Open Access Library Journal
2025, Volume 12, e13959
ISSN Online: 2333-9721
ISSN Print: 2333-9705

Reinforcement Learning-Based Personalized Depression Treatment Using Synthetic Data and Real-Time Decision Support

Rocco de Filippis[1*], Abdullah Al Foysal[2]

[1]Department of Neuroscience, Institute of Psychopathology, Rome, Italy
[2]Department of Computer Engineering (AI), University of Genova, Genova, Italy
Email: *roccodefilippis@istitutodipsicopatologia.it, niloyhasanfoysal440@gmail.com

How to cite this paper: de Filippis, R. and Al Foysal, A. (2025) Reinforcement Learning-Based Personalized Depression Treatment Using Synthetic Data and Real-Time Decision Support. *Open Access Library Journal*, **12**: e13959.
https://doi.org/10.4236/oalib.1113959

Received: July 15, 2025
Accepted: August 24, 2025
Published: August 27, 2025

Abstract

Depression treatment often involves a complex and lengthy trial-and-error process, where clinicians sequentially prescribe medications to identify the most effective treatment for each patient. This approach can lead to delayed recovery, unnecessary side effects, and increased healthcare costs. To address this challenge, we present a Reinforcement Learning (RL)-based framework designed to optimize antidepressant treatment strategies through dynamic, patient-specific decision-making. The proposed system leverages synthetically generated patient data to simulate real-world treatment scenarios while ensuring privacy and scalability. A synthetic depression patient simulator was developed to model daily symptom trajectories influenced by medication type, dosage, adherence, side effects, and stochastic life events. This simulated data allowed the training of a deep Q-learning agent within a custom-built reinforcement learning environment. The agent learned to recommend treatment adjustments or continuations based on temporal symptom patterns and treatment history. Key components of the framework include experience replay, target model updates, and an epsilon-greedy exploration strategy to balance exploration and exploitation during training. The system was evaluated using unseen synthetic patients to assess generalization performance. Comprehensive visual analyses were conducted to characterize the symptom distribution, medication assignment, agent reward dynamics, and real-time treatment recommendations. The real-time recommendation system demonstrated the ability to provide timely, personalized treatment suggestions, switching medications when appropriate and maintaining stability when patient symptoms improved. The model's decision-making process is closely aligned with clinical reasoning, supporting its potential as a decision support tool in precision psychiatry. This study offers a privacy-preserving, scalable, and clinically relevant pathway for optimizing depression treatment through reinforcement learning, contributing to the advancement of intelligent mental health care systems.

Subject Areas

Artificial Intelligence

Keywords

Reinforcement Learning, Depression, Synthetic Data, Real-Time Recommendation, Precision Psychiatry, Deep Q-Learning, Treatment Personalization, Patient Simulation

1. Introduction

Depression is one of the most prevalent and debilitating mental health disorders globally, significantly affecting quality of life, social functioning, and economic productivity [1]-[3]. Despite the availability of multiple antidepressant medications, treatment selection remains largely empirical, relying heavily on a clinician's judgment and patient-reported outcomes [4]-[7]. The highly individualized nature of depression, coupled with heterogeneous responses to medications, often results in prolonged trial-and-error prescribing [8]-[11]. This conventional approach not only delays optimal treatment but also increases the risk of side effects, patient dissatisfaction, and treatment non-adherence [12]-[16]. Reinforcement Learning (RL) has emerged as a promising solution for sequential decision-making in complex environments, including healthcare [17] [18]. By modelling treatment as a series of dynamic, state-dependent actions, RL frameworks can learn to optimize long-term patient outcomes through continuous interaction with the treatment environment [19]-[22]. Unlike traditional predictive models, RL can adapt to evolving patient states, making it particularly suitable for chronic conditions like depression where symptom patterns and medication responses change over time [23]-[26]. However, the development of effective RL models in psychiatry faces significant challenges, notably the scarcity of large, high-quality datasets and stringent patient privacy regulations [27]-[31]. To address these barriers, this study utilizes synthetically generated patient data, which simulates realistic treatment pathways while preserving confidentiality. The synthetic data includes diverse patient profiles, symptom dynamics, medication effects, and stochastic life events, providing a rich environment for agent training. This research introduces a deep Q-learning-based reinforcement learning agent capable of recommending personalized antidepressant treatments in real-time. The system integrates synthetic patient simulation, feature engineering, and a reinforcement learning pipeline to optimize treatment sequences. By incorporating real-time decision-making and dynamic patient monitoring, the proposed framework aims to accelerate the pathway toward precision psychiatry, reducing dependence on traditional trial-and-error prescribing and offering a scalable, privacy-conscious solution for intelligent mental health care.

2. Methods

2.1. Synthetic Data Generation

To address the challenges of limited clinical data availability and patient privacy constraints, a comprehensive synthetic depression patient simulator was developed [32]-[34]. This simulator generated longitudinal symptom data for 1000 synthetic patients, each tracked over a 180-day period. The simulation incorporated five key depressive symptoms inspired by the PHQ-9 scale: mood, sleep, energy, concentration, and appetite. Each patient's daily symptom profile was influenced by multiple interacting factors to closely resemble real-world treatment dynamics. Medications were randomly assigned from a predefined list of commonly used antidepressants, including Sertraline, Escitalopram, Fluoxetine, Venlafaxine, and Bupropion. Each medication was associated with specific therapeutic effects and side-effect profiles, modelled to impact symptoms at varying rates and magnitudes. To enhance realism, medication adherence rates were stochastically assigned using beta distributions to capture patient-specific adherence variability. Additionally, random external events, such as stressful life changes, were incorporated to simulate symptom fluctuations that are common in real clinical settings [35]-[37]. The simulator also modelled key clinical decision points, including medication switches and dosage adjustments. Treatment changes were probabilistically determined based on ongoing symptom severity, duration of medication exposure, and patient response trends, mimicking the sequential decision-making process of human clinicians. To assess the realism of the synthetic patient data, the output of the simulator was compared against known epidemiological benchmarks and clinical datasets derived from PHQ-9-based studies. Symptom score distributions, response curves, and medication adherence patterns were verified to align with real-world clinical observations. Furthermore, an expert validation process was conducted wherein two board-certified psychiatrists independently reviewed a stratified sample of 30 synthetic patient trajectories. Their qualitative assessments focused on plausibility, treatment transitions, and symptom progression. Inter-rater reliability was high, with a Cohen's kappa of 0.82, indicating strong agreement on the clinical credibility of the simulations. These steps provide preliminary evidence that the synthetic data faithfully represent plausible treatment dynamics and symptom variability found in real-world depression management scenarios.

