aerospace

MDPI

Article

A Dynamic Bayesian Network-Based Prediction Method for Cognitive Ability of Long-Duration Operators

Si Li [1,2], Xiaoyi Zhou [1,2], Yang Yu [1,2], Pei Li [3,*], Liping Pang [4], Liang Zhang [1,2] and Baiqian Hui [4]

[1] China Institute of Marine Technology and Economy, Beijing 100081, China
[2] National Key Laboratory of Human Factors Engineering, China Astronaut Research and Training Center, Beijing 100094, China
[3] College of Aeronautical Engineering, Civil Aviation University of China, Tianjin 300300, China
[4] School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China
* Correspondence: p_li@cauc.edu.cn; Tel.: +86-15910594100

Abstract

The accurate prediction of cognitive ability variation over time is critical for operator performance and safety management. To address the time-dependent degradation of the cognitive ability of long-duration operators working in confined, shift-based and otherwise demanding environments, this paper proposes a prediction method based on a Dynamic Bayesian Network (DBN). Four cause-layer factors (confinement duration, shift cycle, circadian rhythm and sleep, and working environment) are first analyzed, and five result-layer indicators (perceptual stability, attention-and-vigilance stability, working-memory stability, reasoning-and-decision stability, and executive-function stability) are selected to characterize the cognitive ability of long-duration operators. After the indicators are quantified and normalized, a Dynamic Bayesian Network (DBN) is constructed under a first-order Markov assumption to describe the temporal evolution of cognitive ability. The inference process consists of two stages: state estimation using current observations, followed by prediction through the time-dependent transition. A cognitive ability level, denoted as $Pr \in [0, 1]$, is treated as the target variable. A conditional-probability correction factor is further introduced to capture inter-operator heterogeneity, thereby establishing a prediction model for cognitive ability level. A 288-h simulated-task experiment was designed, and cognitive-test data were collected from 15 participants, who were divided into three groups (A1~A3). The model prediction accuracy was evaluated at both the individual and group levels. The results suggest that the prediction errors remain generally low. Furthermore, the predicted cognitive ability exhibits a continuous decline with task time, with partial recovery occurring during rest periods. The results show that the proposed DBN generally outperforms the Static Bayesian Network (SBN) baseline at both the individual and group levels. In the group-level evaluation, the Root Mean Square Error (RMSE) values for all three groups are below 0.011, the Mean Percentage Error (MPE) values are within −0.84%, and the Maximum Cross-Correlation Coefficient (MCCC) values exceed 0.759, indicating good numerical accuracy and stability. The DBN also outperforms the mean predictor, confirming that it captures meaningful temporal variation rather than merely fitting the overall mean level. At the individual level, the DBN generally achieves lower RMSE and MPE and higher MCCC than the SBN across the 15 participants. These results demonstrate the effectiveness of the proposed DBN framework for the dynamic monitoring and early warning of cognitive ability in long-duration operations.

Check for updates

Academic Editor: Miroslav Kelemen

Received: 20 August 2026
Revised: 1 October 2026
Accepted: 3 October 2026
Published: 7 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: Dynamic Bayesian Network; cognitive ability prediction; long-duration operation; causal analysis

1. Introduction

In long-duration confined missions such as manned spaceflight, space stations, and deep-sea exploration, maintaining operational safety and efficiency relies heavily on human cognitive performance [1]. Personnel operating in such extreme, isolated, and restricted environments often suffer from circadian disruption, fatigue, and intense psychological stress, which can lead to significant cognitive decline across attention, working memory, and decision-making capabilities [2,3]. In critical operations, even a minor cognitive failure can trigger catastrophic consequences [4–6].

Despite its importance, accurately predicting dynamic cognitive performance in confined environments remains a major challenge. Existing assessment methods mostly rely on static evaluations of short-term test results or population-level regression models, and fail to capture the continuous, dynamic variations in cognitive state under complex multi-factor interactions (e.g., circadian rhythm misalignment, isolated confinement, and shift work) [7]. Moreover, most universal models overlook significant individual differences in stress tolerance and fatigue sensitivity [8]. Although DBNs have shown great promise in modeling temporal dynamics and handling multi-source uncertainties [9–11], a systematic DBN-based framework tailored for continuous cognitive prediction under extreme confined conditions is still lacking.

To address these limitations, this paper aims to develop a novel, dynamic, and individualized cognitive ability prediction framework based on Dynamic Bayesian Networks for long-duration confined missions. The primary objective is to continuously monitor and forecast the multi-dimensional cognitive status of operators while accommodating individual variability and environmental stressors.

This paper proposes a DBN-based method for predicting the cognitive ability level, denoted as $Pr \in [0, 1]$, of long-duration confined tasks. A causal analysis is first conducted to construct a variation structure comprising four causal-layer factors and five result-layer indicators; multi-source heterogeneous indicators are then quantified and normalized to the unit interval to enable probabilistic inference within the DBN framework.

The main contributions of this study are threefold:

(1) Causal network architecture: A multi-layer causal architecture is established to link four cause-layer factors to five result-layer cognitive indicators, which is then extended into a Dynamic Bayesian Network.

(2) Dynamic modeling: A DBN-based quantitative model is introduced to capture the temporal evolution of cognitive ability across duty periods, enabling sequential prediction of cognitive state.

(3) Experimental evaluation: The proposed model is evaluated using a 288-h simulated extreme-mission experiment with 15 participants, showing improved accuracy and prediction performance over the SBN baseline.

The rest of this paper is organized as follows: Section 2 provides the theoretical background and reviews related work on cognitive evaluation and prediction methods. Section 3 details the proposed DBN-based prediction method. Section 4 presents the experimental design, data collection, and model evaluation. Section 5 presents the results and discussion, including the prediction accuracy evaluation and comparison with the SBN baseline. Finally, Section 6 concludes the paper and outlines future research directions.

2. Theoretical Background

2.1. Cognitive Performance and Assessment in Confined Missions

Cognitive ability is a multi-dimensional construct comprising perception, attention, working memory, reasoning, and executive control. In long-duration confined environments, extended operation severely impairs these cognitive dimensions, leading to prolonged reaction times, elevated error rates, and degraded decision quality [4]. For example, studies on ISS astronauts during 6-month missions revealed marked slowdowns in processing speed, visual working memory, and sustained attention during early flight phases [5]. Similarly, dense shift work in low-Earth-orbit missions has been strongly linked to decreased vigilance and working-memory efficiency [6].

Standard psychological battery tests evaluate these core dimensions through specific tasks:

- **Perceptual stability**: Assessed via speed-perception or visual contrast-sensitivity tests.
- **Attention and vigilance**: Evaluated using the Psychomotor Vigilance Task (PVT) or continuous performance tasks [12].
- **Working memory**: Measured using verbal/spatial n-back tasks or operation-span tasks [13].
- **Reasoning and decision-making**: Gauged via the Iowa Gambling Task or the Balloon Analogue Risk Task (BART) [14,15].
- **Executive function**: Examined through the color–word Stroop test, Go/No-Go tasks, or the Wisconsin Card Sorting Test (WCST) [16].

These standardized tests yield performance metrics (e.g., reaction time, accuracy, omission rate) that serve as foundational quantitative indicators for cognitive state assessment.

2.2. Current Cognitive Prediction Methods and Limitations

Existing approaches to cognitive assessment and prediction from behavioral and physiological data fall into three main categories: statistical regression modeling, physiological-signal threshold judgment, and task-performance analysis.

Statistical regression methods map performance metrics to external stressors (e.g., duty duration, sleep loss) using linear or logistic regression [17]. Threshold-based methods utilize physiological features, such as heart rate variability, EEG and skin conductance, to infer fatigue states when pre-set empirical limits are breached. While computationally light and suitable for simple engineering applications, these traditional methods suffer from several critical shortcomings:

- **Static assessment**: They evaluate cognitive states at discrete time points, ignoring continuous temporal trajectories and past state dependencies.
- **Independence assumptions**: They treat influencing factors (e.g., sleep deprivation [18], ambient noise, shift phase) as independent, failing to model multi-factor interactions and feedback loops.
- **Overlooking individual differences**: Most existing models aim for a universal population-level fit, ignoring individual variations in stress tolerance and recovery speed [8].
- **Lack of explainability and robustness**: Conventional regression models lack explicit causal structures to explain why cognitive performance degrades, and their accuracy drops significantly when dealing with skewed data or missing values.

Machine learning methods, such as support vector machines and neural networks, can model time-series patterns but often lack interpretability. Regarding Bayesian network-based approaches (including SBNs and DBNs) and data-driven machine learning methods, SBNs can represent causal relationships but are time-invariant, while Dynamic Bayesian Networks extend them with temporal transitions. DBNs combine probabilistic graphical

models with temporal dynamics, making them uniquely suitable for overcoming the aforementioned challenges. By replicating static Bayesian structures across time slices and introducing temporal transition probabilities, DBNs naturally model continuous temporal processes and causal dependencies.

Recent applications show the strength of DBNs in complex human factor modeling. For example, DBNs have been used to dynamically estimate operator fatigue in plateau deep-mine environments using physiological and contextual inputs [9], as well as to score subjective cognitive decline changes over time [10]. Compared to static models, DBNs provide a flexible framework capable of fusing multi-modal heterogeneous data (environmental, physiological, and behavioral), parameterizing individual differences via conditional-probability distributions, and performing probabilistic inference even under incomplete observation. Building upon these advantages, this study adapts DBN modeling to the complex dynamics of long-duration confined missions.

Table 1 summarizes the key characteristics and limitations of existing methods. As shown, most existing approaches satisfy only some of the requirements for long-duration confined missions. This motivates the DBN-based framework proposed in this study.

Table 1. Comparison of cognitive prediction methods.

Method	Temporal Modeling	Causal Structure	Limitation for Confined Missions
Statistical regression [3,17]	Discrete time points	No (independence assumption)	Cannot model temporal evolution or multi-factor interactions
Physiological threshold [8,18]	Real-time monitoring	No (symptom-based)	Lacks causal explanation and predictive capability
Task-performance analysis [12–16]	Scheduled tests	No (post hoc interpretation)	Intrusive, misses continuous dynamics
SBN [11]	Time-invariant	Yes (explicit causal structure)	Cannot model temporal evolution
Existing DBNs [9,10]	Temporal transitions	Yes (explicit causal structure)	Industrial/clinical scenarios; no multi-dimensional cognitive outcomes
Machine learning [19,20]	Time-series-capable	No (black-box)	Often requires large datasets, may not generalize
Proposed DBN (this study)	Per duty period	Yes (explicit causal structure)	—

2.3. Positioning of This Study

As summarized in Table 1, the existing methods each address some but not all requirements for cognitive monitoring in long-duration confined missions. First, regarding temporal modeling, regression, threshold, and task-analysis methods evaluate cognitive state only at discrete time points, while machine learning time-series models can capture temporal patterns but often lack interpretability and require large training datasets. None of these methods provides an interpretable, continuously updated prediction across a mission. Second, regarding causal structure, most methods do not explicitly link environmental and physiological stressors to cognitive outcomes. Static Bayesian Networks offer causal structures but cannot model temporal dynamics, and existing DBN applications focus on industrial or clinical settings without systematically mapping multi-dimensional stressors to specific cognitive sub-domains. Third, regarding group-level adaptation, prior methods

either ignore group differences, require per-individual training data, or do not account for systematic differences across shift schedules.

