

Simulated Light Schedules for Sleep-Wake Adjustments: Exploring the Incorporation of Reinforcement Learning

Ming-yi Seah

Monash University Malaysia

Joanne Mun Yee Lim

Monash University Malaysia

Shiuan-Ni Liang

Monash University Malaysia

Yatinesh Kumari

Monash University Malaysia

Vineetha Kalavally

vineetha@monash.edu

Monash University Malaysia

Article

Keywords: Light schedule, sleep-wake adjustment, machine learning, reinforcement learning

Posted Date: October 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10233757/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

This study investigates the potential of a development framework for generating personalized light schedules aimed at achieving specific sleep-wake cycle targets. Unlike existing approaches that rely on predefined heuristics or static lighting recommendations, the framework formulates light scheduling as a sequential decision-making problem and employs a reinforcement learning–based deep deterministic policy gradient (DDPG) algorithm to iteratively modify real-world light exposure patterns toward achieving a desired sleep-wake class. This process is assisted by a convolutional neural network-based sleep-wake classification model trained on ecologically collected data. The circadian impacts of the generated schedules are evaluated using dim light melatonin onset (DLMO) simulations from a validated circadian model. By incorporating melanopic alongside photopic illuminances and embedding industry-oriented constraints, this work explores possibilities of a novel method in generating and modifying light schedules for sleep-wake adjustments using modern approaches, standards and findings.

Introduction

Light has been known to be the main contributor for circadian entrainment when compared to other non-photoc factors such as temperature, noise, food, exercise, and melatonin [1, 2]. This led to various research investigating the relationships between light and systems related to the circadian rhythm, especially dim light melatonin onset (DLMO) and sleep. Investigations varied from experiments carried out to determine the general effects of light on circadian rhythms [3–5] to more targeted inspections such as individual sensitivity to light and the impacts of spectral differences in light [6–9]. While most of these research studies were carried out under controlled laboratory settings, light exposure in the real-world is more intricate and complicated, motivating similar research to be carried out under ecological settings [10, 11]. Reviews were also published to summarize the effects of light on the circadian rhythm that had been previously investigated [12, 13].

Knowing the precise effects of light on human circadian rhythm is important as it is usually measured via proxies such as core body temperature minima (CBT_{min}) or dim light melatonin onsets (DLMO) which is the clock time at which melatonin levels first rise above a threshold in the evening, marking the start of the biological night. The corresponding measurement processes for both of which are both resource- and time-consuming. These difficulties could be circumvented if models capable of estimating circadian rhythms based on parameters that are easily accessible were developed. This naturally led to developments of various photobiological models that replicated the effects of light on the circadian rhythm since light exposure is significantly easier and less intrusive to measure while also being the most influential [14–17].

The capabilities of robust circadian models to estimate circadian rhythms are also particularly useful in designing light interventions or ‘light diets’ where they would be used to simulate the resulting circadian rhythms of different light schedules. In the day and age when travelling across time-zones is relatively common and shift work is a norm, these interventions serve to assist in targeted phase shifting of the circadian rhythm, hence avoiding the downsides of having to operate while having a mismatched circadian rhythm. The design process of past light intervention research often revolved around a methodology called the “bang-bang” control, which involves requiring individuals to be exposed to alternating extremes of brightness and darkness, as the control method allowed for the fastest circadian adjustments [18, 19]. However, these light diets could be impractical since it is unlikely for individuals to compromise their daily activities severely in the process of committing to light diets [20]. This motivates the need for curations of light diets that are more practical [20–22].

Light diets are essentially recommendations for daily light exposure. This inspired the idea of incorporating machine learning methods into light diet design, more specifically reinforcement learning as the algorithm is currently very widely used in recommendation systems of various industries [23–26]. For machine learning algorithms to work efficiently, large training data sizes are preferred to minimize the possibility of underfitting. This could be difficult since it is demanding to carry out circadian rhythm assessing experiments for large pools of individuals. Therefore, we attempted to curate frameworks around the concept of using sleep-wake behavioral metrics, which are more practical to measure than dim light melatonin onsets (DLMO), as indirect surrogate markers of circadian phase for generating personalized light exposure schedules. Prior studies have shown that certain sleep–wake indicators correlate with circadian rhythms, particularly in individuals with regular schedules, non-shift workers, and those without circadian misalignment, suggesting their potential utility as feasible alternatives to physiologically based assessments [4, 27–31].

This work addresses a key gap in the development of practical light diets for sleep–wake adjustment by introducing a reinforcement learning–based approach to generate adaptive lighting schedules that target specific sleep–wake outcomes. To the best of the authors’ knowledge, while machine learning has been utilized to develop circadian prediction models, this represents one of the early explorations of machine learning to light diet design and generation, despite the proven success of such methods in many other domains. By integrating industry-recommended light exposure levels into the optimization process and novel machine learning methodologies, the curation process of the proposed frameworks serves as a “prototype” to kickstart the advancement of light diet generation towards the modern age of big data, ideally maturing alongside the constant growth of our knowledge in the field of practical circadian light.

Methods

In this study, a data-driven development framework which incorporated a variation of reinforcement learning (RL) called the deep deterministic policy gradient (DDPG) algorithm was introduced and investigated (Fig. 1). In this framework, data collection of experimental daily light exposure and sleep-wake parameter values from participants was required. The starting phase of the introduced framework involved defining sleep-wake classes wherein collected sleep-wake parameters were used to group participants’ light exposure schedules into different sleep-wake classes. Sleep-wake class label generation was carried out using a data-driven clustering algorithm which ensured that the dataset was optimally grouped into appropriate classes according to the collected sleep-wake times while also removing potential outliers, minimizing the effects of outlier pollution. The second phase involved developing a model

that predicted the sleep-wake class of an input light schedule. Then, this model would be integrated into the DDPG algorithm during the final phase where light schedules were generated to achieve a desired sleep-wake class.

Data Collection

An experiment was conducted for 20 participants where their daily light exposure data and sleep-wake parameters were collected. The experiment was approved by Monash Human Research Ethics Committee (MUHREC, Project No. 0794) and was carried out in accordance with MUHREC's guidelines and regulations. Written informed consents were obtained from participating individuals, and they were financially reimbursed for their time at the end of the experiment. They were allowed to leave the study at any point of the study while conceding reimbursement. Individuals were eligible for the experiment if they were aged between 18–30 years old and did not fall under the following exclusion criteria [16, 32, 33]:

1. Daily caffeine consumption > 400 mg (3–4 cups) per day
2. Alcohol consumption > 2 drinks per day or 14 drinks a week
3. History of drug abuse
4. Pregnant or breast-feeding
5. Smoke
6. Under daily medication
7. Suffering from sleep disorders
8. Suffering from vision problems (e.g. colour blindness)
9. Suffered spinal cord injuries
10. Suffering from liver or kidney diseases
11. Suffering from periodontal diseases (e.g. chronic inflammation)
12. Inter-meridian travel in the past two months

The selection of young adults aged between 18–30 years old was done to minimize the effects of conditions associated with old age on their circadian rhythms. The participants mainly consisted of university students from Monash University Malaysia, and no shift worker was recruited for this experiment.

The experiment lasted for approximately seven days in which participants were required to wear two dosimeters, namely a wearable mini spectral sensor SPECCY (Photospeclabs Pty Ltd, Melbourne, Australia) and GENEActiv accelerometer (ActivInsights Ltd, Kimbolton, Cambridgeshire, U.K.). Light exposure data collected from SPECCY were quantified using photopic illuminance as well as α -opic illuminances where each illuminance is measured based on the spectral sensitivities of the corresponding α photoreceptor [CIE S2026/2018]. This notably includes melanopic EDI (MEDI), measured according to the spectral sensitivity of intrinsically photosensitive ganglion cells (ipRGCs), which have been found to be influential on the circadian system [34, 35]. SPECCY was worn around the participants' neck at chest level similarly to a pendant to collect their daily light exposure data at 60-s intervals, whilst the GENEActiv accelerometer was worn on the participants' non-dominant wrist to monitor their actigraphy data and in turn assess their sleep-wake parameters. The accelerometer was configured to collect data at 85.70 Hz, which was the maximum frequency available that allowed for data collection lasting more than 7 days.