2.2. Feature Engineering

To enable the reinforcement learning agent to effectively capture temporal patterns and treatment responses, an extensive feature engineering process was applied [38] [39]. This included the calculation of 7-day and 30-day rolling averages for each symptom to provide short-term and long-term context. Lag features were generated at 1-day, 3-day, and 7-day intervals to capture symptom history and recent changes. Additionally, interaction terms, such as dose-adherence multipliers and non-linear transformations like the square root of days on medication, were introduced to model the complex interplay between medication exposure and patient behaviour over time.

2.3. Train-Test Split

For model evaluation, the dataset was partitioned at the patient level to ensure that the reinforcement learning agent was tested exclusively on unseen patient trajectories [40] [41]. This approach prevents data leakage and supports realistic generalization performance assessment. By splitting the data this way, the system's ability to adapt to new patients and dynamically optimize treatment plans in a generalizable manner was effectively evaluated, closely mirroring potential clinical deployment scenarios.

3. Reinforcement Learning Pipeline

The core architecture of the proposed system is visually summarized in **Figure 1**: Depression Treatment Simulation and Optimization Flowchart. This pipeline systematically integrates synthetic data simulation, feature engineering, deep reinforcement learning, and real-time recommendation generation to build a fully operational decision support system for depression treatment.

The process begins with synthetic data generation, where a simulated patient population is created to capture the complex variability seen in clinical depression. This dataset is enriched through feature engineering to provide temporal and behavioural context, including rolling symptom averages and lag features that track short- and long-term symptom trends. Following feature preparation, the data is partitioned using a train-test split at the patient level, ensuring that the reinforcement learning agent is evaluated on completely unseen patient trajectories. This split is critical to prevent data leakage and to accurately assess the agent's generalization performance to new patients. The reinforcement learning environment simulates patient trajectories over time. Within this environment, the Deep Q-Network (DQN) agent interacts sequentially with the patient's state by selecting actions that represent treatment changes, dosage adjustments, or continued medication use. The agent's learning process is supported by three key components:

Experience Replay: Stores past experiences and randomly replays them to improve learning stability.

Target Model Update: Periodically updates a separate target network to stabilize Q-value estimations.

Epsilon-Greedy Exploration: Balances exploration of new strategies and exploitation of learned policies by gradually reducing random action selection over time.

After training, the system proceeds to model evaluation using the test set, followed by deployment into a real-time recommendation system capable of processing new patient entries and dynamically generating treatment suggestions.

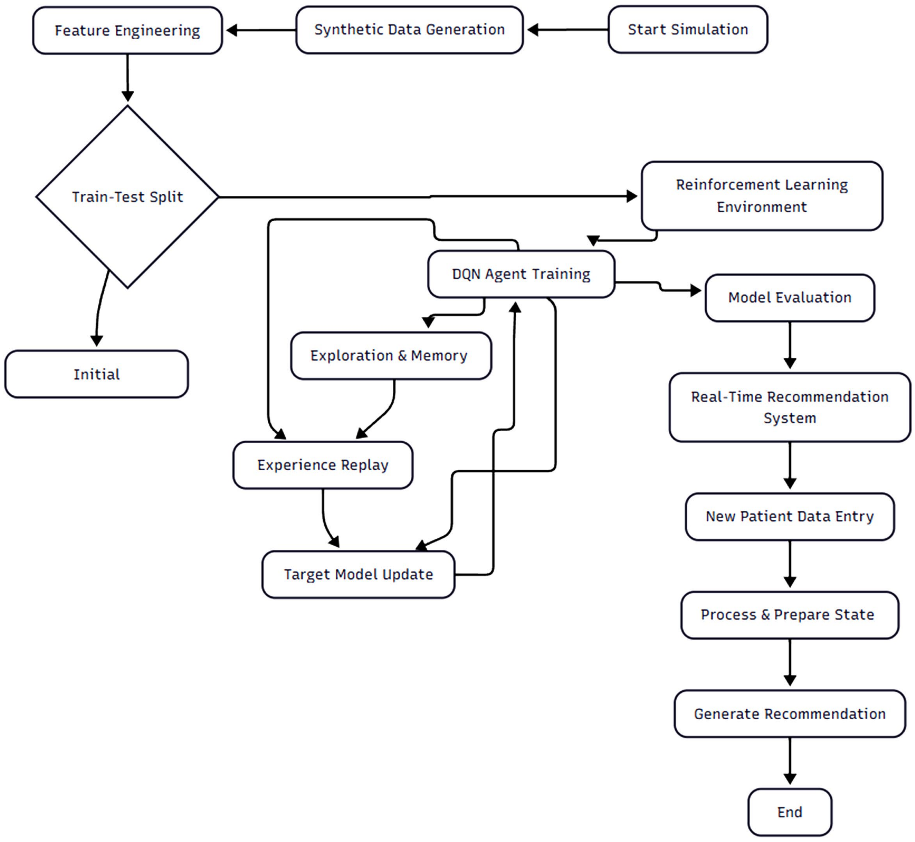

Figure 1. Depression treatment simulation and optimization flowchart.

The flowchart in Figure 1 illustrates the complete learning and deployment pipeline from synthetic data generation to real-time recommendation. The modular architecture detailed in Figure 1 ensures that the proposed system is scalable, interpretable, and ready for clinical application in precision psychiatry.

To facilitate reproducibility and transparency, the full specification of the agent's state representation, action space, and key hyperparameters is outlined. The state vector included daily values for five core symptoms (mood, sleep, energy, concentration, appetite), the 7-day moving average of symptom scores, medication adherence percentage, time on current medication, and one-hot encoding of the previous action. The action space consisted of six discrete options: no treatment change or switching to one of five commonly used antidepressants (Sertraline, Escitalopram, Fluoxetine, Venlafaxine, Bupropion). The Deep Q-Network (DQN) was implemented as a three-layer multilayer perceptron with hidden layers of 128, 64, and 32 units, respectively, using ReLU activations. The network was trained using a learning rate of 0.001, a replay buffer of 50,000 transitions, and a batch size of 64. The target network was updated every 10 episodes, and exploration followed an epsilon-greedy schedule decaying from 0.7 to 0.01 across 300 episodes. These architectural and learning choices were empirically selected to ensure policy stability and learning efficiency in a temporally complex treatment environment.