To the best of our knowledge, no prior study has developed a DBN-based framework that simultaneously provides: (a) cognitive prediction across duty periods in confined missions, (b) an explicit causal structure mapping multi-dimensional stressors to multi-dimensional cognitive outcomes, and (c) conditional-probability correction for different shift schedules. This study addresses these gaps by proposing a tailored DBN-based framework, validated in a 288-h simulated extreme-mission experiment with 15 participants across three shift schedules.

3. Methods

3.1. Cognitive Ability Variations in Long-Duration Operators

3.1.1. Key Cognitive Abilities

Cognitive ability is the basic capacity of long-duration operators to perceive environmental changes and acquire task-relevant information during operation. Information processing involves perceptual processing, attention, memory, decision-making, and response execution. Cognitive ability is the joint result of multiple cause-layer factors. The following five indicators are used to reflect the operators' key cognitive abilities in this study.

(1) Perceptual stability: The ability to perceive external stimuli through vision, hearing, touch, etc., during operation; it is reflected by evaluating the sensitivity of sensory channels and the constancy of perception, indicating whether perceptual quality remains stable under environmental change.

(2) Attention-and-vigilance stability: The ability to maintain concentrated attention and stay alert to key signals during prolonged operation; it measures the persistence, selectivity and anti-interference of attention and the fluctuation of vigilance.

(3) Working-memory stability: The ability to temporarily store, process and update information during a task; its stability reflects how well working-memory capacity and operational efficiency are maintained as time and task load change.

(4) Reasoning-and-decision stability: The degree to which the ability to perform logical reasoning, option evaluation, and decision judgment in complex or uncertain situations remains stable as task time increases; it examines the changes in decision speed, accuracy and strategy consistency.

(5) Executive-function stability: The control ability to plan, monitor, regulate and flexibly switch one's own cognitive processes; task management, error monitoring, inhibitory control and cognitive flexibility are all affected by the executive-function state.

3.1.2. Causal Analysis of Cognitive Ability Variation

A Bayesian network-based prediction method for the cognitive ability of long-duration operators can be established by analyzing the cause-layer influencing factors and result-layer assessment indicators. To build a model that describes the variation in key cognitive abilities of long-duration operators, it is first necessary to identify the factors influencing cognitive performance and the corresponding assessment indicators. Based on previous studies in this field, the causal relationship between influencing factors and outcome variables during cognitive ability variation is established. This relationship provides a basis for the Bayesian network-based prediction method. Accordingly, this section defines two categories of indicators: cause-layer influencing factors and result-layer cognitive ability assessment indicators.

(1) Cause-layer influencing-factor indicators

Based on an analysis of the working environment, shift cycle, and personal state of long-duration operators during their tasks, the following indicators for cognitive decline-influencing factors are determined.

(a) Confinement duration: Operators perform tasks for long periods in confined and isolated environments. Compared with ordinary workplaces, the psychological and physiological stress of prolonged confined operation challenges daily life and working state, thereby affecting the level of cognitive ability. Accordingly, the number of confined-operation days (or hours) is selected as an influencing-factor indicator.

(b) Shift cycle: Long-duration operators work in groups across different time periods covering the whole day, and the arrangement of the shift cycle affects the level of cognitive ability. Therefore, two indicators are selected based on the shift-cycle arrangement: shift timing and daily cumulative work hours.

(c) Circadian rhythm and sleep: An operator's own biological rhythm and sleep quality directly affect alertness, emotional state, fatigue and cognitive ability. Therefore, two indicators are selected as influencing factors: pre-shift sleep duration and circadian arousal level [21].

(d) Working environment: The working environment of long-duration operators differs from ordinary workplaces in illumination, noise level, gas concentration, etc. Their psychological and physiological states are affected by these environmental factors [22]. Therefore, the noise level, CO_2 concentration, and the illuminance and color temperature of lighting in the working environment are selected as influencing-factor indicators.

(2) Result-layer cognitive ability assessment indicators

As discussed above, five assessment indicators are established for the operators' key cognitive ability at the result layer, namely perceptual stability, attention-and-vigilance stability, working-memory stability, decision-making stability, and executive-function stability. These five indicators serve as the result-layer cognitive ability assessment indicators of the Bayesian network.

(a) Perceptual stability can be measured using the speed-perception test.

(b) Attention-and-vigilance stability can be measured using the PVT [12].

(c) Working-memory stability can be measured using the 2-back test, including both verbal and spatial versions [13].

(d) Reasoning-and-decision stability can be measured using the BART [15].

(e) Executive-function stability can be measured using the color–word Stroop test [16].

Based on the above analysis of cause-layer influencing factors and result-layer assessment indicators, the basis for the Bayesian network-based prediction method is established.

3.1.3. Quantification and Normalization of Factors and Indicators

To build the cognitive ability variation model, a unified quantification method is needed to normalize indicators of different dimensions to a common scale. Each indicator value is mapped to the interval [0, 1], where 1 denotes the highest contribution to cognitive decline risk and 0 denotes the lowest. To normalize each continuous cause-layer factor, different transformations are applied depending on the indicator's directional relationship with risk. The cause-layer indicators are categorized into positive and negative risk factors accordingly.

For a positive risk factor (where a larger raw value implies higher risk, e.g., reaction time, noise, CO_2, and cumulative work hours), the normalized risk contribution is

$$c_{i,t} = \frac{C_{i,t} - C_{i,\min}}{C_{i,\max} - C_{i,\min}} \tag{1a}$$

For an indicator where a larger raw value corresponds to lower risk (e.g., sleep duration and circadian alertness level), it is classified as a negative risk factor and normalized by reversal:

$$c_{i,t} = \frac{C_{i,\max} - C_{i,t}}{C_{i,\max} - C_{i,\min}} \tag{1b}$$

where $c_i \in [0, 1]$ denotes the normalized risk contribution factor of the i-th cause-layer indicator, with larger values representing greater risk contribution; $C_{i,min}$ and $C_{i,max}$ are the lower and upper bounds of the raw indicator C_i; and the subscript t of the variable represents simple time.

(1) Quantification and normalization of cause-layer factors

The cause-layer factors are described as follows.

(a) Confinement duration c_1: Quantified by the number of confined-operation hours. Based on the multi-day confined-operation experiments, the relationship between c_1 and the operator key cognitive ability is established.

(b) Shift cycle c_2: Quantified by the shift timing (c_{21}) and the daily cumulative work hours (c_{22}); the equal-weight average of the two sub-indicators yields c_2.

(c) Circadian rhythm and sleep c_3: Quantified by the duration of the most recent sleep (c_{31}) and the circadian alertness level (c_{32}); the equal-weight average of the two sub-indicators yields c_3.

(d) Working environment c_4: Based on prior single-factor studies on environmental effects on cognitive ability, three key environmental factors in the confined environment are identified: noise (c_{41}), CO_2 concentration (c_{42}), and lighting (c_{43}). The weighted average of these three sub-indicators yields c_4 [23–25].

Equal weights are assigned to the sub-indicators within each factor. Specifically, the three environmental sub-indicators c_{41} (noise), c_{42} (CO_2 concentration), and c_{43} (lighting) are equally weighted, as no evidence from prior studies suggests a systematic difference in their relative contribution to cognitive decline in confined environments. Similarly, equal weights are assigned to the two sub-indicators of c_2 (shift timing and daily cumulative work hours) and c_3 (pre-shift sleep duration and circadian arousal level), since both sub-indicators within each factor are considered equally important dimensions.

(2) Quantification and normalization of result-layer indicators

Before and after each duty period, participants are tested on five dimensions: perceptual stability, attention-and-vigilance stability, working-memory stability, reasoning-and-decision stability, and executive-function stability. The five cognitive ability dimensions are denoted as g_1–g_5, respectively.

To eliminate the effect of differing dimensions and value ranges among the test indicators, the raw reaction-time results of each test are normalized; accuracy is already within 0~1.

The lower and upper bounds (min/max) used for normalizing each cause-layer factor are summarized in Table 2.

Table 2. Normalization parameters of reaction-time test results.

No.	Test	Max	Min
g_1	Speed-perception test	900	250
g_2	PVT test	1000	200
g_3	2-back verbal & spatial memory test	1800	650
g_4	BART test	70	10
g_5	Color–word Stroop test	1000	350

3.2. Bayesian Network-Based Prediction of Cognitive Ability Level

3.2.1. Bayesian Network Structure

The Bayesian network method fits the causal structure of cognitive ability and can serve as the mathematical basis for model construction. Depending on whether node states change over time, Bayesian networks are divided into a Static Bayesian Network (SBN) and a DBN. In this paper, the DBN is used to predict the temporal variation in cognitive ability in this work.

First, an SBN is constructed at a fixed time t as the structural basis. The node, $P_{cr,t} \in [0, 1]$, denotes the prior risk probability of cognitive decline at time t, derived from the normalized weighted aggregation of four indicators. A higher $P_{cr,t}$ indicates a greater risk of decline, corresponding to a lower cognitive ability level. It directly influences the cognitive state node M_t, which in turn affects the result indicator nodes $G_{1,t}$~$G_{5,t}$. Then, by treating time t as a variable, the SBN is extended along the time axis into a DBN. Under a first-order Markov assumption, the current state M_t depends only on its immediate predecessor M_{t-1}. M_{t-1} represents the posterior state of cognitive decline from the previous time step, serving as the historical memory in recursive prediction, where 1 indicates cognitive decline and 0 indicates the normal state. Thus, M_{t-1} is introduced as an additional parent node to capture this time dependence. The structure of the Dynamic Bayesian Network for the temporal variation in cognitive ability is shown in Figure 1, and its slices (BN slices) are shown in Figure 2.