The dosimeters were always worn except for when the participants would be exposed to significant amounts of water. There were no restrictions on the participants' sleep schedules, but they were advised not to have extreme schedules, for example pulling all-nighters. Additionally, participants were asked to refrain from exercising during the night before their bedtime.

Participants collected and started wearing their assigned dosimeters on day 0 of the experiment. During collection of the dosimeters, participants answered three sleep-related questionnaires to assess their chronotypes, namely the Munich Chronotype Questionnaire (MCTQ), the Morningness-Eveningness Questionnaire (MEQ), and the Pittsburgh Sleep Quality Index (PSQI). During the experiment, participants filled in sleep diaries before they went to sleep and after they woke up. The sleep diary filled in before they slept enquired about consumption of medications, caffeinated and alcoholic drinks, whereas the sleep diary answered when they woke up enquired information about their sleep-wake times. This study's data collection concluded after six days of light and actigraphy data were collected for each participant.

Data Retention and Cleaning

Data starting from midnight on day 1 to midnight of day 6 was used for this research. For a dataset to be retained, it must adhere to the following data retention criteria adapted from those introduced by Stone et al. [33]:

1. Recorded less than 21 h (~ 15% of 6 d × 24 h) of missing data.
2. No continuous time intervals of missing data were greater than 3 hours for the first day of measurement and 18 hours in total for subsequent days.

The 20 participants' light exposure data were cleaned by first identifying the times at which the participants were asleep (sleep onset) or awake (wake time) through their recorded sleep-wake times in their sleep diaries and actigraphy data collected through the GENEActiv accelerometer. If there existed discrepancies that were greater than 60 minutes between the participants subjective sleep/wake times and the recorded reduction/rise in actigraphy, adjustments to their sleep-wake times were made to match the observed time of decrease/increase in activity [16].

Next, noise and artefacts were identified using the following rules and removed from the dataset:

1. Photopic illuminance < 1 lx during wake as this was most likely caused by blocking of the sensing element [33].
2. Sudden drops where photopic illuminance of a data point is < 10 lx and < 10% of those immediately before and after [6].

The removed data points were replaced according to the following methods:

1. For noise removed for having photopic illuminance < 1 lx during wake, the continuous time gap of missing data was determined. If the interval was > 1 h, it was filled with illuminance levels of the same time from the previous day [33]. If the interval was < 1 h, it was instead replaced with average illuminance levels obtained from the two hours prior to the interval.
2. The average illuminance levels of light data points immediately before and after were used to replace sudden drops.

Finally, light intensity values were set to 0 lx when the participants were asleep as light should be blocked from entering their eyes by closed eyelids.

Based on the criteria listed above, 3 participants' light exposure data were recorded to be missing 1.30% (131/10081), 11.41% (1150/10081), and 73.02% (7361/10081). The participant with 73.02% missing data was excluded as more than 21 h of data was missing. Hence, the final dataset consisted of light exposure and sleep-wake data from 19 participants (aged 21.11 ± 2.64 years old, 7 males, 12 females).

Sleep-Wake Class Definition using Hierarchical Agglomerative Clustering

After the experiment, the participants' average sleep-wake parameters across 6 days were obtained. The sleep-wake parameters that were analyzed in this research include sleep onset (SO), wake time (WT), mid-sleep (MS), total sleep time (TST) and sleep efficiency (SE), which were calculated with the assistance of the GENEActiv R markdown tool (example of GENEActiv provided in Supplementary Material 1.0).

SO and WT were obtained from the participants' recorded sleep/wake times in their sleep diaries and adjusted to observed times of reduction/rise in actigraphy if there were discrepancies greater than 60 minutes between the participants' subjective sleep-wake times and the recorded reduction/rise in actigraphy by the GENEActiv accelerometer. MS was calculated as the midpoint between SO and WT. TST and SE were obtained from the R markdown tool, where TST was the time participants were asleep during the period between SO and WT, and SE was the percentage ratio of the participants' TST to the period between SO and WT. TST and SE were calculated using the sleep/wake times identified by the markdown tool. One correction was made to TST and SE throughout the whole dataset where there was a clear mistake made by the markdown tool (two sleep/wake times recorded for the same day). Sleep-wake metric values for each day were determined as those occurring at the end of the day transversing to the start of the next day.

In the first phase of the proposed framework, the participants were grouped into sleep-wake classes based on their average sleep-wake parameters throughout the experiment (minima and maxima used to standardized sleep-wake parameter values are provided in Supplementary Material 2.0). We opted to use hierarchical agglomerative clustering [36] to suit the relatively small sample size of our dataset. To identify the optimal number of classes to group the participants, three cluster quality scores were utilized: Silhouette Score [37], the Calinski-Harabasz Score [38], and the Davies-Bouldin Score [39]. As a result, five sleep-wake classes were obtained from the clustering algorithm and were labeled Class 1–5, sorted from earliest average MS to latest average MS (Fig. 2).

Most participants were grouped into Class 3 with a moderately early average MS of 05:00, long average TST of 6 hours and 5 minutes, and a high average sleep efficiency of 78.28%. Class 2 and 3 had similar MSs, but Class 2 had a shorter average TST and lower average SE. Class 5 had the latest average MS but the highest average SE. This hinted at the possibility that participants in this class had sleep-wake habits that were least affected by sociological factors when compared to other classes.

Data Augmentation

Data augmentation was applied to our dataset to increase its sample size and ensure that the algorithms and models developed upon this dataset were more robust. This was done by cyclically rotating the participants' daily light schedules which was inspired by Brown et al.'s classification approach to estimating human circadian phase from actigraphy and photometry data [40]. This method is justified because the average sleep-wake parameters of the participants would remain the same when the sequences of their daily light schedules were changed throughout the week. This effectively increased the dataset by 6-fold as a result of having six different starting days for each sample. After augmentation, all the classes were down sampled to follow the number of samples in the minority class (Class 5) to ensure class balance. This led to our final dataset having a total of 12 samples per class (60 samples in total).

Light Schedule Generation using Reinforcement Learning

As shown in Fig. 1, a data-driven model was developed to predict the users' sleep-wake class based on their daily light exposure after an optimal number of sleep-wake classes was defined via clustering. Aside from the obvious use case of predicting one's sleep-wake time, it was also used to verify the light schedules generated in the final phase of the framework. The final phase of this framework involved the generation of artificial light schedules using reinforcement learning. A user would choose a desired sleep-wake class, and the framework would ideally output a light schedule with the desired sleep-wake class that was modified from the user's original daily routine.

Sleep-Wake Classification using Convolutional Neural Network

For the second phase of our proposed light schedule generation framework, a classification model that predicts sleep-wake classes based on daily light exposure was developed. We chose the convolutional neural network (CNN) architecture to form the basis of our classification model following our designed structure for light schedule input. The collected α -opic and photopic illuminance values were averaged across each hour, resulting in 24 hourly averages for six illuminances across a 6-day period. This resulted in the data having dimensions of $24 \times 6 \times 6$, which was similar to colored images with six color channels instead of three. This motivated the usage of convolutional neural networks (CNNs) as the architecture has been shown to be effective in imagery-related applications [41–49]. The CNN would take this $24 \times 6 \times 6$ input and output one of five classes obtained from the previous phase.

To train the CNN, our augmented dataset was split into training and test dataset following an 80:20 ratio (48 training samples, 12 test samples). 20% of the training dataset was used to validate the model's classification results during training. The architecture and training results of the CNN were shown in Table 1. The model demonstrated a classification accuracy of 91.67% (11/12) when it was evaluated using the test dataset.

Table 1
Architecture and training results of sleep-wake classification CNN model.