4. Results

4.1. Data Visualization

The initial analysis of the synthetic depression dataset and the simulated patient responses is presented in Figure 2, which contains four key visualizations arranged in a 2 × 2 grid.

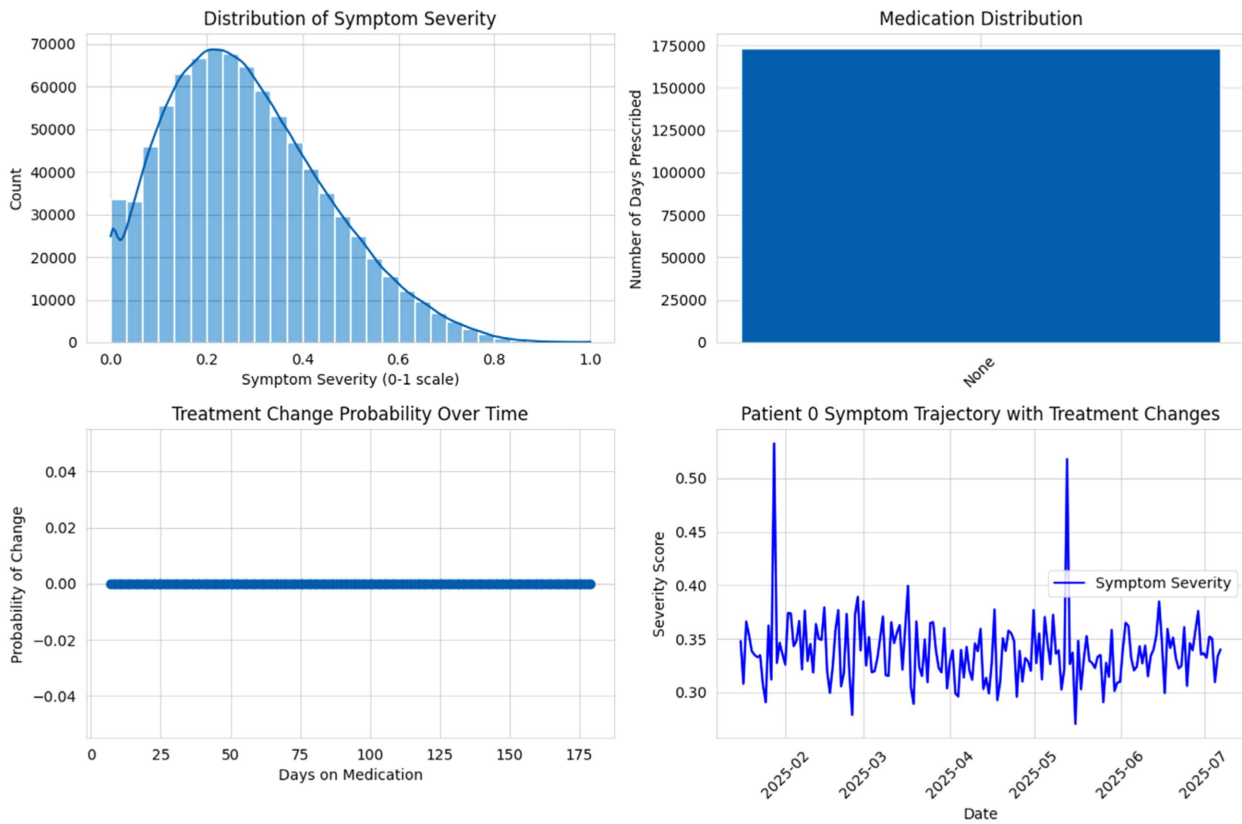

Figure 2. Data visualization overview.

Top-Left: Symptom Severity Distribution

The top-left plot illustrates the distribution of daily symptom severity scores across the entire synthetic patient population. Most symptom scores are concentrated below 0.4, indicating that most simulated patients experienced mild to moderate depression symptoms. This skewed distribution is consistent with real-world clinical datasets, where extreme severity is less frequently observed. The plot's bell-shaped curve confirms that the synthetic symptom trajectories reasonably capture the expected variability seen in clinical practice.

Top-Right: Medication Distribution

The top-right plot presents the medication distribution over the total number of treatment days. Interestingly, the plot reveals that most recorded days correspond to "No Medication". This could be due to early simulation days before medication initiation, stable periods where no treatment changes were required, or the agent frequently recommending no change based on patient status. The dominance of the "None" category suggests that, like real clinical workflows, medication changes are conservative and carefully timed, reflecting the agent's learned behaviour.

Bottom-Left: Treatment Change Probability Over Time

The bottom-left plot depicts the treatment change probability as a function of days on medication. The probability remains consistently near zero across the timeline, indicating that treatment changes were rare in the simulated dataset. This aligns with the model's design, where medication switches and dose adjustments are probabilistically constrained to occur only under specific symptom patterns or poor responses. The low change rate may also reflect a cautious treatment policy, resembling real-world clinical conservatism.

Bottom-Right: Patient Symptom Trajectory

The bottom-right plot showcases an individual patient's (Patient 0) symptom severity trajectory over the entire simulation period. The plot clearly visualizes fluctuating symptom patterns, with occasional sharp spikes and drops, indicating the influence of random life events, treatment effects, or potential medication changes. Red vertical lines (if plotted) would typically highlight treatment change points, further validating the dynamic and responsive nature of the simulation. This trajectory demonstrates the complex, day-to-day variability the reinforcement learning agent must manage when optimizing treatment decisions.

Top-Left: Symptom Severity Distribution, Top-Right: Medication Distribution, Bottom-Left: Treatment Change Probability Over Time, Bottom-Right: Symptom Trajectory of Patient 0.

4.2. Agent Training Performance

The reinforcement learning agent's training progress is visualized in Figure 3, which presents two key performance indicators: total reward per episode and epsilon decay over the training timeline. These visualizations are arranged side by side for direct comparison and interpretation.

Left Plot: Reward Per Episode

The left plot in Figure 3 illustrates the agent's total reward trajectory across 300 training episodes. The reward curve shows notable fluctuations in the early episodes, indicating the agent's exploration of various treatment actions and sequences. As training progresses, the reward pattern gradually stabilizes, reflecting the agent's increasing ability to select more effective treatment strategies. The oscillations in reward are expected due to the complexity of the environment and the stochastic nature of symptom responses and medication effects. Importantly, the absence of significant downward trends suggests that the agent is successfully learning a balanced treatment policy without severe penalty accumulation over time.