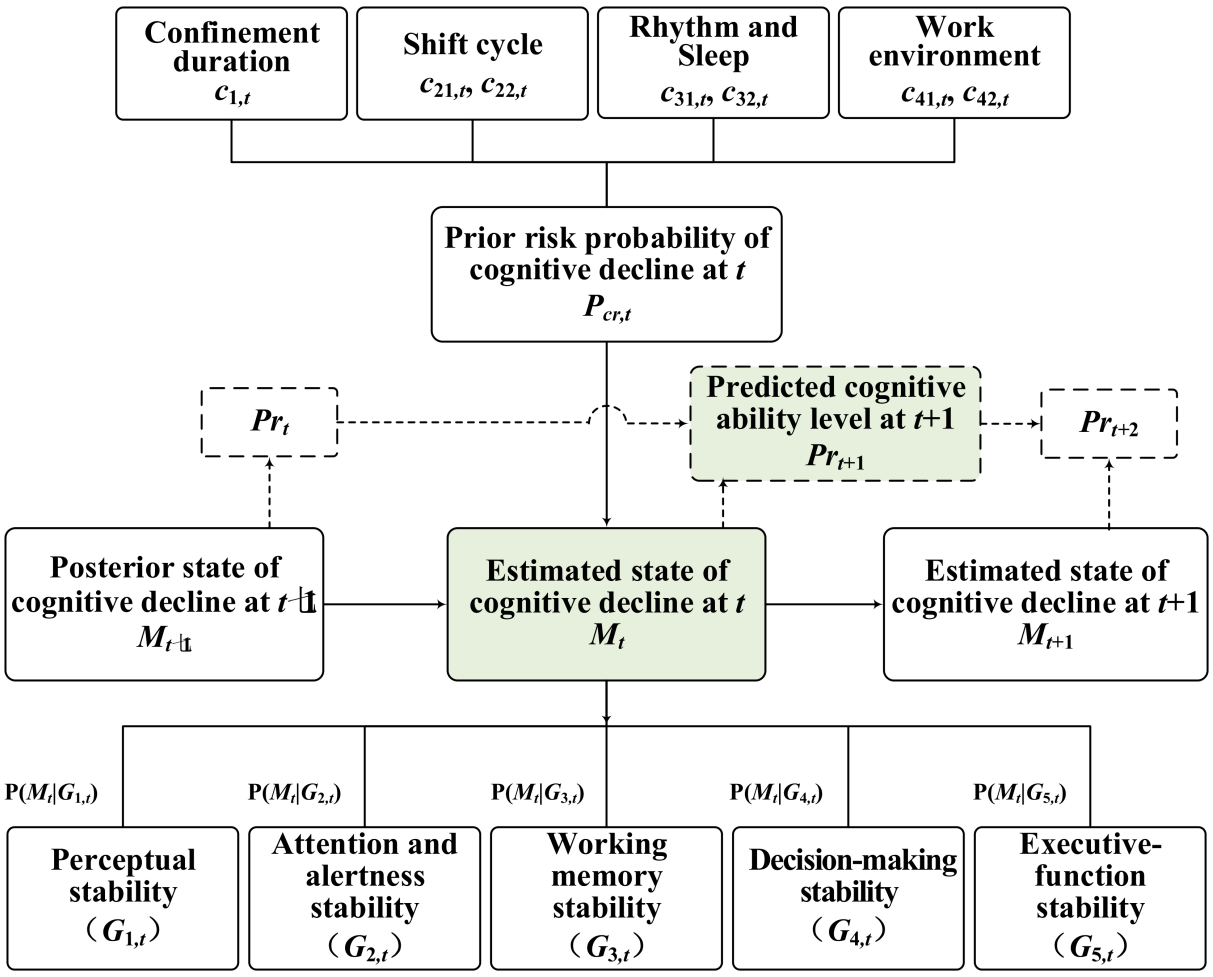


Figure 1. The DBN for cognitive ability level variation.

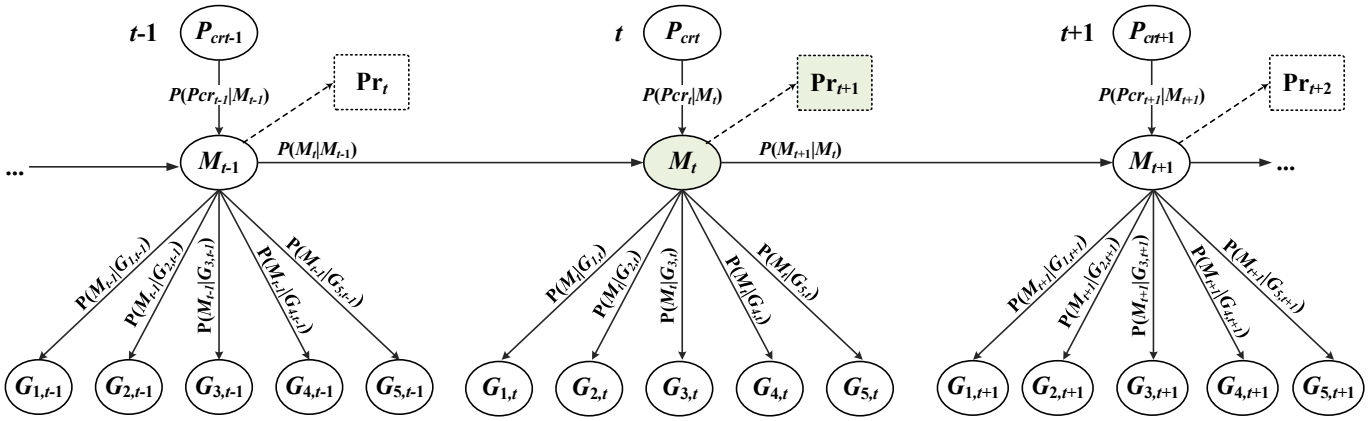


Figure 2. The BN slices in the DBN.

The Static Bayesian Network at each of the times $t-1$, t and $t+1$ can be regarded as one slice of the dynamic network, intuitively illustrating the temporal variation in cognitive ability level, *Pr*. The detailed calculation of *Pr* over time is presented in Equations (6)~(8). In particular, the decline deviation $\triangle Pr_t$ is obtained from the posterior probability and the current ability level. The inference process can be divided into two stages. In the first stage, Equations (4) and (5) perform state estimation, where the current cognitive state M_{t-1} is inferred from observations up to time t. In the second stage, Equations (6) and (7) perform prediction, where the filtered state is propagated forward to obtain the next-step cognitive ability level Pr_{t+1}. This two-stage structure enables the model to both track the current state and forecast its near-term evolution.

3.2.2. Prior Risk Probability of Cognitive Decline

The prior risk probability of cognitive decline, $P_{cr,t}$, as the parent of M_t, reflects the overall effect of external factors on cognitive ability level. A larger $P_{cr,t}$ indicates a higher risk of cognitive decline, which is then used to determine the prior probability. It is obtained by weighting the quantified results of the cause-layer indicators:

$$P_{cr,t} = \sum_i w_i c_i \tag{2}$$

where $P_{cr,t}$ is the prior risk probability of cognitive decline at time t; w_i is weight.

The weights w_1~w_4 were determined using the order relation analysis method (G1 method). The evaluation indicators were first qualitatively ranked, and adjacent indicators were then quantitatively scored on a 5-point influence scale from 0 (no influence) to 4 (very strong influence). It should be noted that, unlike the Analytic Hierarchy Process (AHP), the G1 method does not require a consistency ratio check, as it relies on sequential ranking and adjacent comparisons rather than pairwise comparison matrices. A total of 12 respondents were invited, including eight experts with research experience in confined-environment human factors and four experienced operators from long-duration isolated environments. Before the survey, all respondents were briefed on the indicators and scoring procedure.

Applying this method, the weights w_j are determined as $w_1 = 0.31$ (confinement duration), $w_2 = 0.22$ (shift cycle), $w_3 = 0.20$ (circadian rhythm and sleep) and $w_4 = 0.27$ (working environment). The rank order ($w_1 > w_4 > w_2 > w_3$) indicates that confinement duration has the highest weight, followed by working environment, shift cycle, and circadian rhythm and sleep. It is worth noting that this ranking differs from the emphasis in the human factor literature, where shift work and circadian rhythm are often identified as primary drivers of cognitive decline [6,7]. Prior work on environmental factors [22] also suggests their influence, although the experts in our survey ranked environmental factors lower

than confinement duration. This discrepancy may reflect the specific context of the present study: the 288-h simulated mission involved a relatively moderate shift schedule, and the experts may have weighted the cumulative psychological and physiological burden of prolonged confinement more heavily. This discrepancy warrants further investigation in future work.

3.2.3. Determination of Conditional Probabilities

Besides specifying the network structure, the conditional probabilities must be defined. Based on existing experimental results, the proportion of samples satisfying a given condition is used as the conditional probability. For example, when the parent node $\pi(X_i)$ is in state j, the conditional probability of node X_i is

$$P(X_i = k \mid \pi(X_i) = j) = N_{ijk}/N_{ij} \quad (3)$$

where N_{ijk} is the number of samples for which the parent state is j and the child node X_i is in state k, and N_{ij} is the total number of samples for which the parent is in state j. The cognitive ability node M_t takes states 0 and 1, denoting no decline and decline, respectively.

For the five result indicator nodes, $G_{1,t}$~$G_{5,t}$, a measured value better than the relative baseline is coded as 1 (less likely to cause decline), and a value worse than the baseline is coded as 0 (more likely to cause decline). This yields the conditional probabilities in Table 3.

Table 3. Conditional probabilities.

Result Variable	$P(G_i = 0 \mid M_t = 0)$	$P(G_i = 1 \mid M_t = 0)$	$P(G_i = 0 \mid M_t = 1)$	$P(G_i = 1 \mid M_t = 1)$
$G_{1,t}$	0.55	0.45	0.45	0.55
$G_{2,t}$	0.8	0.20	0.20	0.80
$G_{3,t}$	0.55	0.45	0.45	0.55
$G_{4,t}$	0.7	0.30	0.30	0.70
$G_{5,t}$	0.65	0.35	0.35	0.65

Note: Each row satisfies the sum-to-one axiom. The discriminative gaps reflect the expected sensitivity of each test to cognitive state changes, as inferred from task-specific longitudinal characteristics.