Sleep-Wake Classification CNN Model Architecture		
Type of Layer	Hyperparameter	Value
Input	No. of Neuron	$24 \times 6 \times 6$
Convolutional	No. of Filters	128
	Kernel Size	3×3
	Strides	1×1
	Padding	Same
	Activation Function	Rectified Linear Unit (ReLU)
Pooling	Type	Max
	Size	2×2
Convolutional	No. of Filters	128
	Kernel Size	2×2
	Strides	1×1
	Padding	Same
	Activation Function	Rectified Linear Unit (ReLU)
Pooling	Type	Max
	Size	2×2
Flatten	-	-
Dense	No. of Neuron	256
	Batch Normalization	True
	Activation Function	ReLU
Dense	No. of Neuron	124
	Activation Function	Linear
Output	No. of Neuron	5
	Activation Function	Softmax
Loss Function		Sparse Categorical Cross Entropy
Optimizer		Adam
Training Result		
No. of epoch		390
Accuracy		1.0000
Loss		1.8217e-05
Validation Accuracy		1.0000
Validation Loss		0.0093

Incorporation of Deep Deterministic Policy Gradient Reinforcement Learning

The classification model was essential to the reinforcement learning (RL) light schedule generation framework in this study to determine whether a generated schedule could achieve a targeted sleep-wake class. RL is a machine learning algorithm often used in recommendations systems. This learning algorithm consists of several key components that interact with each other to ensure that the best action could be carried out for maximum rewards. We adapted a variation of reinforcement learning called deep deterministic gradient policy (DDPG) which is an extension of deterministic gradient policy (DPG) that incorporates deep Q-learning (DQN) [50]. DPG enables reinforcement learning to be applied over continuous action spaces whereas the DQN component allows for high dimensionality, which both were aspects required for our reinforcement learning implementation in the context of light schedule modification. The framework and training process of our algorithm are shown in Algorithm 1. The general architecture of a DDPG algorithm and a flowchart of the DDPG light generation algorithm are provided in Supplementary Material 3.0.

The first component in RL is the environment, which is the system that the model or algorithm interacts with. The environment in our context of light schedule generation was defined as the individual's 6-day light schedule processed in the logarithmic scale. The next component is the state of the environment represented by s_t where t is the time step. This component describes the circumstances that the environment is experiencing at a certain time. The state of our environment consisted of 6 days x 24 hourly averages x six photopic and α -opic illuminances. The illuminances were bounded by two upper limits: (i) $10^{2.7}$ lx (~ 501 lx) during the night (20:00–07:00) and (ii) $10^{3.2}$ lx ($\sim 1,585$ lx) during the day (08:00–19:00) to ensure that the generated schedules were practical. These upper bounds were chosen because they were still achievable via the assistance of artificial lightings during the respective periods of time, ensuring that the generated schedules remained practical with minimal dependence on natural circumstances.

The action, a_t , is an act carried out by the agent to change the environment at time step t . Each action taken in the environment is defined using 13 continuous values. The 1st value depicts whether a light schedule should be delayed or advanced. The next 6 values (2nd – 7th) and the final 6 values (8th – 13th) are factors used to scale each illuminance values during the night (20:00–07:00) and during the day (08:00–19:00) respectively. These factors were limited to range between 0.90–1.10 to control the changes introduced to the schedules at each iteration. This was done to reduce the probability of the states not converging towards a solution or saturating towards prohibited conditions.

One of the most important components in RL is the reward, represented by r_t . It provides feedback on the action a_t , and it directly contributes to the efficiency of the implemented RL algorithm. The sleep-wake classification model developed in phase 2 of the framework was used to generate this reward to the environment. The value was quantified as the difference between the predicted sleep-wake class of s_t and the targeted sleep-wake class. Additionally, training was awarded a bonus if the modified light schedule achieved the target sleep-wake class and penalized if it entered a prohibited state. Prohibited states were defined as follows: (i) all values were lower than 1, implying that the schedule was digressing towards an all-zero state; (ii) number of unique values was less than 50, indicating that the schedule was saturating at the introduced upper bounds. Once a prohibited state was detected, the training episode was terminated, and a new episode was started with a fresh environment. The policy in RL is the main driving force when it comes to learning the solutions to a problem. It encompasses a strategy that maps s_t to a_t and the goal of an RL algorithm's training is to curate a policy that maximizes cumulative rewards.

In DDPG, the actor-critic algorithm is utilized to refine its policy. The actor-critic algorithm, as its name suggests, is composed of an actor and a critic model. The “deep” component in DDPG signifies the usage of multilayer-perceptron (MLP) networks for these two models. The actor network takes in s_t and outputs a_t based on policy μ , and its parameters are denoted by θ^μ . It is responsible for deciding the appropriate a_t that should be taken after observing s_t based on the policy learnt. On the other hand, the critic network parameterized by θ^Q , takes in s_t and a_t and outputs $Q(s_t, a_t)$, which is the expected return, also called the Q-value, for a_t given s_t . The general idea of the actor-critic algorithm's workings is that the critic network will provide feedback to the actor network's decisions, and the actor network will learn and update its policy according to the critic network's assessment to yield higher expected rewards.

In addition to the aforementioned components, target actor and critic networks are also utilized in DDPG. They are essentially delayed copies of the actor and critic networks which are updated less frequently. Their purpose is to stabilize the algorithm and improve performance. Additionally, there is also an experience replay memory, which is a memory buffer that stores transitions (s_t, a_t, r_t, s_{t+1}) from previous episodes. Mini batches would be randomly sampled from the buffer to update the networks' parameters and policies. To promote exploration, noise is generated using the Ornstein-Uhlenbeck process [51] and added to the output of the actor model during training.

The completed DDPG algorithm was used to generate 100 light schedules for each sleep-wake class, all derived from schedules randomly selected from the original dataset. To gauge the efficacy of the DDPG algorithm in achieving targeted sleep-wake classes, an open-source prototype assessment tool encapsulating the implementation of a model that predicted dim light melatonin onsets (DLMOs) based on light and activity input provided by Cheng et al. [28] was utilized to simulate the DLMOs of these schedules (<http://www.predictdlmo.com>). Additionally, the model was used to simulate the DLMOs of the schedules collected from the participants and correlation between the simulated DLMOs and the participants' sleep-wake parameters were calculated and shown in Supplementary Material 4.0.

Algorithm 1: Training Process of Deep Deterministic Policy Gradient (DDPG) Algorithm.		
1.	C:	List to contain shuffled sleep-wake class labels.
2.		34 light source combinations, indices
3.		correspond to the labels assigned to each
4.		light source.
5.	DDPG:	DDPG algorithm, takes in original light schedule and targeted sleep-wake class. Outputs
5.		modified light schedule.
6.	D:	Dataset containing original light schedules from experiment.
7.	s_t :	24 x 6 x 6 light schedule.
8.	r_e :	Episodic reward.
9.	r_t :	Reward at time step t.
11.		$size \leftarrow 200$
12.		$\backslash \text{varvec}f \backslash \text{varveco} \backslash \text{varvecr}class \backslash \text{varveci} \backslash \text{varvecn} \{1, 2, 3, 4, 5\} \backslash \text{varvecd} \backslash \text{varveco}$
13.		$addclasstoC \backslash \text{varvec}f \backslash \text{varveco} \backslash \text{varvecr}size$
14.		$\backslash \text{varvece} \backslash \text{varvecn} \backslash \text{varvecd} \backslash \text{varvec}f \backslash \text{varveco} \backslash \text{varvecr}$
15.		$C \leftarrow shuffle(C)$
16.		$r_e \leftarrow 0$
17.		$\backslash \text{varvec}f \backslash \text{varveco} \backslash \text{varvecr} \backslash \text{varveca} \backslash \text{varvecl} \backslash \text{varvecl}c \in C \backslash \text{varvecd} \backslash \text{varveco}$
18.		$c_{original} \leftarrow c$
19.		$c_{target} \leftarrow random(\{1, 2, 3, 4, 5\} exclude c_{original})$
20.		$S_{original} \leftarrow random(D where c_D == c_{original})$
21.		$c_{predict}, S_{modified}, r_t \leftarrow DDPG(S_{original}, c_{target})$
22.		$r_e \leftarrow r_e + r_t$
23.		$\backslash \text{varvec}w \backslash \text{varvec}h \backslash \text{varveci} \backslash \text{varvecl} \backslash \text{varvece}c_{target} \neq c_{predict} \backslash \text{varveco} \backslash \text{varvecr}S_{modified}isvalid \backslash \text{varveca} \backslash \text{varvecn} \backslash \text{varvecd}r_e > -20$
24.		$updateDDPGparameters$
25.		$c_{predict}, S_{modified}, r_t \leftarrow DDPG(S_{original}, c_{target})$
26.		$\backslash \text{varvece} \backslash \text{varvecn} \backslash \text{varvecd} \backslash \text{varvec}w \backslash \text{varvec}h \backslash \text{varveci} \backslash \text{varvecl} \backslash \text{varvece}$
27.		$\backslash \text{varvece} \backslash \text{varvecn} \backslash \text{varvecd} \backslash \text{varvec}f \backslash \text{varveco} \backslash \text{varvecr}$