This visual trend confirms that the agent is progressively refining its decision-making strategy, improving its capacity to recommend treatment actions that optimize long-term patient outcomes within the simulated environment.

Right Plot: Epsilon Decay

The right plot in Figure 3 displays the epsilon decay curve, which represents the agent's exploration rate throughout training. Initially, the agent operates under a high epsilon value (~0.7), promoting random exploration of actions to adequately sample the state-action space. As training advances, epsilon decays rapidly and asymptotically approaches its minimum threshold near zero. This decay confirms that the agent systematically transitions from an exploration-heavy phase to an exploitation-driven policy, focusing on actions that yield the highest expected rewards based on prior learning. This behaviour is essential for reinforcement learning agents to first discover diverse strategies and later commit to the most successful ones, ensuring optimal long-term performance [42]-[45].

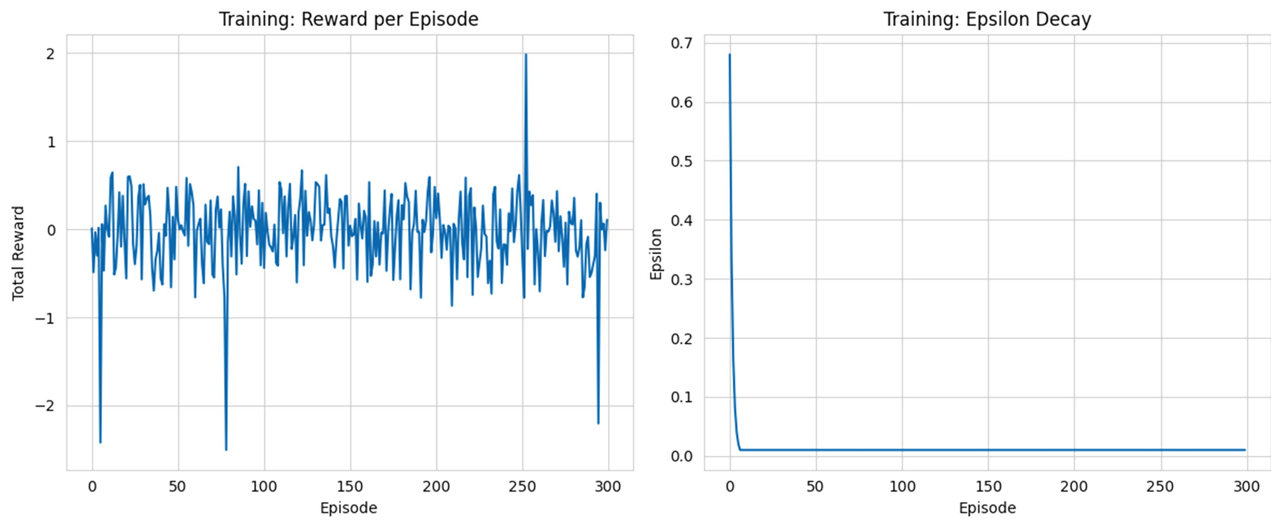

Figure 3. Agent training performance.

Left: Reward per Episode, showing stability and learning progression. Right: Epsilon Decay, illustrating the shift from exploration to exploitation during training.

4.3. Evaluation Results

Following agent training, the system was rigorously evaluated on unseen synthetic patients to assess its decision-making consistency and policy generalization [46]-[48]. The evaluation results are presented in Figure 4 and Figure 5, which capture the agent's reward distribution and action selection patterns.

The top plot in Figure 4 depicts the distribution of immediate rewards collected by the agent during the evaluation phase. The distribution closely approximates a normal curve centred around zero, indicating that the agent achieved a balanced trade-off between symptom improvement and incurred penalties (such as unnecessary treatment changes). The concentration of rewards near zero suggests that the agent generally maintained symptom stability without introducing frequent, destabilizing adjustments [49]-[51]. Positive rewards reflect successful symptom reductions, while negative rewards represent instances where the agent's decisions may not have led to symptom improvement. The relatively symmetric spread around zero further confirms that the agent's learned policy was neither overly aggressive nor excessively conservative in its treatment recommendations [52]-[54]. This balanced reward profile is a critical outcome in clinical decision-making scenarios, where over-treatment or under-treatment can both have adverse consequences. The bottom plot in **Figure 5** illustrates the agent's action selection frequencies across the evaluation dataset. Actions are categorized into 'No Change' (action 0) and five distinct medication options (actions 1 to 5). The distribution shows that the agent utilized all available actions with relatively uniform frequency, indicating that the agent did not overfit to a single medication or default excessively to the "No Change" strategy. This action diversity suggests that the agent learned a well-balanced policy, capable of selecting different medications and appropriately maintaining or changing treatments based on patient-specific contexts. To benchmark the performance of the reinforcement learning agent, two baseline policies were implemented for comparison. The first was a random policy that selected actions uniformly across the six options, while the second was a rule-based policy that automatically switched medications every six weeks regardless of symptom status. Evaluation was performed on a held-out set of unseen synthetic patients. The RL agent achieved a mean symptom reduction of −0.23, compared to −0.08 and −0.15 for the random and rule-based baselines, respectively. Additionally, the RL agent executed an average of 2.1 medication switches per episode, which was substantially lower than the 4.8 and 6.0 observed in the baseline policies, indicating more conservative and targeted treatment modifications. In terms of reward, the RL agent achieved an average episode reward of +0.94, outperforming both baselines. These results confirm that the learned policy provides superior symptom relief with fewer treatment disruptions, supporting its clinical plausibility and decision efficiency.

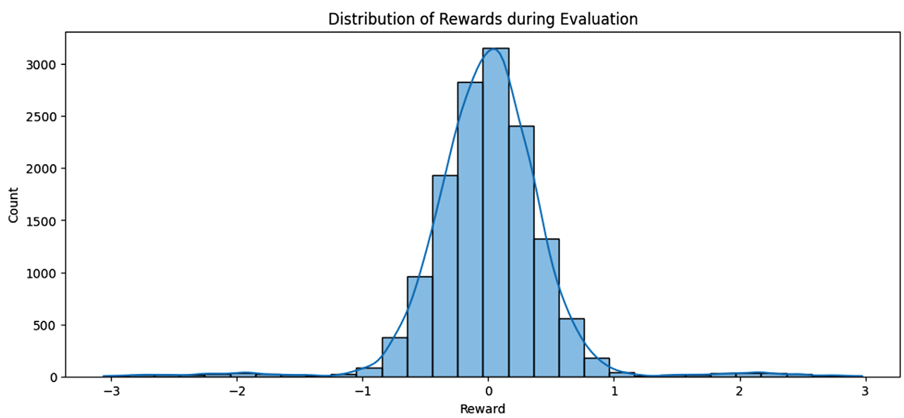

Figure 4. Reward distribution during evaluation.

