±16.30**	5.51±3.37	28.63±14.36**	151.18±18.26**	3.99±2.43**
DRL-DH	78.83±6.60**,†	2.64±1.96**,‡	18.53±6.48**,†	149.96±8.83**	2.94±0.99**,‡

Keterangan: Simbol ∗ menunjukkan signifikansi ($p \leq 0,05$) untuk suspensi glukosa rendah (LGS) dan † menunjukkan signifikansi ($p \leq 0,05$) untuk single-hormon DRL (DRL-SH). Simbol ganda (misalnya, ‡) menunjukkan signifikansi statistik ($p \leq 0,01$). Keterangan ini berlaku untuk tabel 1 dan tabel 2 [19].

5. KESIMPULAN

Kesimpulannya, penelitian ini menyoroti potensi beberapa metode SRL dalam meningkatkan kepatuhan pasien terhadap pengobatan diabetes dan juga menyediakan pemahaman lebih lanjut tentang performa beberapa metodologi dalam hal memberikan saran pengobatan terbaik bagi pasien. Namun, setiap metode memiliki tantangan dan keterbatasannya sendiri yang perlu diatasi. Penelitian selanjutnya diharapkan peneliti yang lain dapat mengombinasi beberapa metode ini agar dapat memberikan solusi yang lebih baik, komprehensif dan efektif dalam mengelola kadar glukosa darah pada pasien diabetes.

UCAPAN TERIMAKASIH

Terima kasih kepada dr. Fusia Meidiawaty dan dr. Anisa Syafitri untuk diskusinya mengenai diabetes melitus dan penanganan klinisnya. Penelitian ini dapat digunakan sebagai bahan pengajaran pada mata kuliah sistem pengambilan keputusan dan Teknik Riset Operasional. Penelitian ini juga merupakan bagian

REFERENSI

[1] R. Kumar, P. Saha, Y. Kumar, S. Sahana, A. Dubey, And O. Prakash, "A Review On Diabetes Mellitus: Type 1 & Type 2," World Journal of Pharmacy and Pharmaceutical Sciences, Vol. 9, No. 10, Pp. 838-850, Oct. 2020. Issn 2278 - 4357.

[2] M. Tejedor, A. Z. Woldaregay, and F. Godtliebsen, "Reinforcement learning application in diabetes blood glucose control: A systematic review," Artificial Intelligence in Medicine, vol. 104, p. 101836, Apr. 2020.

[3] F. Zohora, M. H. Tania, M. S. Kaiser, and M. Mahmud, "Forecasting the Risk of Type II Diabetes using Reinforcement Learning," presented at the 2020 Joint 9th International Conference on Informatics, Electronics & Vision (ICIEV) and 2020 4th International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Kitakyushu, Japan, Aug. 26-29, 2020, doi: 10.1109/ICIEVicIVPR48672.2020.9306653.

[4] C. Shiranthika, K.-W. Chen, C.-Y. Wang, C.-Y. Yang, B. H. Sudantha, and W.-F. Li, "Supervised Optimal Chemotherapy Regimen Based on Offline Reinforcement Learning," IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 9, pp. 4763-4772, Sep. 2022, doi: 10.1109/JBHI.2022.3183854.

[5] S. Neaz, N. Hussain, M. F. Hossain, and T. Rahman, "Diabetes Mellitus: Insights from Epidemiology, Biochemistry, Risk Factors, Diagnosis, Complications and Comprehensive Management," Diabetology 2021, vol. 2, pp. 36–50, 2021. MDPI Journal. DOI: 10.3390/diabetology2020004.

[6] Y. L. Tinungki and J. S. H. Hinonaung, "Deteksi Dini Penyakit Diabetes Mellitus (DM) dan Obat Tradisional DM Pada Lansia di Kabupaten Kepulauan Sangihe," PT. Sonpedia Publishing Indonesia, Nov. 22, 2023.

[7] L. Cloete, "Diabetes mellitus: an overview of the types, symptoms, complications and management," Nursing Standard (Royal College of Nursing (Great Britain) : 1987), vol. 37, no. 1, pp. 61-66, Jan. 2022. DOI: 10.7748/ns.2021.e11709. PMID: 34708622.