Results

Daily Average Illuminances of Experimental and Generated Light Schedules

While all α -opic and photopic illuminances had been recorded by SPECCY during our experiment, our results focused on melanopic EDI and photopic illuminance for this analysis as the circadian effects of other α -opic illuminances are still unclear at the point of writing. First, the daily average illuminances of light schedules collected from our experiment and generated from the DDPG algorithm were calculated and plotted (Fig. 3). Mean light timing above 150 lx ($MLiT_{150}$) and total duration above 150 lx (TAT_{150}) were calculated using LightLogR [52] for both experimental and generated light schedules for comparison purposes ($MLiT$ and TAT are light quantification metrics used in past research and their usages were summarized and described by Hartmeyer et al. [53]). DDPG-generated schedules for Class 1 and Class 4 exhibited similar mean light timings above 150 lx ($MLiT_{150}$) and total durations above 150 lx (TAT_{150}) to those of the experimental schedules (Class 1: $\Delta MLiT_{150, MEDl} = 30$ minutes, $\Delta MLiT_{150, Photopic} = 30$ minutes, $\Delta TAT_{150, MEDl} = 1$ hour, $\Delta TAT_{150, MEDl} = 1$ hour; Class 4: $\Delta MLiT_{150, MEDl} = 30$ minutes, $\Delta MLiT_{150, Photopic} = 0$ minute, $\Delta TAT_{150, MEDl} = 1$ hour, $\Delta TAT_{150, Photopic} = 2$ hours). Class 2 and Class 5 generated schedules deviated furthest from the original dataset (Class 2: $\Delta MLiT_{150, MEDl} = 90$ minutes, $\Delta TAT_{150, MEDl} = 5$ hours, $\Delta MLiT_{150, Photopic} = 1$ hour, $\Delta TAT_{150, Photopic} = 4$ hours; Class 5: $\Delta MLiT_{150, MEDl} = 30$ minutes, $\Delta TAT_{150, MEDl} = 8$ hours, $\Delta MLiT_{150, Photopic} = 51.5$ minutes, $\Delta TAT_{150, Photopic} = 9$ hours).

Dim Light Melatonin Onset (DLMO) Simulations from Experimental and Generated Schedules

The DDPG-generated and experimental schedules were passed into the photobiological model by Cheng et al. [28] via their open-source assessment tool. While the model takes in photopic illuminance, activity and sleep-wake status as input, activity counts were not generated by our DDPG algorithm, hence the variable was left out for this investigation. For sleep-wake statuses of the generated schedules, it was assumed that any illuminance values below 1 lx indicated sleep. The generated schedules' DLMOs ($DLMO_{DDPG}$) were floored to their respective hours for visual clarity and plotted in a circular histogram (Fig. 4a). In addition, the simulated DLMOs of the experimental schedules ($DLMO_{Exp}$) were shown on the same plot as a scatter plot with their exact values for better comparisons with their DDPG-generated counterparts.

The simulated $DLMO_{DDPG}$ values of Class 2 deviated significantly from simulated Class 2 $DLMO_{Exp}$ values, which confirmed with the $MLiT_{150}$ and TAT_{150} results. It could also be observed that simulated $DLMO_{DDPG}$ values for Class 2 consisted of a high number of outliers which also featured DLMO values not typically exhibited by individual adhering to diurnal schedules (before 20:00), indicating that schedules generated for this class were most likely not ideal (Fig. 4b). For Class 5, the simulated DLMOs for the generated schedules mostly ranged between those of the two experimental schedules which were situated relatively far from each other. The spread across the simulated $DLMO_{Exp}$ were relatively more concentrated for Classes 1, 3 and 4, with their DDPG counterparts having similar simulated DLMOs.

Discussion

The daily average illuminances of experimental and DDPG-generated schedules were plotted and compared (Fig. 3). Given that Class 1 and Class 2 had the same number of experimental samples, it was expected that the two classes would show relatively similar metric values with those of the experimental schedules. However, this was not the case as only Class 1 DDPG-generated schedules exhibited similar metric values to those of the experimental schedules, whereas generated schedules for Class 2 deviated furthest from the original dataset along with Class 5 which had the smallest experimental sample size. This indicated that the differences between the experimental and DDPG generated schedules were most likely caused by the algorithm itself. This occurrence could be caused by the upper bounds introduced in the DDPG algorithm, where schedules from Class 2 consisted of data that exceeded the limits more frequently, resulting in more discrepancies between experimental and generated schedules. This finding was crucial as it emphasizes the concept of compromise between accuracy and practicality when crafting a DDPG algorithm for light schedule generation as without proper bounds, light schedules could be modified to have extreme values, which may enable achieving targeted sleep-wake classes in theory, but are unfeasible via existing artificial lights.

A photobiological model, implemented as an open-source assessment tool, by Cheng et al. [28] was used to simulate the DLMOs of the DDPG-modified and experimental schedules. For DDPG schedules of Class 5, the difference in simulated DLMO values was mostly caused by the dataset having only two samples as it could be observed that the simulated DLMO values were closely following one of the experiment samples (Fig. 4). Despite the spread of experimental DLMOs being similar for both Class 1 and Class 2 and both the classes having the same sample size, the DLMOs of the DDPG schedules deviated significantly from experimental DLMOs for Class 2. The simulated DLMOs for generated and experimental schedules from Class 3 and Class 4 were comparable resultantly from their larger and more concentrated samples sizes and schedules.

While our exploratory investigation showed that there were merits to the application of reinforcement learning in generating and modifying light schedules for the purpose of sleep-wake adjustment, there were still limitations to the methods and adjustments that should be considered. As the framework was developed upon a dataset formed from a small sample size with a limited population type (young adults aged 18–30 years old), it is important to note that emphasis should be put on the methods incorporated into the frameworks and not interpret the obtained results as concrete proof of associations between light and sleep-wake cycles or circadian rhythms. In addition, the light diets generated in this work should not be treated as solutions for adjustments to the corresponding sleep-wake classes but should be analyzed as hypotheticals in relation to the data used to generate them, similarly to the discussions described. To extend the introduced framework to deployable states, extensive work is still required in data collection as well as validation. Both data collection and validation should be carried out for a large population with varying individuals, ranging from healthy children to elderly with impacted circadian rhythms to determine modifications required to refine the frameworks, for example whether accurate classifications of sleep-wake classes is possible with a singular model, or should there be multiple models tailored towards targeted populations. A larger dataset would also remove or reduce the necessity of data augmentation and potentially increase cluster qualities of obtained sleep-wake classes.

Light diets generated from the framework need to be validated to establish whether they are effective in adjusting sleep-wake timings towards targeted phases and their efficiencies in doing so. In research carried out by Walch et al. [10] and Christensen et al. [20], a smartphone app called 'Entrain' was used to collect large-scale sleep-wake and light exposure data across the world. If a similar app could be developed with the integration of wearable light dosimeters such as SPECCY, the light diet algorithms introduced in this study could be vastly refined and improved with a significantly larger dataset, given that a viable method is deployed to ensure user compliance.