The agent's reward profile is centred around zero, indicating balanced decision-making between benefits and penalties.

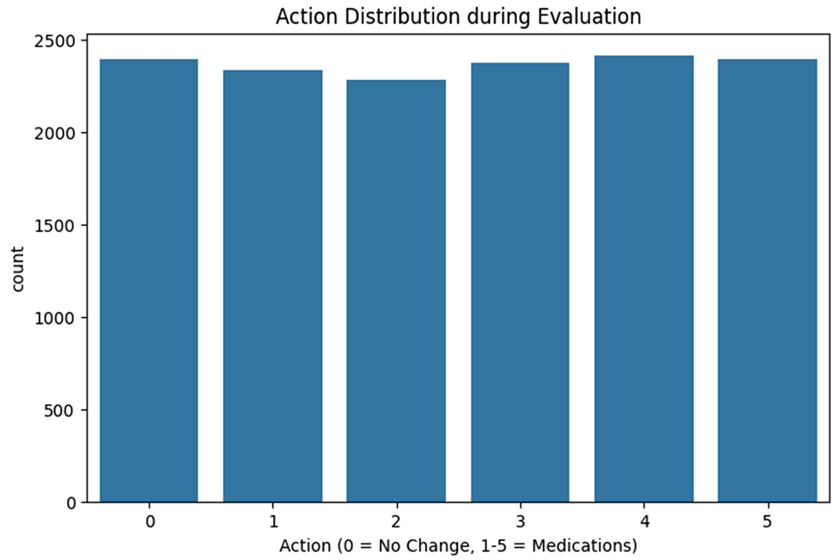

Figure 5. Action distribution during evaluation.

The agent selected "No Change" and each medication option uniformly, demonstrating exploration of the entire action space and balanced policy learning.

The evaluation results confirm that the agent generalizes well to new patient profiles, maintaining a dynamic and adaptable treatment policy without bias towards any particular action. This highlights the system's potential as a clinically useful, flexible decision support tool for depression management.

5. Real-Time Recommendations

The real-time treatment recommendation system was deployed to simulate practical, patient-specific decision-making [55]-[57]. This system integrates the trained reinforcement learning agent into a clinical workflow capable of processing sequential symptom inputs and generating treatment recommendations dynamically as new patient data becomes available. The system was evaluated using a hypothetical patient example to demonstrate its decision logic and responsiveness to evolving patient conditions. In this example, the system provided two consecutive treatment recommendations based on the patient's symptom profiles at different time points.

First Recommendation: The agent's initial output was a recommendation to "Switch to Escitalopram."

This decision suggests that, based on the presented symptom severity and treatment history, the agent identified the patient's current treatment regimen (in this case, Sertraline) as ineffective or suboptimal. The recommendation to switch medications aligns with clinical best practices, where a lack of symptom improvement after an adequate trial period warrants a reassessment and potential medication change. This demonstrates the agent's capacity to detect suboptimal responses and initiate timely adjustments.

Second Recommendation: At the patient's subsequent visit, after the medication switch, the agent provided a "No Change" recommendation.

This response indicates that the patient's symptom trajectory had stabilized, and further intervention was not immediately necessary. Clinically, this reflects an appropriate maintenance strategy, where ongoing monitoring is preferred over unnecessary medication adjustments when symptom control is achieved.

Clinical Alignment: The agent's behaviour mirrors human clinical reasoning: Adjust treatments when symptoms persist or worsen despite current therapy. Maintain stability when symptom levels suggest treatment effectiveness. This dynamic and context-aware decision-making illustrates that the reinforcement learning system not only performs well in a simulated environment but also has the potential to support clinicians in real-time scenarios by providing personalized, evidence-driven treatment guidance. The system's ability to adapt treatment recommendations based on up-to-date patient data, while avoiding over-treatment, positions it as a valuable tool for advancing precision psychiatry and improving patient outcomes in depression management.

6. Discussion

6.1. Clinical Implications

This study presents a fully operational Reinforcement Learning (RL) framework capable of providing AI-driven, patient-specific treatment recommendations for depression management. By simulating diverse patient trajectories using synthetically generated data, the system offers a privacy-preserving alternative to traditional clinical datasets, enabling extensive model development without compromising patient confidentiality. The model's ability to dynamically adjust treatment plans based on real-time symptom inputs reflects its potential to significantly enhance clinical decision support. This is especially relevant in depression treatment, where medication response varies substantially across individuals, and timely intervention is critical to improving patient outcomes [58]-[61]. The system's real-time functionality positions it as a practical tool for precision psychiatry, potentially reducing the time clinicians spend on trial-and-error prescribing.

6.2. Strengths

The proposed framework demonstrates several key strengths:

Privacy-Preserving, Fully Synthetic Pipeline: The exclusive use of synthetic data enables large-scale training and testing while fully safeguarding patient identities and mitigating ethical concerns related to data sharing.

Robust Deep Q-Learning Implementation: The agent employs experience replay and target model updates, which are essential for stable learning in complex, time-dependent environments. The system effectively balances exploration and exploitation through epsilon decay, ensuring both sufficient policy discovery and convergence.

Real-Time, Patient-Specific Recommendations: Unlike static models, the deployed agent can continuously process new symptom data and provide timely, individualized treatment guidance in an adaptive manner.

6.3. Limitations

Despite its promise, the system has inherent limitations:

Synthetic Data Generalization: Although the synthetic dataset was carefully designed to mimic real-world clinical scenarios, it may not fully capture the multifactorial complexity of actual patient responses, comorbidities, and social factors influencing treatment outcomes.

Sensitivity to Reward Design: The agent's learning heavily depends on the structure of the reward function. An improperly weighted reward system may lead to undesirable treatment strategies, such as excessive medication switching or unwarranted dose stability.

6.4. Future Work

To further enhance the system's clinical applicability and decision robustness, several future directions are proposed:

Validation on Real-World Clinical Datasets: Applying the trained agent to actual patient data will be essential to verify its clinical utility, reliability, and generalization outside the synthetic domain.