The conditional probabilities in Table 3 are prior probabilities specified based on qualitative sensitivity trends reported in the literature and domain expert knowledge, since exact numerical values are not directly available in the existing literature. They are not arbitrarily assigned but are synthesized from multiple converging sources: ① the longitudinal trends observed in our current simulated-task experiment, ② patterns repeatedly identified in our previous simulation-based cognitive studies under similar confined operational scenarios, and ③ established findings in the open literature on cognitive-test sensitivity to cognitive state transitions. The discriminative gap between $P(G_i = 0 \mid M_t = 0)$ and $P(G_i = 0 \mid M_t = 1)$ is intentionally differentiated to reflect the actual behavioral characteristics of each indicator, as detailed below:

(1) For G_2 (attention/vigilance, assessed by PVT), which exhibits the most pronounced and stable downward trend over time with minimal fluctuation, we assign the largest discriminative gap (0.80 vs. 0.20). This is consistent with the well-documented high sensitivity of PVT to sustained attention decrements and its established role as a gold-standard measure of vigilance in fatigue and cognitive-deprivation studies [3]. This setting indicates that PVT is the most reliable single indicator for detecting cognitive decline in our context.

(2) For G_4 (reasoning/decision, assessed by BART), which shows a clear mid-phase decline but with substantial inter-session variability, we assign a moderately large gap (0.70 vs. 0.30). The probabilities are intentionally pulled slightly toward 0.5 (compared

to G_2) to account for the higher measurement noise, reflecting that while the indicator is sensitive, its single-point observation is less trustworthy. This treatment is also in line with prior reports that risk-taking tasks like BART are susceptible to contextual and emotional fluctuations, which increase within-subject variability.

(3) For G_5 (executive function, assessed by Stroop), which displays a moderate downward trend with low variability, we assign an intermediate gap (0.65 vs. 0.35). This is consistent with the Stroop effect's known sensitivity to executive-control deterioration under sustained cognitive load, while its relatively stable test–retest reliability supports a moderate rather than extreme discriminative setting [16].

(4) For G_1 (perceptual stability, assessed using the speed-perception test) and G_3 (working memory, assessed using the 2-back task), where the long-term decline is not pronounced in the experimental data, we assign small gaps (0.55 vs. 0.45). These near-equal probabilities suggest that these two indicators are relatively insensitive to the overall cognitive state change in this study, and their observed variations are largely attributable to task-inherent difficulty or baseline noise rather than state-dependent decline. This is consistent with findings that simple perceptual tests often plateau under moderate stress, and that high-load working-memory tasks like 2-back may exhibit floor effects that compress the observable performance range even under cognitive fatigue [13].

These prior settings ensure that the Bayesian inference places appropriate weights on each indicator according to its empirically observed reliability and sensitivity, as corroborated by both our experimental trends and the broader cognitive-science literature, without requiring any fitting to the current dataset.

3.2.4. Prediction of Cognitive Decline Risk

Analyzing the variation in cognitive ability essentially amounts to solving for the probability that the result node M_t is in state 1, i.e., determining $P(M_t = 1 \mid G_{1,t}, \ldots, G_{5,t}, M_{t-1}, P_{cr,t})$. Using conditional independence and the law of total probability, the posterior probability of M_t in the Static Bayesian Network is

$$P(M_t = 1|G_{1:5}, P_{cr,t}) = \frac{P(M_t = 1|P_{cr,t}) \cdot \prod_{i=1}^{5} P(G_{i,t}|M_t = 1)}{\sum_{m\in\{0,1\}} P(M_t = m|P_{cr,t}) \cdot \prod_{i=1}^{5} P(G_{i,t}|M_t = m)} \quad (4)$$

After incorporating the temporal effect, the posterior probability in the Dynamic Bayesian Network (proportional to the normalizing constant) is

$$P(M_t = 1|G_{1:5}, P_{cr,t}, M_{t-1}) \propto P(M_t = 1|P_{cr,t}, M_{t-1}) \cdot \prod_{i=1}^{5} P(G_{i,t}|M_t = 1) \quad (5)$$

3.2.5. Prediction of Cognitive Ability Level

Existing studies mostly build universal models and rarely capture the specific patterns of a work group or an individual operator. Based on the constructed cognitive decline risk probability distribution, a conditional-probability correction factor is used to characterize the uniqueness of different operators. On the basis of Equation (5), the current prior probability and the posterior probability $P(M_t = 1 \mid G_{i,t}, P_{cr,t}, M_{t-1})$ at the measured nodes are combined and converted into the current cognitive ability. The cognitive ability level, Pr_t, is analyzed at per duty period. Let $Pr_t \in [0, 1]$ denote the cognitive ability level at time t, where 1 indicates the best performance and 0 the worst.

The upper and lower bounds of the cognitive ability level fluctuation, $Pr_{\max}$ and $Pr_{\min}$, and the initial cognitive ability Pr_0 are defined for each work group or a participant. Combining the attenuated posterior probability $P(M_t = 1 \mid G_{i,t}, P_{cr,t}, M_{t-1})$ and setting a

maximum-decay-deviation constraint $\Delta Pr_{\max}$, the cognitive ability decline deviation ΔPr_t within a working period is calculated from the current ability level Pr_t:

$$\Delta Pr_t = \min\{P(M_t = 1 \mid G_{i,t}, P_{cr,t}, M_{t-1}) \cdot \gamma_1 \cdot (Pr_t - Pr_{\min}), \Delta Pr_{\max}\} \quad (6)$$

where $P(M_t = 1 \mid G_{i,t}, P_{cr,t}, M_{t-1})$ is the attenuated posterior probability, ranging 0~1; γ_1 is the decay coefficient.

The cognitive ability level prediction for the next time step, P_{rt+1}, is then updated as

$$Pr_{t+1} = Pr_t - \Delta Pr_t \quad (7)$$

In this study, factors such as cognitive rest are the main means of alleviating cognitive decline. Therefore, given the recovery rate γ_2 of cognitive ability over time, the cognitive ability level for different individual or group during a rest period can be obtained as

$$Pr_{t+1} = Pr_t + \min\{\gamma_2 \cdot (Pr_{\max} - Pr_t) \cdot \Delta T, \Delta Pr_{\max}\} \quad (8)$$

where γ_2 is the recovery coefficient; ΔT is the rest duration, h.

In this study, Pr_0, $Pr_{\max}$, $Pr_{\min}$ and $\Delta Pr_{\max}$ are global parameters determined from the observed extremes across all participants: $Pr_0 = 0.90$, $Pr_{\max} = 0.90$, $Pr_{\min} = 0.75$, and $\Delta Pr_{\max} = 0.01$. γ_1 and γ_2 are group-specific parameters that reflect the different decay and recovery characteristics associated with each shift schedule. Group A1 operates during 00:00~04:00 and 12:00~16:00, a schedule that is more disruptive to circadian rhythms and may contribute to the different decline and recovery patterns. Groups A2 and A3 have shift schedules that more closely follow normal circadian rhythms, resulting in slower but steadier changes. Accordingly, the values of γ_1 for Group A1~A3 are 2, 0.4 and 0.4, respectively, and the values of γ_2 for Group A1~A3 are 1.5, 0.05 and 0.05, respectively.

4. Experimental Design

To further analyze the variation in cognitive ability during tasks and validate the accuracy of the constructed model, 15 participants (Participants 1~15) were invited to take part in a simulated task. A 15-day experiment was conducted, consisting of a 3-day training period for the participants and a 12-day formal experiment (288 h in total). The participants were divided into three groups by duty time, named A1~A3: Group A1 comprised Participants 1~5, Group A2 comprised Participants 6~10, and Group A3 comprised Participants 11~15. Each group performed two task sessions per day; the duty times of the groups are shown in Figure 3. For example, the duty times of Group A1 were 0:00~4:00 and 12:00~16:00. Each group completed the assigned tasks within its assigned duty time, ensuring that one group was always on duty at any time of day. Apart from the assigned tasks, participants could rest or engage in recreational activities during the remaining time. This study was approved by the Institutional Review Board (IRB) of Beihang University.

As shown in Figure 4, the experimental task was designed around a typical monitoring scenario and comprised four sub-tasks: perception, comprehension, prediction, and decision. Five cognitive tests, including the speed-perception test, PVT, verbal and spatial 2-back tests, BART, and the Stroop test, were conducted before and after each duty period. In addition, the working environment was continuously monitored, and mental-state scales and physiological indicators were measured at regular intervals. In the experimental task, cognitive tests were conducted before and after each duty period. For each duty period, the five cognitive indicators $G_{1,t}$~$G_{5,t}$ were obtained by averaging the pre-duty and post-duty test results. These averaged values were used as the model inputs and assigned to the

midpoint of the corresponding duty period. Thus, each group contributes four data points per day, corresponding to its two daily duty periods.

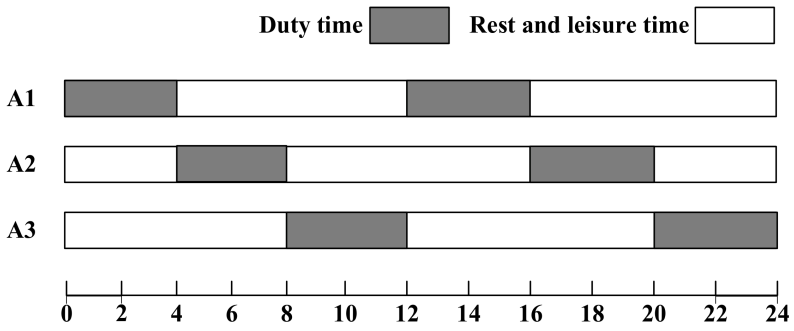


Figure 3. Duty-shift time frame during the experiment.



Figure 4. Measurement platform.

5. Results and Discussion

5.1. Experimental Measured Results of Cognitive Ability Level

For each group, the experimental cognitive ability level, $Prex_i$ (i = 1, 2, 3), is obtained by averaging the five indicator results (G_1~G_5) with equal weights across participants within the group. The measured cognitive ability data for the three groups (A1, A2, and A3) are presented in Figure 5.

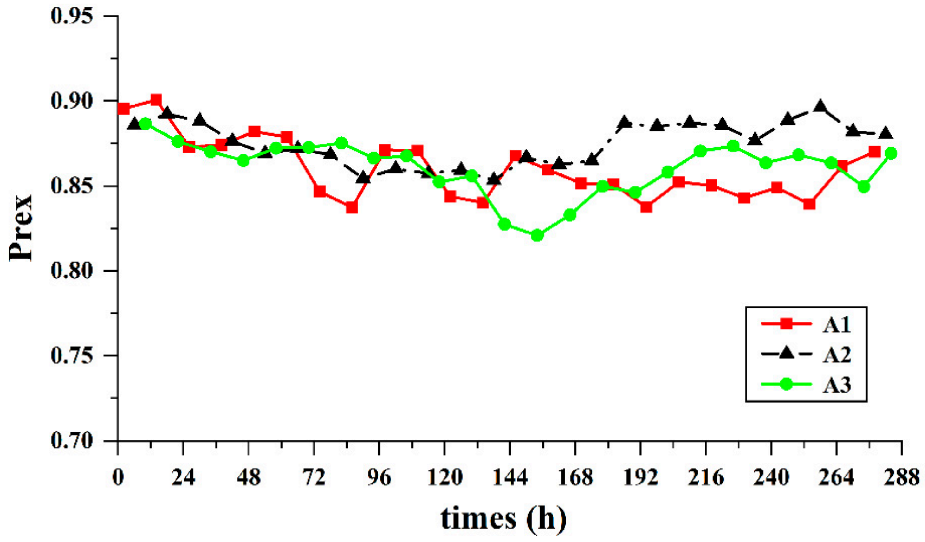


Figure 5. Measured cognitive ability variation for Groups A1, A2, and A3.