The apparent difference between the work presented in this exploratory investigation and similar circadian research conducted in the past is the usage of sleep-wake parameters in place of circadian proxies such as DLMO or CBT_{min}. This difference inherently implies that the introduced framework operates within a boundary that is more dominantly affected by behavioral factors rather than circadian zeitgebers. While this means that no causal relationship could be conclusively obtained between daily light exposure and circadian rhythms during the analysis of the data collected, it enables the framework to be scaled and deployed to ecological settings. Assessments of circadian proxies require high-levels of precision and execution in comparison to recording subjective or objective sleep-wake parameters. In addition, daily sleep-wake behaviors of individuals in the modern age are often dictated by sociological factors rather than biological processes. Hence, there is merit to utilizations of sleep-wake parameters in our framework where it ideally would progress towards a generalized application that generates light diets capable of assisting in sleep-wake adjustments.

Another crucial aspect that is often emphasized in the field of circadian research is interindividual differences. Investigations and studies have been carried out extensively to refine circadian models to accommodate interindividual differences which include variables from light sensitivity [54, 55] to retinal size variations [56, 57]. It is necessary to accurately quantify these as it has been shown that large variations exist between individuals, therefore a physiologically informed model of the circadian system must account for these factors to be reliable, accurate and generalizable. However, individual-specific responses to light are, again, not easily assessable via ecological data collection processes. This does not necessarily imply that models or frameworks derived from ecological sleep-wake data would not be on par with sophisticated physiological models, but rather different approaches need to be explored and tested to be used in place of these variables. For example, the changes in sleep-wake timings in response to modifications in daily light exposure could be quantified as interindividual responses to light. Additionally, non-light factors such as work schedules, demographic information, daytime naps or sleep irregularities, caffeine intake, and pre-bedtime activities all contribute to interindividual differences in daily sleep-wake timings. While this may result in the proposed frameworks deviating from being purely light-based methods, this direction of advancement should be inevitable for them to mature into efficient and effective light diet generators for sleep-wake adjustments meant for the general population.

Conclusion

In this work, we developed a light schedule generation framework comprising of a method novel to the application: reinforcement learning, specifically the DDPG algorithm. We first defined different sleep-wake classes based on experimental sleep-wake parameters, namely MS, TST, SE, SO and WT. Then, we developed a classification CNN that takes in a 6-day hourly photopic and α -opic illuminance schedule and predicts the resulting sleep-wake class. This CNN is used in tandem with our DDPG algorithm, newly tailored for the purpose of light schedule modification. While there were clear improvements that could still be achieved via future work, the framework could be easily scaled to larger datasets collected in the future via substitutions of clustering methods or have additional variable components considered as machine learning methods were utilized. The light diets generated also encompass all α -opic and photopic illuminances, ensuring that the framework could be updated when our knowledge of the effects of the remaining illuminances on the human circadian rhythm mature to more advanced levels.

Declarations

Data Availability

Access to data collected in this research can be requested by emailing the corresponding author subject to approval from MUHREC.

Author Contributions

M.S., J.M.Y.L., S.N.L., Y.K., and V.K. contributed to the conception and design of the study. M.S. and Y.K. contributed to the acquisition of data. M.S., J.M.Y.L., S.N.L., and V.K. contributed to the analysis and interpretation of data. M.S. drafted the manuscript document, and all authors reviewed the manuscript.

Acknowledgements

We acknowledge the Monash Malaysia Intelligent Lighting Laboratory (MMILL) and the Advanced Computing Platform (ACP) at Monash University Malaysia for providing the computational resources and equipment essential for this research.

Funding

This research was supported by the Fundamental Research Grant Scheme (FRGS) under Grant No. FRGS/1/2020/SKK06/MUSM/02/1 from the Ministry of Higher Education (MOHE), Malaysia.

References

1. L. Fuhr, M. Abreu, P. Pett, and A. Relógio, "Circadian systems biology: When time matters," *Comput Struct Biotechnol J*, vol. 13, no. C, pp. 417-426, 2015, doi: 10.1016/j.csbj.2015.07.001.

2. P. L. Lowrey and J. S. Takahashi, "Chapter 6 - Genetics of Circadian Rhythms in Mammalian Model Organisms," in *Advances in Genetics*, vol. 74, S. Brody Ed.: Academic Press, 2011, pp. 175-230, doi: 10.1016/B978-0-12-387690-4.00006-4.
3. S. J. Crowley, T. A. Molina, and H. J. Burgess, "A week in the life of full-time office workers: Work day and weekend light exposure in summer and winter," *Appl Ergon*, vol. 46, pp. 193-200, 2015, doi: 10.1016/j.apergo.2014.08.006.
4. E. J. Wams et al., "Linking light exposure and subsequent sleep: A field polysomnography study in humans," *Sleep*, vol. 40, no. 12, 2017, doi: 10.1093/sleep/zsx165.
5. M. G. Figueiro, C. Jarboe, and L. Sahin, "The sleep maths: A strong correlation between more daytime light and better night-time sleep," *Lighting research & technology (London, England : 2001)*, vol. 53, no. 5, pp. 423-435, 2021, doi: 10.1177/14771535211005832.
6. S. W. Cain et al., "Evening home lighting adversely impacts the circadian system and sleep," *Sci Rep*, vol. 10, no. 1, pp. 19110-19110, 2020, doi: 10.1038/s41598-020-75622-4.
7. S. L. Chellappa, "Individual differences in light sensitivity affect sleep and circadian rhythms," *Sleep*, vol. 44, no. 2, p. 1, 2021, doi: 10.1093/sleep/zsaa214.
8. M. T. Aguilar-Carrasco, A. I. Álvarez-López, P. Bustamante, A. I. Álvarez Ríos, and S. Domínguez-Amarillo, "The impact of light adjustment versus tuneable white mode on ICU staff's circadian rhythms," *Building research and information : the international journal of research, development and demonstration*, vol. 53, no. 1-2, pp. 120-138, 2025, doi: 10.1080/09613218.2024.2394479.
9. H. Jo, H. R. Park, S. J. Choi, S.-Y. Lee, S. J. Kim, and E. Y. Joo, "Effects of Organic Light-Emitting Diodes on Circadian Rhythm and Sleep," *Psychiatry investigation*, vol. 18, no. 5, pp. 471-477, 2021, doi: 10.30773/pi.2020.0348.
10. O. J. Walch, A. Cochran, and D. B. Forger, "A global quantification of "normal" sleep schedules using smartphone data," *Science Advances*, vol. 2, no. 5, p. e1501705, 2016, doi: doi:10.1126/sciadv.1501705.
11. S. T. Peeters, K. C. H. J. Smolders, and Y. A. W. de Kort, "What you set is (not) what you get: How a light intervention in the field translates to personal light exposure," *Building and environment*, vol. 185, p. 107288, 2020, doi: 10.1016/j.buildenv.2020.107288.
12. C. Blume, C. Garbazza, and M. Spitschan, "Effects of light on human circadian rhythms, sleep and mood," *Somnologie : Schlafforschung und Schlafmedizin = Somnology : sleep research and sleep medicine*, vol. 23, no. 3, pp. 147-156, 2019, doi: 10.1007/s11818-019-00215-x.
13. L. Tähkämö, T. Partonen, and A.-K. Pesonen, "Systematic review of light exposure impact on human circadian rhythm," *Chronobiology international*, vol. 36, no. 2, pp. 151-170, 2019, doi: 10.1080/07420528.2018.1527773.
14. J. Hesse, D. Malhan, M. Yalcin, O. Aboumanify, A. Basti, and A. Relogio, "An Optimal Time for Treatment-Predicting Circadian Time by Machine Learning and Mathematical Modelling," *Cancers*, vol. 12, no. 11, p. 1, 2020, doi: 10.3390/cancers12113103.
15. J. E. Stone, S. Postnova, T. L. Sletten, S. M. W. Rajaratnam, and A. J. K. Phillips, "Computational approaches for individual circadian phase prediction in field settings," *Current opinion in systems biology*, vol. 22, pp. 39-51, 2020, doi: 10.1016/j.coisb.2020.07.011.
16. J. M. Murray et al., "Light-based methods for predicting circadian phase in delayed sleep–wake phase disorder," *Scientific reports*, vol. 11, no. 1, pp. 10878-10878, 2021, doi: 10.1038/s41598-021-89924-8.
17. Y. Huang et al., "Predicting circadian phase across populations: a comparison of mathematical models and wearable devices," *Sleep*, vol. 44, no. 10, p. 1, 2021, doi: 10.1093/sleep/zsab126.
18. K. Serkh and D. B. Forger, "Optimal schedules of light exposure for rapidly correcting circadian misalignment," *PLoS Comput Biol*, vol. 10, no. 4, pp. e1003523-e1003523, 2014, doi: 10.1371/journal.pcbi.1003523.
19. A. Mohamed, V. Kalavally, Y. J. Khong, C. P. Tan, S. W. Cain, and A. J. K. Phillips, "A simulation-based framework for multi-individual circadian re-entrainment using a single light schedule," *Building and environment*, vol. 229, p. 109943, 2023, doi: 10.1016/j.buildenv.2022.109943.
20. S. Christensen, Y. Huang, O. J. Walch, and D. B. Forger, "Optimal adjustment of the human circadian clock in the real world," *PLoS Comput Biol*, vol. 16, no. 12, pp. e1008445-e1008445, 2020, doi: 10.1371/journal.pcbi.1008445.
21. A. Guyett et al., "A circadian-informed lighting intervention accelerates circadian adjustment to a night work schedule in a submarine lighting environment," *Sleep (New York, N.Y.)*, vol. 47, no. 11, p. 1, 2024, doi: 10.1093/sleep/zsae146.
22. M. Xu, C. Papatsimpa, L. Schlangen, and J.-P. Linnartz, "Improving adjustment to daylight saving time transitions with light," *Scientific reports*, vol. 14, no. 1, pp. 15001-13, 2024, doi: 10.1038/s41598-024-65705-x.
23. C. You, J. Lu, D. Filev, and P. Tsiotras, "Advanced planning for autonomous vehicles using reinforcement learning and deep inverse reinforcement learning," *Robotics and autonomous systems*, vol. 114, pp. 1-18, 2019, doi: 10.1016/j.robot.2019.01.003.
24. F. Ahmad, X. Zhang, Z. Tang, F. Sabah, M. Azam, and R. Sarwar, "Deep deterministic policy gradients with a self-adaptive reward mechanism for image retrieval: Deep deterministic policy gradients," *The Journal of supercomputing*, vol. 81, no. 1, 2025, doi: 10.1007/s11227-024-06764-9.
25. Y. Matsuo et al., "Deep learning, reinforcement learning, and world models," *Neural networks*, vol. 152, pp. 267-275, 2022, doi: 10.1016/j.neunet.2022.03.037.
26. M. Panzer and B. Bender, "Deep reinforcement learning in production systems: a systematic literature review," *International journal of production research*, vol. 60, no. 13, pp. 4316-4341, 2022, doi: 10.1080/00207543.2021.1973138.
27. S. J. Crowley, C. Acebo, G. Fallone, and M. A. Carskadon, "Estimating Dim Light Melatonin Onset (DLMO) Phase in Adolescents Using Summer or School-Year Sleep/Wake Schedules," *Sleep (New York, N.Y.)*, vol. 29, no. 12, pp. 1632-1641, 2006, doi: 10.1093/sleep/29.12.1632.
28. P. Cheng et al., "Predicting circadian misalignment with wearable technology: validation of wrist-worn actigraphy and photometry in night shift workers," *Sleep (New York, N.Y.)*, vol. 44, no. 2, p. 1, 2021, doi: 10.1093/sleep/zsaa180.
29. A. W. McHill et al., "Robust stability of melatonin circadian phase, sleep metrics, and chronotype across months in young adults living in real-world settings," *Journal of pineal research*, vol. 70, no. 3, pp. e12720-n/a, 2021, doi: 10.1111/jpi.12720.