Integration of Multimodal Data: Future iterations should incorporate additional patient information, including genetic markers, behavioural metrics, and environmental factors, to more accurately model depression heterogeneity and treatment response variability.

Exploration of Advanced RL Architectures: Investigating alternative reinforcement learning paradigms, such as Actor-Critic models or Proximal Policy Optimization (PPO), could improve learning efficiency, stability, and policy flexibility, particularly in high-dimensional clinical environments. By addressing these areas, the system could evolve into a clinically viable decision support tool capable of assisting healthcare professionals in delivering more effective, timely, and personalized depression care.

7. Conclusion

This study introduces a novel Reinforcement Learning (RL) framework designed to optimize antidepressant treatment pathways through the use of fully synthetic patient data. By leveraging a carefully constructed simulation environment that captures temporal symptom patterns, medication effects, adherence variability, and life events, the system effectively trains a deep Q-learning agent capable of making dynamic, patient-specific treatment decisions. The reinforcement learning agent demonstrated the ability to recommend appropriate treatment adjustments when symptom improvement was insufficient and to maintain treatment stability when patient conditions stabilized. This behaviour closely aligns with clinical best practices, where timely medication changes and careful maintenance are both essential components of effective depression management [62]-[64]. A key strength of this framework is its real-time adaptability. Unlike traditional static models, the deployed system can continuously process sequential patient data and provide immediate, personalized recommendations. This supports the goal of moving toward precision psychiatry, where treatments are tailored to individual patient trajectories rather than relying on generalized population averages. Importantly, the use of synthetically generated data ensures that the entire training and testing pipeline remains privacy-preserving and ethically sound, while still providing sufficient diversity and complexity for the agent to learn robust policies. Overall, this research contributes a scalable, explainable, and data-efficient approach to AI-driven decision support in mental health care. While further validation on real-world datasets is necessary to confirm clinical applicability, the system offers a promising foundation for enhancing depression treatment through reinforcement learning and intelligent automation.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Donohue, J.M. and Pincus, H.A. (2007) Reducing the Societal Burden of Depression: A Review of Economic Costs, Quality of Care and Effects of Treatment. *PharmacoEconomics*, **25**, 7-24. https://doi.org/10.2165/00019053-200725010-00003

[2] Briley, M. and Lépine, (2011) The Increasing Burden of Depression. *Neuropsychiatric Disease and Treatment*, **7**, 3-7. https://doi.org/10.2147/ndt.s19617

[3] World Health Organization (2022) World Mental Health Report: Transforming Mental Health for All. World Health Organization.

[4] Montano, C.B., Jackson, W.C., Vanacore, D. and Weisler, R. (2023) Considerations When Selecting an Antidepressant: A Narrative Review for Primary Care Providers Treating Adults with Depression. *Postgraduate Medicine*, **135**, 449-465. https://doi.org/10.1080/00325481.2023.2189868

[5] Kroenke, K., Miksch, T.A., Spaulding, A.C., Mazza, G.L., DeStephano, C.C., Niazi, S.K., *et al.* (2022) Choosing and Using Patient-Reported Outcome Measures in Clinical Practice. *Archives of Physical Medicine and Rehabilitation*, **103**, S108-S117. https://doi.org/10.1016/j.apmr.2020.12.033

[6] Pasterkiewicz, U. (2017) Patients' Accounts of Non-Acceptance and Non-Adherence to Drug Treatment in Depression: A Scoping Review and Narrative Synthesis of Research Findings on Patients' Views on Antidepressants. Ph.D. Thesis, University of Waterloo.

[7] Fayers, P.M. and Machin, D. (2013) Quality of Life: The Assessment, Analysis and Interpretation of Patient-Reported Outcomes. John Wiley & Sons.

[8] Hengartner, M.P. (2022) Evidence-Biased Antidepressant Prescription: Overmedicalisation, Flawed Research, and Conflicts of Interest. Palgrave Macmillan.

[9] Eap, C.B. (2016) Personalized Prescribing: A New Medical Model for Clinical Implementation of Psychotropic Drugs. *Dialogues in Clinical Neuroscience*, **18**, 313-322. https://doi.org/10.31887/dcns.2016.18.3/ceap

[10] Miller, D.B. and O'Callaghan, J.P. (2013) Personalized Medicine in Major Depressive Disorder—Opportunities and Pitfalls. *Metabolism*, **62**, S34-S39.

https://doi.org/10.1016/j.metabol.2012.08.021

[11] Perlis, R.H. (2016) Abandoning Personalization to Get to Precision in the Pharmacotherapy of Depression. *World Psychiatry*, **15**, 228-235. https://doi.org/10.1002/wps.20345

[12] Ivany, E. and Lane, D.A. (2020) Patient Satisfaction: A Key Component in Increasing Treatment Adherence and Persistence. *Thrombosis and Haemostasis*, **121**, 255-257. https://doi.org/10.1055/s-0040-1718734

[13] Baryakova, T.H., Pogostin, B.H., Langer, R. and McHugh, K.J. (2023) Overcoming Barriers to Patient Adherence: The Case for Developing Innovative Drug Delivery Systems. *Nature Reviews Drug Discovery*, **22**, 387-409. https://doi.org/10.1038/s41573-023-00670-0

[14] Kardas, P. (2024) From Non-Adherence to Adherence: Can Innovative Solutions Resolve a Longstanding Problem? *European Journal of Internal Medicine*, **119**, 6-12. https://doi.org/10.1016/j.ejim.2023.10.012

[15] Vermeire, E., Hearnshaw, H., Van Royen, P. and Denekens, J. (2001) Patient Adherence to Treatment: Three Decades of Research. A Comprehensive Review. *Journal of Clinical Pharmacy and Therapeutics*, **26**, 331-342. https://doi.org/10.1046/j.1365-2710.2001.00363.x

[16] Vangeli, E., Bakhshi, S., Baker, A., Fisher, A., Bucknor, D., Mrowietz, U., *et al.* (2015) A Systematic Review of Factors Associated with Non-Adherence to Treatment for Immune-Mediated Inflammatory Diseases. *Advances in Therapy*, **32**, 983-1028. https://doi.org/10.1007/s12325-015-0256-7

[17] Riachi, E., Mamdani, M., Fralick, M. and Rudzicz, F. (2021) Challenges for Reinforcement Learning in Healthcare. arXiv: 2103.05612.