The three groups (A1, A2, A3) exhibit a consistent decline-then-recovery pattern, but with clear differences in temporal phase. Group A1 reaches its minimum earliest at approximately 86 h, followed by Group A2 at approximately 138 h and Group A3 at approximately 154 h, indicating a progressive phase delay from A1 to A3. In terms of fluctuation amplitude, Group A3 shows the largest peak-to-trough variation, while Group A2 shows the smallest. Despite these differences, all three groups recover to similar levels by the end of the 288 h experiment, with final *Prex* values around 0.87~0.88.

5.2. Compared Model and Evaluation Metrics

(1) Baseline models

(a) DBN-based prediction model

To compare with the DBN-based prediction method for cognitive ability level proposed in this paper, additional comparative analyses are carried out using the same experimental dataset. A reference baseline model is generated by applying a Static Bayesian Network (SBN). Although the proposed DBN-based prediction method is not a data-trained neural network, we also conduct a variant analysis by removing its temporal structure to verify the contribution of the dynamic component to the overall predictive performance. Specifically, the proposed DBN-based prediction method is degraded to an SBN by removing the temporal transition term M_{t-1} in Equation (5), i.e., assuming independence across time slices. The results of the SBN are represented as Pr_{SBN}. By comparing the SBN results with those of the proposed DBN-based prediction method, the contribution of the dynamic structure to the predictive performance can be shown.

(b) Mean predictor

In addition to the SBN, a mean predictor is included as a simple lower-bound baseline. For each group, the mean predictor estimates the cognitive ability level at time $t + 1$ as the average of all observed P_{rex} values up to time t, i.e., $Prm_{t+1} = (\sum_{i=1}^{t} Prex_i)/t$. This baseline uses only historical information and involves no temporal modeling or prior knowledge. It provides a reference point for evaluating whether the DBN captures meaningful temporal variation beyond the overall mean level.

(2) Evaluation metrics

To comprehensively evaluate the predictive performance of the proposed model, three complementary metrics are adopted: the RMSE, the MPE and the MCCC.

The RMSE can evaluate the fitting stability and robustness against fluctuations. A smaller RMSE indicates smaller deviation variability. It is calculated as

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(Pr_t - Prex_t)^2} \tag{9}$$

The MPE provides a directional indication of systematic prediction bias, revealing whether the model tends to overestimate or underestimate the measured values. It is defined as

$$MPE = \frac{1}{N}\sum_{t=1}^{N}\frac{Pr_t - Prex_t}{Prex_t} \times 100\% \tag{10}$$

A positive MPE indicates a tendency to overestimate, while a negative MPE indicates that the model tends to underestimate the cognitive ability. The absolute value of MPE reflects the average relative deviation, offering a scale-independent interpretation of the prediction bias.

The MCCC can measure the maximum linear correlation between a predicted sequence and a measured sequence, allowing for a small time shift between them. In this study, the maximum number of lag steps is set to $K = 2$. For each lag k ($|k| \leq K$), the predicted sequence is shifted relative to the measured sequence by k time steps. The overlapping portions of the two sequences are then used to compute the Pearson correlation coefficient r_k, calculated as

$$r_k = \frac{\sum_{t=1}^{N} \left(Prex_t - \overline{Prex}\right)\left(Pr_t - \overline{Pr}\right)}{\sqrt{\sum_{t=1}^{N} \left(Prex_t - \overline{Prex}\right)^2} \cdot \sqrt{\sum_{t=1}^{N} \left(Pr_t - \overline{Pr}\right)^2}} \tag{11}$$

$$MCCC = \max_{|k| \leq K}(r_k) \tag{12}$$

where $Prex_t$ and Pr_t are the t-th overlapping samples of the measured sub-sequence and the predicted sub-sequence, respectively, and $\overline{Prex}$ and $\overline{Pr}$ are their respective means.

The closer the value of MCCC is to 1, the stronger the positive linear synchronization between the predicted and measured sequences under optimal time alignment.

Together, RMSE, MPE, and MCCC provide a balanced assessment of the model predictive accuracy: RMSE measures the magnitude of numerical deviation, MPE indicates the direction of systematic bias (overestimation or underestimation), and MCCC evaluates the consistency of the temporal trend.

5.3. Comparison Results

5.3.1. Group-Level Prediction Results

Building upon the individual-level illustrations, this subsection extends the evaluation to the group level, which is the primary focus of this study. The group-level predictions are generated by applying the proposed DBN-based prediction method to the group-averaged input data for each shift condition (Groups A1, A2, and A3). The predicted cognitive ability level variations for the three groups are shown in Figures 6–8. Each figure presents a four-curve comparison among *Prex*, *Pr*, *Prm* and Pr_{SBN}.

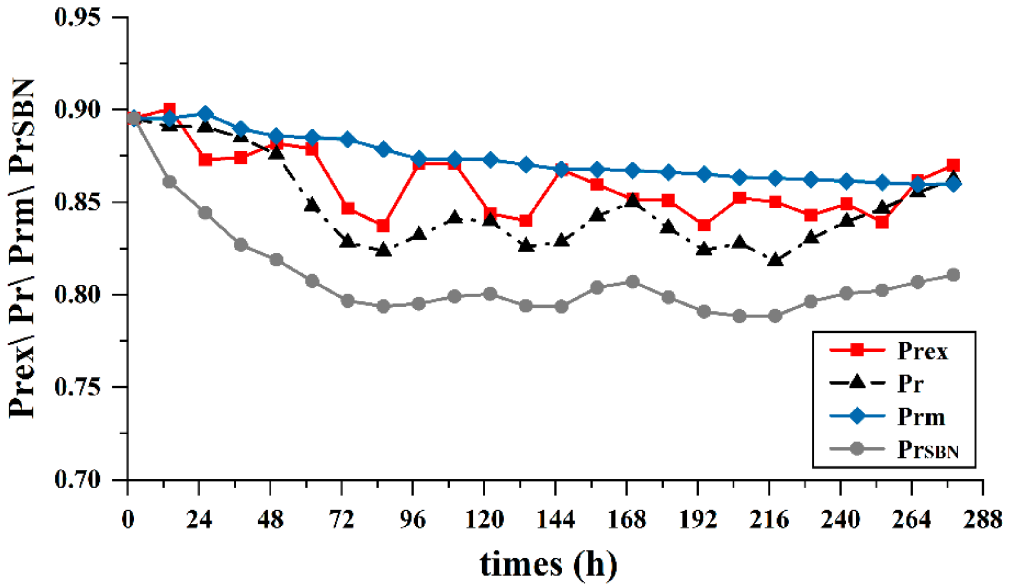


Figure 6. Comparison of *Prex*, *Pr*, *Prm* and Pr_{SBN} for Group A1.

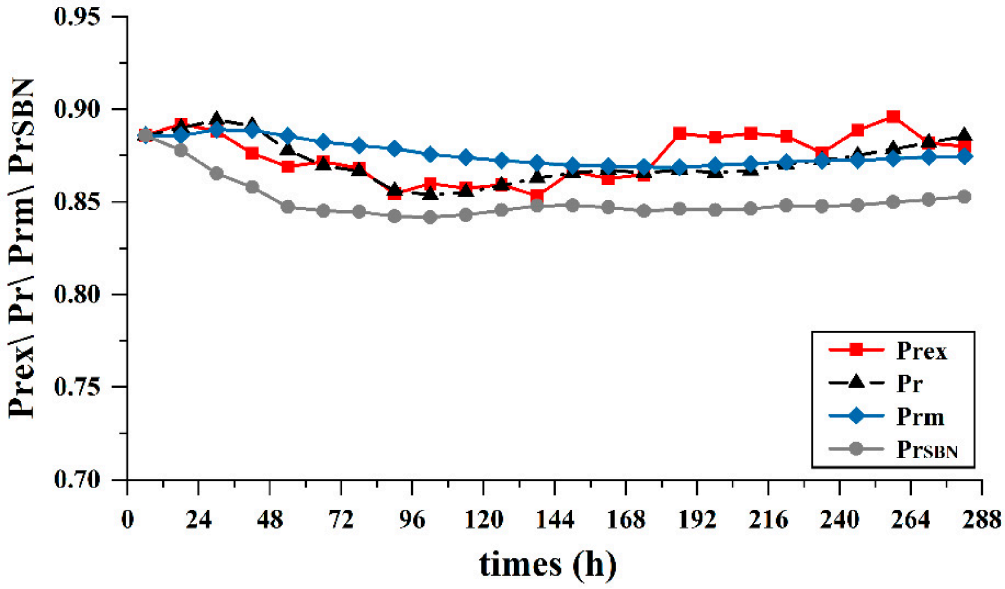


Figure 7. Comparison of *Prex*, *Pr*, *Prm* and Pr_{SBN} for Group A2.

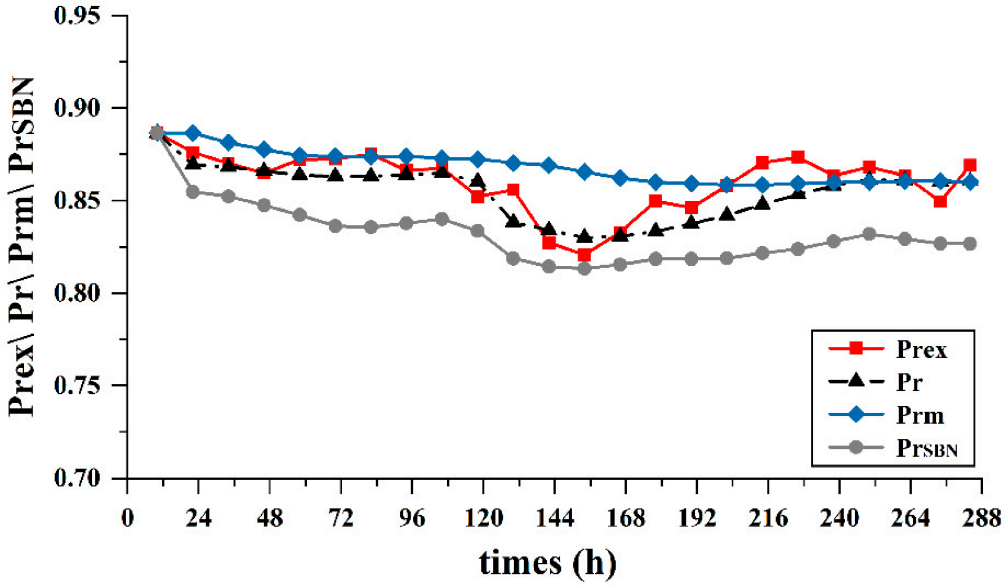


Figure 8. Comparison of *Prex*, *Pr*, *Prm* and Pr_{SBN} for Group A3.