30. D.-J. Dijk and J. F. Duffy, "Novel Approaches for Assessing Circadian Rhythmicity in Humans: A Review," *J Biol Rhythms*, vol. 35, no. 5, pp. 421-438, 2020, doi: 10.1177/0748730420940483.
31. A. Khademi, Y. El-Manzalawy, L. Master, O. M. Buxton, and V. G. Honavar, "Personalized Sleep Parameters Estimation from Actigraphy: A Machine Learning Approach," *Nat Sci Sleep*, vol. 11, pp. 387-399, 2019, doi: 10.2147/NSS.S220716.
32. B. Rodriguez-Morilla *et al.*, "Application of Machine Learning Methods to Ambulatory Circadian Monitoring (ACM) for Discriminating Sleep and Circadian Disorders," *Front Neurosci*, vol. 13, pp. 1318-1318, 2019, doi: 10.3389/fnins.2019.01318.
33. J. E. Stone *et al.*, "Generalizability of A Neural Network Model for Circadian Phase Prediction in Real-World Conditions," *Sci Rep*, vol. 9, no. 1, pp. 11001-17, 2019, doi: 10.1038/s41598-019-47311-4.
34. T. M. Brown, "Melanopic illuminance defines the magnitude of human circadian light responses under a wide range of conditions," *J Pineal Res*, vol. 69, no. 1, pp. e12655-n/a, 2020, doi: 10.1111/jpi.12655.
35. R. J. Lucas *et al.*, "Measuring and using light in the melanopsin age," *Trends Neurosci*, vol. 37, no. 1, pp. 1-9, 2013, doi: 10.1016/j.tins.2013.10.004.
36. A. K. e. Kurtz, "Hierarchical Clustering Schemes," *Psychometrika*, vol. 32, no. 3, pp. 241-254, 1967, doi: 10.1007/BF02289588.
37. P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53-65, 1987/11/01/ 1987, doi: [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
38. T. Caliński and J. and Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1-27, 1974/01/01 1974, doi: 10.1080/03610927408827101.
39. D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224-227, 1979, doi: 10.1109/TPAMI.1979.4766909.
40. L. S. Brown *et al.*, "A classification approach to estimating human circadian phase under circadian alignment from actigraphy and photometry data," *J Pineal Res*, vol. 71, no. 1, pp. e12745-n/a, 2021, doi: 10.1111/jpi.12745.
41. P. Yang and X. Zhang, "A Dual-Branch Fusion of a Graph Convolutional Network and a Convolutional Neural Network for Hyperspectral Image Classification," *Sensors (Basel, Switzerland)*, vol. 24, no. 14, p. 4760, 2024, doi: 10.3390/s24144760.
42. D. S. Verma, M. Dafadar, J. K. Mishra, A. Kumar, S. Mahato, and H. Alexander, "AI-Enable Rice Image Classification Using Hybrid Convolutional Neural Network Models," *International journal of intelligent systems*, vol. 2025, no. 1, 2025, doi: 10.1155/int/5571940.
43. R. J. Pally and S. Samadi, "Application of image processing and convolutional neural networks for flood image classification and semantic segmentation," *Environmental modelling & software : with environment data news*, vol. 148, p. 105285, 2022, doi: 10.1016/j.envsoft.2021.105285.
44. W. K. Cho *et al.*, "Diagnostic Accuracies of Laryngeal Diseases Using a Convolutional Neural Network-Based Image Classification System," *The Laryngoscope*, vol. 131, no. 11, pp. 2558-2566, 2021, doi: 10.1002/lary.29595.
45. X. Cao, J. Yao, Z. Xu, and D. Meng, "Hyperspectral Image Classification With Convolutional Neural Network and Active Learning," *IEEE transactions on geoscience and remote sensing*, vol. 58, no. 7, pp. 4604-4616, 2020, doi: 10.1109/TGRS.2020.2964627.
46. M. Hamouda and M. S. Bouhlel, "Modified Convolutional Neural Networks Architecture for Hyperspectral Image Classification (Extra-Convolutional Neural Networks)," *IET image processing*, vol. 19, no. 1, 2025, doi: 10.1049/ipr2.12169.
47. H. Li, X. Xiong, C. Liu, Y. Ma, S. Zeng, and Y. Li, "SFFNet: Staged Feature Fusion Network of Connecting Convolutional Neural Networks and Graph Convolutional Neural Networks for Hyperspectral Image Classification," *Applied sciences*, vol. 14, no. 6, p. 2327, 2024, doi: 10.3390/app14062327.
48. O. L. P. Hansen *et al.*, "Species-level image classification with convolutional neural network enables insect identification from habitus images," *Ecology and evolution*, vol. 10, no. 2, pp. 737-747, 2020, doi: 10.1002/ece3.5921.
49. Y. Sun, B. Xue, M. Zhang, and G. G. Yen, "Evolving Deep Convolutional Neural Networks for Image Classification," *IEEE transactions on evolutionary computation*, vol. 24, no. 2, pp. 394-407, 2020, doi: 10.1109/TEVC.2019.2916183.
50. T. P. Lillicrap *et al.*, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.
51. G. E. Uhlenbeck and L. S. Ornstein, "On the theory of the Brownian motion," *Physical review*, vol. 36, no. 5, p. 823, 1930, doi: 10.1103/PhysRev.36.823.
52. J. Zauner, S. Hartmeyer, and M. Spitschan, "LightLogR: Reproducible analysis of personal light exposure data," (in eng), *J Open Source Softw*, vol. 10, no. 107, p. 7601, Mar 13 2025, doi: 10.21105/joss.07601.
53. S. L. Hartmeyer and M. Andersen, "Towards a framework for light-dosimetry studies: Quantification metrics," *Lighting research & technology (London, England : 2001)*, vol. 56, no. 4, pp. 337-365, 2024, doi: 10.1177/14771535231170500.
54. J. E. Stone *et al.*, "The Role of Light Sensitivity and Intrinsic Circadian Period in Predicting Individual Circadian Timing," *J Biol Rhythms*, vol. 35, no. 6, pp. 628-640, 2020, doi: 10.1177/0748730420962598.
55. A. J. K. Phillips *et al.*, "High sensitivity and interindividual variability in the response of the human circadian system to evening light," *Proc Natl Acad Sci U S A*, vol. 116, no. 24, pp. 12019-12024, 2019, doi: 10.1073/pnas.1901824116.
56. S. Higuchi *et al.*, "Inter-individual difference in pupil size correlates to suppression of melatonin by exposure to light," *Neuroscience letters*, vol. 440, no. 1, pp. 23-26, 2008, doi: 10.1016/j.neulet.2008.05.037.
57. L. A. Watson *et al.*, "Increased sensitivity of the circadian system to light in delayed sleep-wake phase disorder," *The Journal of physiology*, vol. 596, no. 24, pp. 6249-6261, 2018, doi: 10.1113/JP275917.