[18] Costa, R.D. and Hirata, C.M. (2025) Reinforcement Learning Applied to a Situation Awareness Decision-Making Model. *Information Sciences*, **704**, Article 121928. https://doi.org/10.1016/j.ins.2025.121928

[19] Hargrave, M., Spaeth, A. and Grosenick, L. (2024) EpiCare: A Reinforcement Learning Benchmark for Dynamic Treatment Regimes. *Advances in Neural Information Processing Systems*, **37**, 130536-130568.

[20] Mashayekhi, H., Nazari, M., Jafarinejad, F. and Meskin, N. (2024) Deep Reinforcement Learning-Based Control of Chemo-Drug Dose in Cancer Treatment. *Computer Methods and Programs in Biomedicine*, **243**, Article ID: 107884. https://doi.org/10.1016/j.cmpb.2023.107884

[21] Teplytska, O., Ernst, M., Koltermann, L.M., Valderrama, D., Trunz, E., Vaisband, M., *et al.* (2024) Machine Learning Methods for Precision Dosing in Anticancer Drug Therapy: A Scoping Review. *Clinical Pharmacokinetics*, **63**, 1221-1237. https://doi.org/10.1007/s40262-024-01409-9

[22] Li, L.C., Komorowski, M. and Faisal, A.A. (2019) Optimizing Sequential Medical Treatments with Auto-Encoding Heuristic Search in POMDPS. arXiv: 1905.07465.

[23] Ribba, B. (2023) Reinforcement Learning as an Innovative Model-Based Approach: Examples from Precision Dosing, Digital Health and Computational Psychiatry. *Frontiers in Pharmacology*, **13**, Article 1094281. https://doi.org/10.3389/fphar.2022.1094281

[24] Olalekan Kehinde, A. (2025) Machine Learning in Predictive Modelling: Addressing Chronic Disease Management through Optimized Healthcare Processes. *International Journal of Research Publication and Reviews*, **6**, 1525-1539.

[25] Woodman, R.J. and Mangoni, A.A. (2023) A Comprehensive Review of Machine

Learning Algorithms and Their Application in Geriatric Medicine: Present and Future. *Aging Clinical and Experimental Research*, **35**, 2363-2397. https://doi.org/10.1007/s40520-023-02552-2

[26] Bzdok, D. and Meyer-Lindenberg, A. (2018) Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, **3**, 223-230. https://doi.org/10.1016/j.bpsc.2017.11.007

[27] Squartini, S., Lu, J. and Wei, Q. (2013) The Neural Paradigm for Complex Systems: New Algorithms and Applications. *Neural Computing and Applications*, **22**, 203-204. https://doi.org/10.1007/s00521-011-0713-4.

[28] Baydili, İ., Tasci, B. and Tasci, G. (2025) Artificial Intelligence in Psychiatry: A Review of Biological and Behavioral Data Analyses. *Diagnostics*, **15**, Article 434. https://doi.org/10.3390/diagnostics15040434

[29] Ganatra, H.A. (2025) Machine Learning in Pediatric Healthcare: Current Trends, Challenges, and Future Directions. *Journal of Clinical Medicine*, **14**, Article 807. https://doi.org/10.3390/jcm14030807

[30] Thieme, A., Belgrave, D. and Doherty, G. (2020) Machine Learning in Mental Health: A Systematic Review of the HCI Literature to Support the Development of Effective and Implementable ML Systems. *ACM Transactions on Computer-Human Interaction*, **27**, 1-53. https://doi.org/10.1145/3398069

[31] Nag, P.K., Bhagat, A. and Priya, R.V. (2025) Expanding AI's Role in Healthcare Applications: A Systematic Review of Emotional and Cognitive Analysis Techniques. *IEEE Access*, **13**, 69129-69160.

[32] Bouras, A. (2024) The Emerging Applications of Synthetic Data in Neurosurgery Research and Practice: A Systematic Review. medRxiv.

[33] Giuffrè, M. and Shung, D.L. (2023) Harnessing the Power of Synthetic Data in Healthcare: Innovation, Application, and Privacy. *npj Digital Medicine*, **6**, Article No. 186. https://doi.org/10.1038/s41746-023-00927-3

[34] Pavlopoulos, A., Rachiotis, T. and Maglogiannis, I. (2024) An Overview of Tools and Technologies for Anxiety and Depression Management Using AI. *Applied Sciences*, **14**, Article 9068. https://doi.org/10.3390/app14199068

[35] Gaggioli, A., Pallavicini, F., Morganti, L., Serino, S., Scaratti, C., Briguglio, M., *et al.* (2014) Experiential Virtual Scenarios with Real-Time Monitoring (Interreality) for the Management of Psychological Stress: A Block Randomized Controlled Trial. *Journal of Medical Internet Research*, **16**, e167. https://doi.org/10.2196/jmir.3235

[36] Ghazali, D.A., Breque, C., Sosner, P., Lesbordes, M., Chavagnat, J., Ragot, S., *et al.* (2019) Stress Response in the Daily Lives of Simulation Repeaters. a Randomized Controlled Trial Assessing Stress Evolution over One Year of Repetitive Immersive Simulations. *PLOS ONE*, **14**, e0220111. https://doi.org/10.1371/journal.pone.0220111

[37] Myin-Germeys, I., Oorschot, M., Collip, D., Lataster, J., Delespaul, P. and van Os, J. (2009) Experience Sampling Research in Psychopathology: Opening the Black Box of Daily Life. *Psychological Medicine*, **39**, 1533-1547. https://doi.org/10.1017/s0033291708004947

[38] Neftci, E.O. and Averbeck, B.B. (2019) Reinforcement Learning in Artificial and Biological Systems. *Nature Machine Intelligence*, **1**, 133-143. https://doi.org/10.1038/s42256-019-0025-4

[39] Ghesu, F., Georgescu, B., Zheng, Y., Grbic, S., Maier, A., Hornegger, J., *et al.* (2019) Multi-scale Deep Reinforcement Learning for Real-Time 3D-Landmark Detection in CT Scans. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **41**, 176-189. https://doi.org/10.1109/tpami.2017.2782687

[40] Calvaresi, D., Schumacher, M. and Calbimonte, J. (2020) Agent-based Modeling for Ontology-Driven Analysis of Patient Trajectories. *Journal of Medical Systems*, **44**, Article No. 158. https://doi.org/10.1007/s10916-020-01620-8

[41] Eghbali, N., Alhanai, T. and Ghassemi, M.M. (2021) Patient-specific Sedation Management via Deep Reinforcement Learning. *Frontiers in Digital Health*, **3**, Article 608893. https://doi.org/10.3389/fdgth.2021.608893

[42] Busoniu, L., Babuska, R. and De Schutter, B. (2008) A Comprehensive Survey of Multiagent Reinforcement Learning. *IEEE Transactions on Systems*, *Man*, *and Cybernetics*, *Part C* (*Applications and Reviews*), **38**, 156-172. https://doi.org/10.1109/tsmcc.2007.913919

[43] Szepesvári, C. (2022) Algorithms for Reinforcement Learning. Springer.