As shown in Figures 6–8, the predictions of the DBN-based model generally follow the trend of the measured data across all three cases. In contrast, the SBN predictions remain relatively flat and tend to underestimate the measured values, highlighting the advantage of incorporating temporal dependencies in the DBN framework. This pattern is consistent with the group-level observations and suggests that the time-dependent structure enables the DBN to better capture the dynamic fluctuations inherent in cognitive validations. The mean predictor is generally higher than the measured data and shows a smooth trend.

Meanwhile, the SBN, lacking such temporal dependency, fails to respond to short-term variations. The SBN performs inference independently at each time slice without incorporating historical information, so its outputs primarily reflect the immediate influence of the cause-layer factors. Since these external factors tend to vary smoothly over time, the SBN yields relatively flat predictions and tends to stay at a lower level. In contrast, the DBN incorporates temporal dependencies through the transition term $M_{t-1} \rightarrow M_t$. This allows

the DBN to capture both the accumulation of cognitive load and the subsequent recovery during rest periods, resulting in more pronounced fluctuations that better reflect the decline-then-recovery pattern observed in the measured data. This comparison suggests that the time-dependent structure is a key factor in modeling the dynamic nature of cognitive ability during long-duration operations.

At the group level, the DBN-based predictions generally capture the measured cognitive validations across all three shift conditions, with errors remaining at very low levels. The corresponding RMSE, MPE, and MCCC values are summarized in Table 4.

Table 4. Group-level prediction accuracy comparisons.

Group	Model	RMSE	MPE	MCCC
A1	DBN	0.010	−0.31%	0.786
	SBN	0.031	−3.24%	0.763
	Mean predictor	0.019	1.64%	0.761
A2	DBN	0.010	−0.84%	0.759
	SBN	0.027	−2.73%	0.530
	Mean predictor	0.014	0.16%	0.446
A3	DBN	0.011	−0.64%	0.868
	SBN	0.031	−3.28%	0.783
	Mean predictor	0.017	0.96%	0.536

In Table 4, the DBN generally achieves lower RMSE and MPE and higher MCCC than the SBN and the mean predictor across all three groups, which is consistent with the individual-level findings. For the three groups, the RMSE values of the DBN-based model are nearly identical across groups (approximately 0.01 for A1, A2, and A3), compared to 0.027~0.031 for the SBN model; the MPE values of the DBN-based model range from −0.31% to −0.64%, compared to −2.73% to −3.28% for the SBN model; and the MCCC values of the DBN-based model range from 0.759~0.868, compared to 0.530~0.783 for the SBN. These results provide evidence that the DBN provides accurate and stable group-level predictions across different shift schedules, with the dynamic structure offering advantages in the evaluated cases over the static counterpart. The RMSE of the mean predictor ranges from 0.014 to 0.019, the MPE ranges from 0.016% to 1.64%, and the MCCC ranges from 0.446 to 0.761. The mean predictor produces a smooth, nearly constant prediction and cannot capture the temporal fluctuations of cognitive ability within duty periods. Therefore, the subsequent analysis focuses on the DBN and the SBN, which share the same causal structure and differ only in temporal dependency.

From the above comparisons at both the individual and group levels, it can be seen that the full DBN generally outperforms the SBN in terms of numerical accuracy and trend alignment. While more systematic individual-level validation with larger samples is planned for future work, the current results already support the model's applicability at both levels.

5.3.2. Individual-Level Prediction Examples

To illustrate the model's performance at the individual level, three participants (one from each group) are selected as illustrative examples: A1P2, A2P1, and A3P1. Their individual test data are used to predict cognitive variations, as shown in Figures 9–11. Since the individual-level results for all 15 participants are provided in Table 5, these three cases serve as illustrations rather than as statistically representative selections.

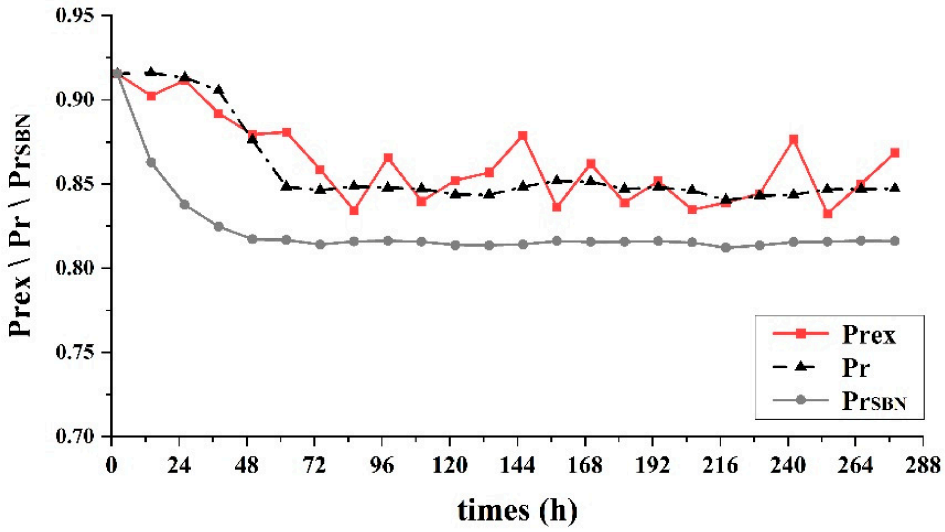


Figure 9. Comparison of *Prex*, *Pr* and Pr_{SBN} for Participant A1P2 from Group A1.

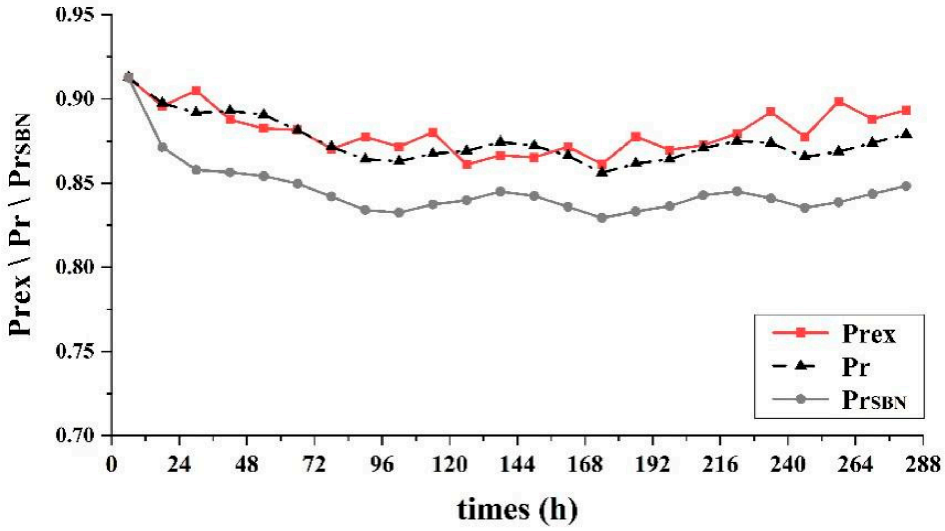


Figure 10. Comparison of *Prex*, *Pr* and Pr_{SBN} for Participant A2P1 from Group A2.

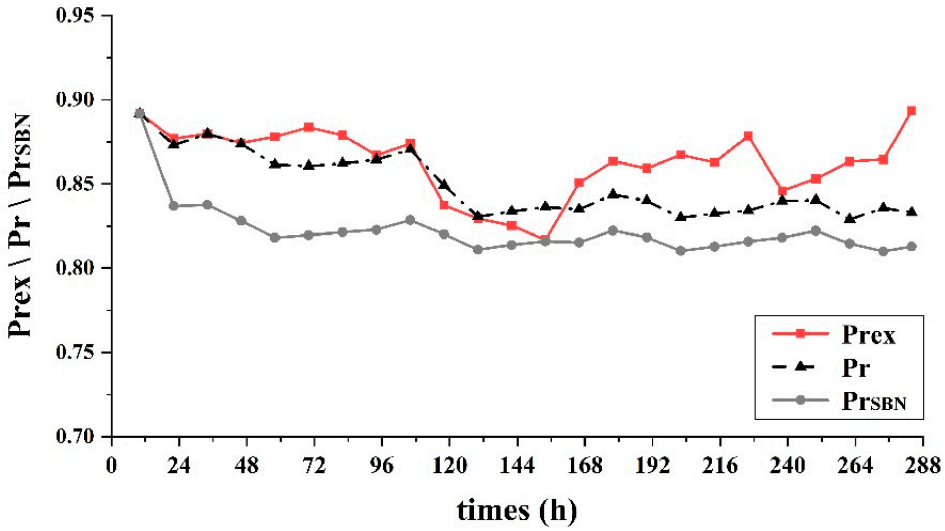


Figure 11. Comparison of *Prex*, *Pr* and Pr_{SBN} for Participant A3P1 from Group A3.

Table 5. Individual-level prediction accuracy comparisons for all participants.

Participant	RMSE		MPE		MCCC	
	DBN	SBN	DBN	SBN	DBN	SBN
A1P1	0.030	0.039	1.769%	−1.848%	0.778	0.605
A1P2	0.015	0.044	−0.421%	−4.587%	0.817	0.715
A1P3	0.033	0.056	−0.977%	−5.184%	0.654	0.498
A1P4	0.028	0.060	−1.450%	−6.169%	0.841	0.641
A1P5	0.019	0.053	0.763%	−5.635%	0.797	0.762
A2P1	0.011	0.037	−0.640%	−3.950%	0.777	0.756
A2P2	0.009	0.022	0.53%	−2.223%	0.758	0.621
A2P3	0.027	0.042	−0.103%	−3.526%	0.594	0.511
A2P4	0.041	0.061	−2.87%	−6.058%	0.772	0.707
A2P5	0.015	0.041	−0.344%	−4.210%	0.535	0.469
A3P1	0.023	0.046	−1.680%	−4.87%	0.657	0.514
A3P2	0.021	0.054	−1.063%	−5.570%	0.630	0.516
A3P3	0.017	0.042	0.846%	−3.526%	0.791	0.511
A3P4	0.031	0.049	−0.775%	−5.013%	0.706	0.590
A3P5	0.034	0.05	0.508%	−3.660%	0.539	0.407

Note: All errors are computed with respect to the measured *Prex* values for the corresponding participant. Lower RMSE and MPE indicate better numerical accuracy, while higher MCCC indicates better temporal trend alignment.