Figures

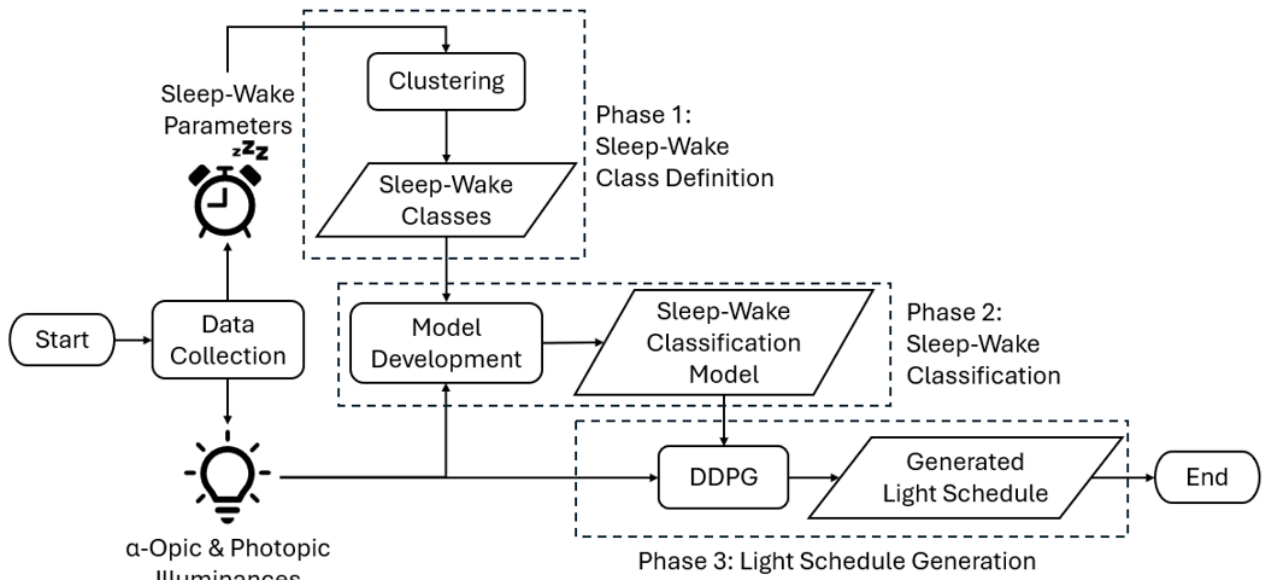


Figure 1

Development framework for light schedule generation using deep deterministic gradient policy (DDPG) reinforcement learning (RL).