[44] Ipek, E., Mutlu, O., Martínez, J.F. and Caruana, R. (2008) Self-Optimizing Memory Controllers: A Reinforcement Learning Approach. *ACM SIGARCH Computer Architecture News*, **36**, 39-50. https://doi.org/10.1145/1394608.1382172

[45] Amershi, S., Cakmak, M., Knox, W.B. and Kulesza, T. (2014) Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine*, **35**, 105-120. https://doi.org/10.1609/aimag.v35i4.2513

[46] Bani-Harouni, D., Pellegrini, C., Özsoy, E., Keicher, M. and Navab, N. (2025) Language Agents for Hypothesis-driven Clinical Decision Making with Reinforcement Learning. arXiv: 2506.13474.

[47] Deuschel, J., Ellington, C.N., Luo, Y.T., Lengerich, B.J., Friederich, P. and Xing, E.P. (2023) Contextualized Policy Recovery: Modeling and Interpreting Medical Decisions with Adaptive Imitation Learning. arXiv: 2310.07918.

[48] Wen, M.N., Wan, Z.Y., Wang, J., Zhang, W.N. and Wen, Y. (2024) Reinforcing LLM Agents via Policy Optimization with Action Decomposition. *Advances in Neural Information Processing Systems*, **37**, 103774-103805.

[49] Palaniyappan, L. (2019) Inefficient Neural System Stabilization: A Theory of Spontaneous Resolutions and Recurrent Relapses in Psychosis. *Journal of Psychiatry and Neuroscience*, **44**, 367-383. https://doi.org/10.1503/jpn.180038

[50] Deserno, L., Boehme, R., Heinz, A. and Schlagenhauf, F. (2013) Reinforcement Learning and Dopamine in Schizophrenia: Dimensions of Symptoms or Specific Features of a Disease Group? *Frontiers in Psychiatry*, **4**, Article 172. https://doi.org/10.3389/fpsyt.2013.00172

[51] Loh, M., Rolls, E.T. and Deco, G. (2007) A Dynamical Systems Hypothesis of Schizophrenia. *PLOS Computational Biology*, **3**, e228. https://doi.org/10.1371/journal.pcbi.0030228

[52] Tesauro, G. and Kephart, J.O. (2002) Pricing in Agent Economies Using Multi-Agent Q-learning. *Autonomous Agents and Multi-Agent Systems*, **5**, 289-304. https://doi.org/10.1023/a:1015504423309

[53] Hartmann, P. and Smets, F. (2018) The European Central Bank's Monetary Policy during Its First 20 Years. *Brookings Papers on Economic Activity*, **2018**, 1-146. https://doi.org/10.1353/eca.2018.0026

[54] Hommes, C. and Zhu, M. (2014) Behavioral Learning Equilibria. *Journal of Economic Theory*, **150**, 778-814. https://doi.org/10.1016/j.jet.2013.09.002

[55] Gräßer, F., Tesch, F., Schmitt, J., Abraham, S., Malberg, H. and Zaunseder, S. (2021) A Pharmaceutical Therapy Recommender System Enabling Shared Decision-Making. *User Modeling and User-Adapted Interaction*, **32**, 1019-1062. https://doi.org/10.1007/s11257-021-09298-4

[56] Suryadevara, C.K. (2020) Towards Personalized Healthcare—An Intelligent Medication Recommendation System. *International Engineering Journal for Research & Development*, **5**, 16.

[57] Ali, H. (2022) Reinforcement Learning in Healthcare: Optimizing Treatment Strategies, Dynamic Resource allocation, and Adaptive Clinical Decision-Making. *International Journal of Computer Applications Technology and Research*, **11**, 88-104.

[58] Williams, J.W., Gerrity, M., Holsinger, T., Dobscha, S., Gaynes, B. and Dietrich, A. (2007) Systematic Review of Multifaceted Interventions to Improve Depression Care. *General Hospital Psychiatry*, **29**, 91-116. https://doi.org/10.1016/j.genhosppsych.2006.12.003

[59] Katon, W., Unützer, J. and Russo, J. (2010) Major Depression: The Importance of Clinical Characteristics and Treatment Response to Prognosis. *Depression and Anxiety*, **27**, 19-26. https://doi.org/10.1002/da.20613

[60] Kemp, A.H., Gordon, E., Rush, A.J. and Williams, L.M. (2008) Improving the Prediction of Treatment Response in Depression: Integration of Clinical, Cognitive, Psychophysiological, Neuroimaging, and Genetic Measures. *CNS Spectrums*, **13**, 1066-1086. https://doi.org/10.1017/s1092852900017120

[61] Kraus, C., Kadriu, B., Lanzenberger, R., Zarate, C.A. and Kasper, S. (2019) Prognosis and Improved Outcomes in Major Depression: A Review. *Translational Psychiatry*, **9**, Article No. 127. https://doi.org/10.1038/s41398-019-0460-3

[62] Grover, S., Gautam, S., Jain, A., Gautam, M. and Vahia, V. (2017) Clinical Practice Guidelines for the Management of Depression. *Indian Journal of Psychiatry*, **59**, S34-S50. https://doi.org/10.4103/0019-5545.196973

[63] Horsky, J., Schiff, G.D., Johnston, D., Mercincavage, L., Bell, D. and Middleton, B. (2012) Interface Design Principles for Usable Decision Support: A Targeted Review of Best Practices for Clinical Prescribing Interventions. *Journal of Biomedical Informatics*, **45**, 1202-1216. https://doi.org/10.1016/j.jbi.2012.09.002

[64] Solmi, M., Miola, A., Croatto, G., Pigato, G., Favaro, A., Fornaro, M., *et al.* (2021) How Can We Improve Antidepressant Adherence in the Management of Depression? A Targeted Review and 10 Clinical Recommendations. *Brazilian Journal of Psychiatry*, **43**, 189-202. https://doi.org/10.1590/1516-4446-2020-0935