In all three cases, the DBN-based predictions generally follow the trend of the measured data, despite the larger fluctuations inherent in individual-level observations due to individual variability.

Before presenting the prediction accuracy, it should be noted that the conditional probabilities in Table 3 and the cause-layer weights were established independently of the 15-participant dataset. Specifically, the conditional probabilities were specified a priori based on literature trends and expert knowledge, and the weights were obtained from an independent expert survey. The group-specific decay and recovery coefficients γ_1 and γ_2, however, were calibrated from the observed trends in the experimental data. Therefore, the evaluation should be interpreted as partially out-of-sample, since the core prior structure was fixed independently while the two group-specific coefficients were calibrated from the data. The corresponding RMSE, MPE, and MCCC values for all participants are summarized in Table 5.

In Table 5, the DBN consistently achieves lower RMSE and MPE and higher MCCC than the SBN across all 15 individuals, which is consistent with the group-level findings. For all participants, the RMSE of the DBN-based model ranges from 0.009 to 0.041, compared to 0.022 to 0.061 for the SBN; the MPE of the DBN-based model ranges from −2.87% to 1.77%, compared to −6.17% to −1.85% for the SBN; and the MCCC of the DBN-based model ranges from 0.535 to 0.841, compared to 0.407 to 0.762 for the SBN. These results suggest that the model is capable of generating reasonable individual predictions when individual input data are available.

Based on all 15 individual-level prediction results, the mean RMSE is 0.024 for the DBN and 0.046 for the SBN, representing a reduction of approximately 48%. The mean MPE is −0.39% for the DBN and −4.34% for the SBN. The mean MCCC is higher for the DBN than for the SBN, with values of 0.682 and 0.588, respectively. These results indicate that the DBN showed better predictive consistency and lower prediction errors than the SBN across all individuals, demonstrating that the proposed method has stable and improved individual-level predictive performance in this sample.

5.3.3. Sensitivity Analysis of *Pr* with Respect to Weights

To evaluate the robustness of the model to variations in the weights, a one-at-a-time sensitivity analysis is conducted. Each weight w_i is perturbed by ±50% in turn. To maintain the normalization constraint ($\sum w_i = 1$), the remaining three weights are rescaled proportionally according to

$$w_i' = w_i \times (1 \pm 10\%)$$

$$w_j' = w_j \times \frac{1 - w_i'}{w_i}, \text{ for all } j \neq i$$

where w_i' is the perturbed weight and w_j' is the adjusted weights for the other three factors.

For each perturbation, the deviation in the predicted cognitive ability level Pr' was computed at each duty-period point for each of the 15 participants, relative to the baseline *Pr* obtained with the original weights. The maximum absolute deviation across all duty-period points and participants is summarized in Table 6.

Table 6. Sensitivity of *Pr* to ±50% variations in each cause-layer weight.

Group	Maximum Absolute Deviation
A1	0.047
A2	0.029
A3	0.025

The results show that, even under ±50% variations in each weight, the maximum absolute deviation in *Pr* remained below 0.047, which is small relative to the typical range of *Pr*. The mean absolute deviations were mostly below 0.01 and are therefore not separately reported. Deviations were computed relative to the baseline *Pr* obtained with the original weights. This indicates that the model predictions are reasonably robust to variations in the expert-derived weights. The reason for this insensitivity is that the prior risk probability P_{cr} remained above 0.5 for all participants throughout the experiment, so weight variations did not push the prior across the decision boundary. The final prediction was therefore primarily driven by the observed cognitive indicators $G_{1,t}$~$G_{5,t}$, which corrected the prior through the Bayesian posterior update. This is a desirable property for operational deployment, as the model can accommodate reasonable variations in expert judgment without producing unstable predictions.

It should be noted that the limited sensitivity of *Pr* to weight variations is a consequence of the moderate conditions of this experiment. Confinement duration, environmental factors, and work hours were all within normal ranges; only shift timing introduced a moderate level of circadian stress. As a result, $P_{cr,t}$ remains within a stable range throughout the experiment, and the final prediction was dominated by the observed cognitive indicators. The causal layer still provides an interpretable prior structure, but its influence on the final prediction was limited under these conditions.

5.4. Study Limitations and Future Work

Several limitations of this study should be acknowledged. First, each shift group comprised only five participants, which limits the statistical power of the analysis and precludes formal significance testing for group differences. Therefore, the observed differences in cognitive validations among the three groups should be interpreted as preliminary observations rather than definitive causal conclusions. Additionally, factors such as chronotype and sleep habits were not systematically controlled, which may contribute to inter-individual variability. Nevertheless, all participants were recruited under a uniform protocol and

maintained consistent living, meal, and rest schedules throughout the experiment, which helped mitigate the influence of lifestyle and environmental confounders.

Second, the conditional-probability correction factor and the individualized prediction mechanism have not been separately evaluated through ablation experiments. This is mainly because these components are relatively tightly integrated into the DBN framework, and isolating them would likely require substantial modifications to the core model logic, which is considered beyond the scope of this study. Instead, the overall performance improvement of the full DBN over the SBN is used as an indirect indication of their contribution. While it is recognized that this is not the most direct form of validation, it is hoped that the consistent superiority of the DBN across all three groups may still offer some support for the effectiveness of the complete dynamic design. In addition, since γ_1 and γ_2 were calibrated from the experimental data, the evaluation is not fully out-of-sample. Future work should validate the framework on an independent dataset to establish its predictive performance more rigorously. The present experiment did not include extreme conditions such as prolonged confinement or severe environmental stressors. Therefore, the influence of the cause-layer weights under such conditions cannot be empirically demonstrated with the current dataset. Validating the causal layer under more extreme conditions is identified as a direction for future work.

The primary contribution of this study remains the validation of the proposed DBN-based prediction framework in a controlled confined-environment setting. The results show that the model can effectively capture the temporal evolution of cognitive ability at the group level. Future work should involve larger-scale studies with balanced experimental designs and randomized assignment to enable more robust statistical inference and to explore individual-level modeling as the dataset size permits.

5.5. Discussion

The results presented above show that cognitive ability declines over the task period and partially recovers during rest, and that the proposed DBN captures this pattern more accurately than the static SBN baseline. These findings can be related to prior studies in several respects.

First, the observed decline trend is consistent in direction with previous research on cognitive performance in extreme environments. Dev et al. [5] reported that performance slowed in early flight on tasks of processing speed, visual working memory, and sustained attention among ISS astronauts during 6-month low-Earth-orbit missions. Flynn-Evans et al. [6] demonstrated that circadian misalignment during spaceflight affects sleep and medication use, which are closely related to cognitive performance. Zong et al. [7] further emphasized that circadian disruption and sleep disorders during long missions impair cognitive function and key task performance. Our results are consistent with these observations in terms of the direction of cognitive change, and extend them by showing that the temporal evolution of cognitive ability can be represented quantitatively through a dynamic probabilistic framework.

Second, the advantage of the DBN over the SBN is consistent with prior applications of DBNs in human factor and cognitive modeling. Chen et al. [9] used a DBN to estimate operator fatigue in plateau deep-mine environments, and Etholén et al. [10] applied a DBN to forecast subjective cognitive decline risk in a general population cohort. Both studies reported that incorporating temporal dependencies improved prediction relative to static models, although their application contexts differ substantially from the present study. Our results extend this line of work to long-duration confined and shift-based operations.

Third, the observed phase delay from Group A1 to Group A3 may be related to differences in circadian alignment. Group A1 operates during 00:00~04:00 and 12:00~16:00,

a schedule that is more disruptive to circadian rhythms, which may accelerate both decline and recovery. Groups A2 and A3 have schedules closer to normal circadian rhythms, which may lead to slower but steadier changes. This interpretation is consistent in direction with prior work on circadian misalignment and cognitive performance [6,7]. However, given the small sample size (five participants per group), these explanations should be regarded as tentative and require validation in larger studies.

Thus far, a DBN-based framework specifically designed for continuous cognitive prediction in long-duration confined and shift-based missions has not yet been systematically established. Most existing prediction methods rely on static regression or threshold-based approaches, which cannot capture temporal dependencies or multi-factor interactions. The results reported here therefore suggest that a dynamic, causally structured, and group-adaptive model may offer a more suitable framework for continuous cognitive monitoring in long-duration operations. Nevertheless, since the present study is based on a relatively small sample, these findings should be interpreted as preliminary evidence rather than definitive conclusions.

6. Conclusions

This study investigates the variation in long-duration operator cognitive ability level, denoted as $Pr \in [0, 1]$, under confined shift-based working conditions. By analyzing the causal relationships between cause-layer influencing factors (confinement duration, shift cycle, circadian rhythm and sleep, and working environment) and result-layer cognitive assessment indicators (perceptual stability, attention-and-vigilance stability, working-memory stability, decision-making stability, and executive-function stability), a DBN-based prediction model is constructed. A conditional-probability correction factor is introduced to enable predictions that reflect differences among groups or operators. The model is validated through a 288-h simulated-task experiment with 15 participants divided into three groups (A1~A3) under different shift timings. The main conclusions are as follows.

(1) The experimental results show that the cognitive ability of all three groups exhibits an overall declining trend over the 12-day task period, with fluctuations and partial recoveries that vary by shift timing. The three groups show a consistent V-shaped trend, with a progressive phase delay from A1 to A3. Group A1 reaches its minimum earliest, followed by Group A2 and then Group A3. All groups recover to similar levels by the end of the experiment. The observed phase differences suggest that shift schedule may influence the timing of cognitive decline and recovery, while the overall pattern remains consistent across groups.

(2) The proposed method incorporates a conditional-probability correction factor to capture group-specific cognitive variation patterns under different shift schedules. The effectiveness of the dynamic structure is further supported by individual-level prediction examples, which show that the DBN can generate reasonable predictions for individual participants when individual input data are available. These findings suggest that the model has potential for practical applications where inter-individual differences are significant, while more systematic individual-level validation with larger samples is planned for future work.