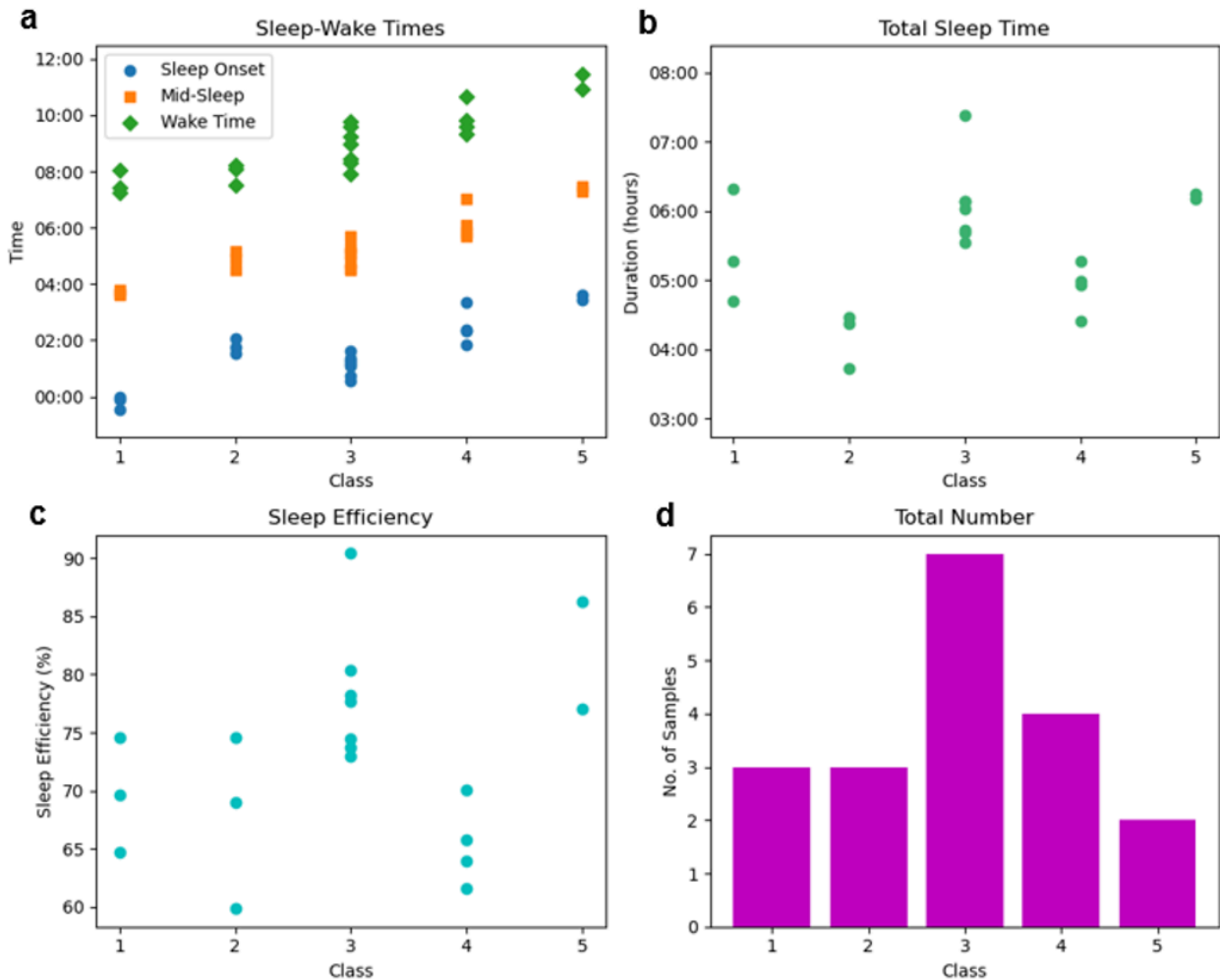


Figure 2

The (a) MS, WT, SO, (b) TST, (c) SE, and (d) number of participants for each sleep-wake class defined via hierarchical clustering.

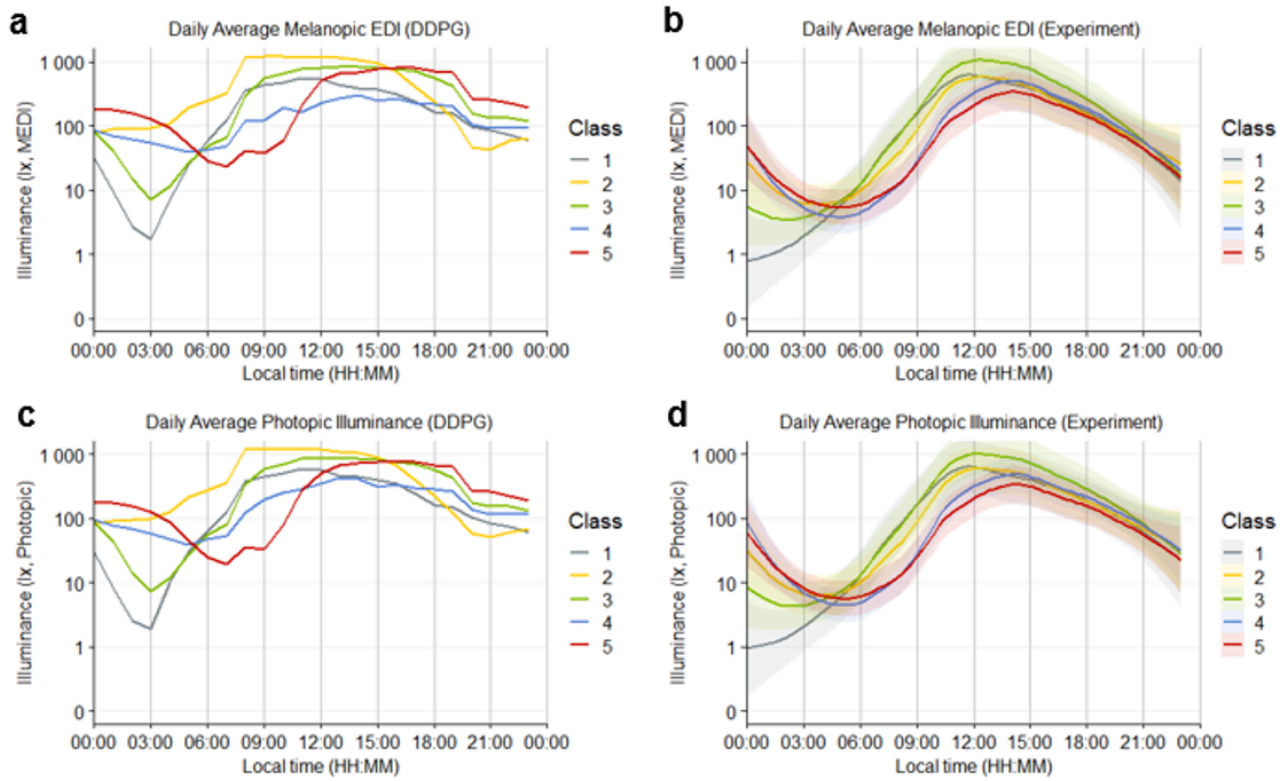


Figure 3

Line plots of daily average (a) melanopic EDI and (c) photopic illuminance of DDPG generated schedules. Smoothen line plots of daily average (b) melanopic EDI and (d) photopic illuminance of schedules collected from the experiment.

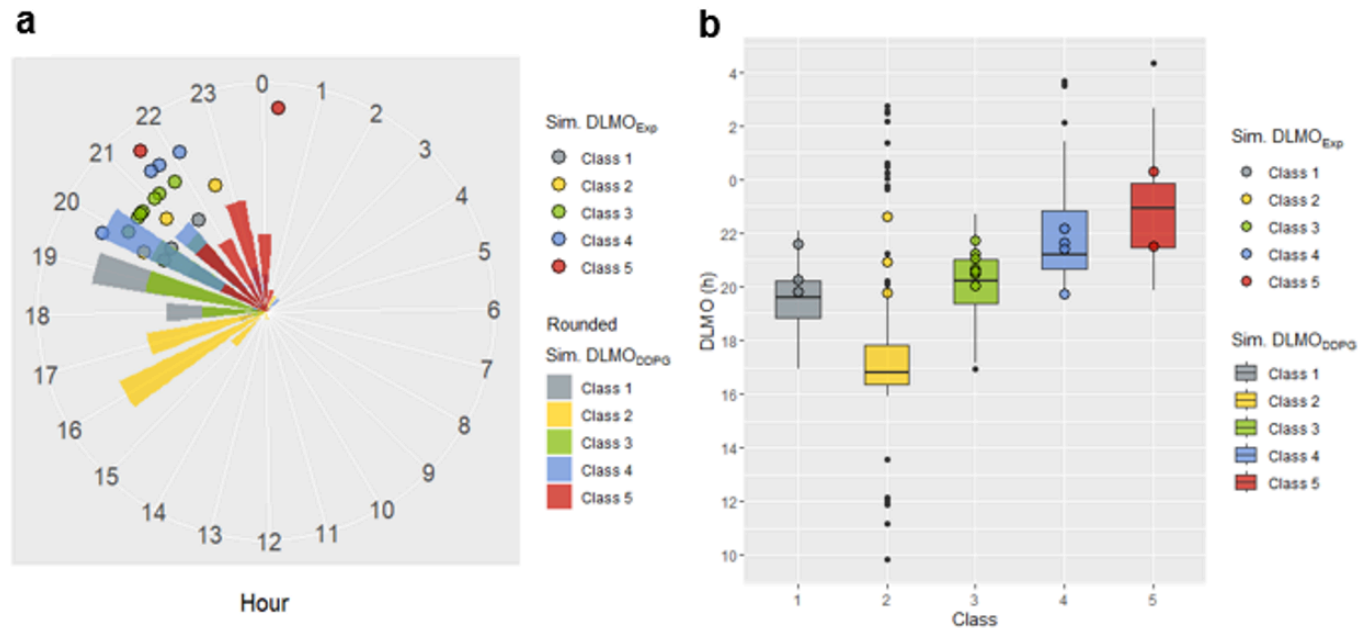


Figure 4

(a) Simulated DLMO values represented using a circular histogram (DLMO_{DDPG}) for the 100 DDPG-generated light schedules and scatter plots (DLMO_{Exp}) for the schedules obtained from the experiment. Simulated DLMO_{DDPG} values are floored to their hours. Amplitudes or “y” values of simulated DLMO_{Exp} do not have any implications and are solely for clearer visualization. (b) Box (DLMO_{DDPG}) and scatter plots (DLMO_{Exp}) of simulated DLMO values of the light schedules.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterialSR.docx](#)