[8] C.S. Lau and T.C. Aw, "HbA1c in The Diagnosis and Management of Diabetes Mellitus: An Update," Diabetes, vol. 6, pp. 1-4, 2020. DOI: 10.15761/DU.1000137

[9] A. Sonabend-W, J. Lu, L. A. Celi, T. Cai, and P. Szolovits, "Expert-Supervised Reinforcement Learning for Offline Policy Learning and Evaluation," in Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada, 2020.

[10] R. S. Sutton and A. G. Barto, "Reinforcement Learning: An Introduction," 2nd ed. MIT Press, Oct. 19, 2018. ISBN: 9780262352703, 0262352702

[11] A. S. Chauhan, M. S. Varre, K. Izuora, M. B. Trabia, and J. S. Dufek, "Prediction of Diabetes Mellitus Progression Using Supervised Machine Learning," Sensors, vol. 23, no. 10, p. 4658, May 11, 2023. DOI: 10.3390/s23104658.

[12] D. Datta, M. Bhattacharya, S. S. Rajest, T. Shynu, R. Regin, and S. S. Priscila, "Development of Predictive Model of Diabetic Using Supervised Machine Learning Classification Algorithm of Ensemble Voting," International Journal of Bioinformatics Research and Applications, vol. 19, no. 3, pp. 151-169, Sep. 28, 2023. DOI: 10.1504/IJBRA.2023.133695.

[13] B. Chen, C. Zhu, P. Agrawal, K. Zhang, and A. Gupta, "Self-Supervised Reinforcement Learning that Transfers using Random Features," presented at the 37th Conference on Neural Information Processing Systems (NeurIPS 2023), Advances in Neural Information Processing Systems 36 (NeurIPS 2023) Main Conference, May 26, 2023. DOI: 10.48550/arXiv.2305.17250.

[14] T. Zhou, T. Guo, C. Dang, and M. Beer, "Bayesian reinforcement learning reliability analysis," Computer Methods in Applied Mechanics and Engineering, vol. 424, p. 116902, May 1, 2024. DOI: 10.1016/j.cma.2024.116902. Published by Elsevier.

[15] X. Xin, A. Karatzoglou, I. Arapakis, and J. M. Jose, "Self-Supervised Reinforcement Learning for Recommender Systems," in SIGIR '20: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, July 2020, pp. 931-940, doi: 10.1145/3397271.3401147.

[16] J. C. Lauffenburger, E. Yom-Tov, P. A. Keller, M. E. McDonnell, L. G. Bessette, C. P. Fontanet, E. S. Sears, E. Kim, K. Hanken, J. J. Buckley, R. A. Barlev, N. Haff, and N. K. Choudhry, "REinforcement Learning to Improve Non-Adherence for Diabetes Treatments by Optimising Response and Customising Engagement (REINFORCE): study protocol of a pragmatic randomised trial," BMJ Open, vol. 11, e052091, Nov. 11, 2021. DOI:10.1136/bmjopen-2021-052091.

[17] J. He, H. Zhao, D. Zhou, and Q. Gu, "Nearly Minimax Optimal Reinforcement Learning for Linear Markov Decision Processes," in Proceedings of the 40th International Conference on Machine Learning, PMLR 202, pp. 12790-12822, 2023.

[18] M. H. Lim, W. H. Lee, B. Jeon, and S. Kim, "A Blood Glucose Control Framework Based on Reinforcement Learning With Safety and Interpretability: In Silico Validation," IEEE Access, vol. 9, pp. 113334-113347, Jul. 26, 2021. DOI: 10.1109/ACCESS.2021.3100007.

[19] T. Zhu and P. Herrero, "Basal Glucose Control in Type 1 Diabetes using Deep Reinforcement Learning: An In Silico Validation," IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 9, pp. 2168-2194, 2020, doi: 10.1109/JBHI.2020.3014556.

[20] J. Jumiyatun, I. Mahmudi, and A. Mustari, "Kontrol Power Elektronik Dan Aplikasinya: Perancangan, Pemodelan, Simulasi dan Implementasi," Media Nusa Creative (MNC Publishing), Nov. 15, 2021.
[21] Z. Liu, W. Zhao, S. Liu, M. Feng, L. Ji, X. Liao, X. Sun, G. Xie, X. Jiang, T. Zhao, and G. Hu, "A Deep Fment Learning Approach for Type 2 Diabetes Mellitus Treatment," in 2020 IEEE International Conference on Healthcare Informatics (ICHI), 2020, DOI: 10.1109/ICHI48887.2020.9374313.