(3) The prediction accuracy of the proposed DBN-based model is evaluated using three complementary metrics: RMSE, MPE, and MCCC. At the individual level, the DBN generally achieves lower RMSE and MPE and higher MCCC than the SBN across the three representative participants. At the group level, the RMSE values for all three groups are below 0.011, the MPE values are below -0.84%, and the MCCC values exceed 0.759. The DBN also outperforms the mean predictor, confirming that it captures meaningful temporal variation rather than merely fitting the overall mean level. These results indicate

that the DBN-based model achieves good accuracy and trend consistency, and that the dynamic structure shows potential to improve prediction performance relative to the static counterpart in both individual and group-level predictions.

(4) The proposed method provides a feasible approach for the dynamic monitoring and early warning of cognitive ability in long-duration operators. The prediction results can inform the scheduling of rest periods, task allocation, and targeted interventions to mitigate cognitive decline in confined shift-based environments.

The practical implications of this study are summarized as follows.

From a practical perspective, the proposed DBN-based prediction framework may support scheduling and fatigue risk management in long-duration confined operations, such as space missions, deep-sea exploration, and polar stations. Specifically, the model can provide quantitative estimates of cognitive ability level *Pr* for different shift schedules, which could help practitioners identify potentially high-risk periods and inform adjustments to duty rosters or rest periods. The group-specific correction mechanism allows the framework to be adapted to different teams or shift patterns, suggesting potential for operational use in settings where individualized monitoring is not feasible.

From a research perspective, the study provides a DBN-based methodology for modeling the temporal evolution of cognitive ability under multiple interacting stressors. It offers a transparent causal structure linking cause-layer factors to multi-dimensional cognitive outcomes and demonstrates how prior knowledge and expert judgment can be integrated into a probabilistic framework without requiring large-scale training data. This provides a way for other researchers to extend the model with additional physiological indicators (e.g., heart rate variability, EEG), to validate it in larger and more diverse samples, and to explore multi-step forecasting and adaptive intervention strategies.

In summary, the DBN-based prediction model developed in this study offers a reliable tool for assessing and forecasting cognitive ability variation in long-duration operations, with potential applications in aerospace, deep-sea missions, polar stations, and other confined high-risk workplaces. Future work will focus on incorporating additional physiological indicators (e.g., heart rate variability, electroencephalogram) and expanding the dataset to further improve prediction accuracy and generalizability.

Author Contributions: Methodology, S.L., Y.Y., P.L. and L.P.; software, X.Z., P.L. and L.P.; validation, S.L., X.Z., Y.Y. and B.H.; formal analysis, S.L., X.Z., P.L. and L.P.; investigation, X.Z., Y.Y., L.Z. and B.H.; resources, X.Z. and L.P.; data curation, S.L. and L.Z.; writing—original draft, S.L.; writing—review & editing, S.L., X.Z., Y.Y., P.L., L.P. and L.Z.; visualization, X.Z.; supervision, S.L., Y.Y., P.L., L.P., L.Z. and B.H.; project administration, S.L., Y.Y. and L.Z.; funding acquisition, S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Foundation of National Key Laboratory of Human Factors Engineering, Grant NO. HFNKL2024W08.

Data Availability Statement: Due to the limitations of the Ethics Review Committee, these data cannot be made public to protect the privacy and confidential information of the subjects. The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors would like to thank all subjects involved in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Stahn, A.C.; Kühn, S. Extreme Environments for Understanding Brain and Cognition. *Trends Cogn. Sci.* **2022**, *26*, 1–3. [CrossRef] [PubMed]
2. Manzey, D.; Lorenz, B.; Poljakov, V. Mental Performance in Extreme Environments: Results from a Performance Monitoring Study during a 438-Day Spaceflight. *Ergonomics* **1998**, *41*, 537–559. [CrossRef] [PubMed]

3. Basner, M.; Savitt, A.; Moore, T.M.; Port, A.M.; McGuire, S.; Ecker, A.J.; Nasrini, J.; Mollicone, D.J.; Mott, C.M.; McCann, T.; et al. Development and Validation of the Cognition Test Battery for Spaceflight. *Aerosp. Med. Hum. Perform.* **2015**, *86*, 942–952. [CrossRef] [PubMed]
4. Strangman, G.E.; Sipes, W.; Beven, G. Human Cognitive Performance in Spaceflight and Analogue Environments. *Aviat. Space Environ. Med.* **2014**, *85*, 1033–1048. [CrossRef] [PubMed]
5. Dev, S.I.; Khader, A.M.; Begerowski, S.R.; Anderson, S.R.; Clément, G.; Bell, S.T. Cognitive Performance in ISS Astronauts on 6-Month Low Earth Orbit Missions. *Front. Physiol.* **2024**, *15*, 1451269. [CrossRef] [PubMed]
6. Flynn-Evans, E.E.; Barger, L.K.; Kubey, A.A.; Sullivan, J.P.; Czeisler, C.A. Circadian Misalignment Affects Sleep and Medication Use before and during Spaceflight. *npj Microgravity* **2016**, *2*, 15019. [CrossRef] [PubMed]
7. Zong, H.; Fei, Y.; Liu, N. Circadian Disruption and Sleep Disorders in Astronauts: A Review of Multi-Disciplinary Interventions for Long-Duration Space Missions. *Int. J. Mol. Sci.* **2025**, *26*, 5179. [CrossRef] [PubMed]
8. Caldwell, J.A. Fatigue in Aviation. *Travel Med. Infect. Dis.* **2005**, *3*, 85–96. [CrossRef] [PubMed]
9. Chen, S.; Pan, L.; Xu, K.; Li, X.; Zuo, Y.; Zhou, Z.; Li, B.; Dai, Z.; Li, Z. Real-Time Monitoring and Prediction of Remote Operator Fatigue in Plateau Deep Mining Based on Dynamic Bayesian Networks. *Sci. Rep.* **2025**, *15*, 1063. [CrossRef] [PubMed]
10. Etholén, A.; Roos, T.; Hänninen, M.; Bouri, I.; Kulmala, J.; Rahkonen, O.; Kouvonen, A.; Lallukka, T. Forecasting Cognitive Functioning: An Artificial Intelligence Approach with Dynamic Bayesian Networks. *Eur. J. Public Health* **2024**, *34*, ckae144.593. [CrossRef]
11. Wang, S.; Pang, L.; Li, P.; Jiao, T.; Wang, X.; Yin, H. A Bayesian Network-Based Prediction Method for Continuous Operational Capability During Team Collaborative Tasks. *Mathematics* **2025**, *13*, 3117. [CrossRef]
12. Dinges, D.F.; Powell, J.W. Microcomputer Analyses of Performance on a Portable, Simple Visual RT Task during Sustained Operations. *Behav. Res. Methods Instrum. Comput.* **1985**, *17*, 652–655. [CrossRef]
13. Kirchner, W. Age Differences in Short-Term Retention of Rapidly Changing Information. *J. Exp. Psychol.* **1958**, *55*, 352–358. [CrossRef] [PubMed]
14. Bechara, A.; Damasio, A.R.; Damasio, H.; Anderson, S.W. Insensitivity to Future Consequences Following Damage to Human Prefrontal Cortex. *Cognition* **1994**, *50*, 7–15. [CrossRef] [PubMed]
15. Lejuez, C.W.; Read, J.P.; Kahler, C.W.; Richards, J.B.; Ramsey, S.E.; Stuart, G.L.; Strong, D.R.; Brown, R.A. Evaluation of a Behavioral Measure of Risk Taking: The Balloon Analogue Risk Task (BART). *J. Exp. Psychol. Appl.* **2002**, *8*, 75–84. [CrossRef] [PubMed]
16. Stroop, J. Studies of interference in serial verbal reactions. *J. Exp. Psychol.* **1935**, *18*, 643–662. [CrossRef]
17. Van Dongen, H.P.A.; Maislin, G.; Mullington, J.M.; Dinges, D.F. The Cumulative Cost of Additional Wakefulness: Dose-Response Effects on Neurobehavioral Functions and Sleep Physiology from Chronic Sleep Restriction and Total Sleep Deprivation. *Sleep* **2003**, *26*, 117–126. [CrossRef] [PubMed]
18. Goel, N.; Rao, H.; Durmer, J.S.; Dinges, D.F. Neurocognitive Consequences of Sleep Deprivation. *Semin. Neurol.* **2009**, *29*, 320–339. [CrossRef] [PubMed]
19. Vogl, J.; O'BRien, K.; Onge, P.S. One size does not fit all: A support vector machine exploration of multiclass cognitive state classifications using physiological measures. *Front. Neuroergon.* **2025**, *6*, 1566431. [CrossRef] [PubMed]
20. Jiang, Y.; Li, J.; Zhang, Y. Fatigue Detection Algorithm for Nuclear Power Plant Operators Based on Random Forest and Back Propagation Neural Networks. *Mathematics* **2025**, *13*, 774. [CrossRef]
21. Basner, M.; Dinges, D.F.; Mollicone, D.; Ecker, A.; Jones, C.W.; Hyder, E.C.; Di Antonio, A.; Savelev, I.; Kan, K.; Goel, N.; et al. Mars 520-d Mission Simulation Reveals Protracted Crew Hypokinesis and Alterations of Sleep Duration and Timing. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2635–2640. [CrossRef] [PubMed]
22. Zhang, J.; Pang, L.; Cao, X.; Wanyan, X.; Wang, X.; Liang, J.; Zhang, L. The Effects of Elevated Carbon Dioxide Concentration and Mental Workload on Task Performance in an Enclosed Environmental Chamber. *Build. Environ.* **2020**, *178*, 106938. [CrossRef]
23. Zhang, J.; Pang, L.; Yang, C.; Fan, Y.; Zhao, B.; Cao, X. Experimental Evaluation of Noise Exposure Effects on Subjective Perceptions and Cognitive Performance. *Buildings* **2024**, *14*, 1100. [CrossRef]
24. Allen, J.G.; MacNaughton, P.; Satish, U.; Santanam, S.; Vallarino, J.; Spengler, J.D. Associations of Cognitive Function Scores with Carbon Dioxide, Ventilation, and Volatile Organic Compound Exposures in Office Workers: A Controlled Exposure Study of Green and Conventional Office Environments. *Environ. Health Perspect.* **2016**, *124*, 805–812. [CrossRef] [PubMed]
25. Sun, J.; Luo, Z.; Gao, Y.; Dong, Q.; Qi, X. The Interactive Effects of Illuminance and Correlated Color Temperature on Mood and Alertness Evaluated through Psychological and Physiological Measures. *Build. Environ.* **2026**, *294*, 114394. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.