Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Novel Applications of Statistical and Machine Learning Methods to Analyze Trial Level Data from Congnitive Measures

Dr. Maheshwari Munigala*

*Chhatrapati Shahu Ji Maharaj University, Kanpur. mahe7munigala@gmail.com

ARTICLE INFO

Received: 22 Dec 2024

Revised: 14 Feb 2025

Accepted: 24 Feb 2025

ABSTRACT

Traditional analytical methods use aggregate metrics, which fail to show the fine-grained patterns that emerge from trial-level fluctuations of cognitive performance due to dynamic internal states like attention, fatigue, and learning processes. This research examines how statistical and machine learning (ML) methods can analyse trial-level behavioural and physiological data to improve understanding of cognitive dynamics. A total of 9,276 trials were obtained from 112 participants who completed Stroop, N-back, and Go/No-Go tasks. The annotation of each trial included reaction time measurements alongside accuracy data, task condition information, and EEG-derived alpha power measurements. Our analysis incorporated Bayesian hierarchical models, generalized linear mixed models, state-space models, Random Forest, XGBoost, deep neural networks, and Long Short-Term Memory (LSTM) networks to forecast both reaction times and task accuracy. The LSTM model demonstrated the best predictive power by achieving $R^2 = 0.862$ for RT prediction and AUC-ROC = 0.925 for accuracy classification. The AUC-ROC score reached 0.925 for classification, while R^2 reached 0.862 for reaction time prediction, which proved superior to all other techniques. The predictive features of trial number, task congruency, and EEG alpha power emerged through Shapley Additive Explanations (SHAP) and LSTM saliency maps. The research demonstrates how combining statistical transparency with ML flexibility helps reveal personalized and time-dependent cognitive patterns. The proposed method provides a powerful structure for modelling trials while creating potential applications for individualized cognitive assessment systems in educational and mental health settings.

1. INTRODUCTION

Real-time cognitive analysis demands methods beyond simple measurement of average reaction times and aggregate accuracy scores. Traditional cognitive models utilize summary metrics as performance indicators but fail to recognize that cognitive processes naturally exhibit variability over time. The performance variables of reaction time (RT), accuracy, and error rates demonstrate trial-by-trial variations due to changes in internal states such as fatigue, attention, and strategic adaptations. Statistical efficiency in modelling can lead to cognitive impoverishment when researchers disregard process variability, resulting in inadequate representations of real-time cognitive operations. The current research literature suggests that scientists should move away from using static aggregate measures toward trial-level modelling to preserve the temporal and contextual integrity of behavioural data (Rouder & Haaf, 2019; Seli et al., 2016).

Trial-by-trial data reveal the dynamic nature of cognitive processes by showing how people adjust their performance to task requirements and how hidden mechanisms like engagement or effort change across time. Performance shows both between-task condition differences and within-session fluctuations, which stem from

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

rising cognitive load, learning effects, and attention-related breakdowns. The analysis of this variability demands modelling approaches that handle intricate within-subject fluctuations and maintain awareness of inter-trial dependencies.

The Bayesian cognitive modelling framework provides a successful method to analyze latent variables through its ability to handle uncertainty and hierarchical data structures. Bayesian models enable researchers to perform trial-level inference while maintaining the ability to generalize findings across individuals. Their dual functionality makes Bayesian cognitive models ideal for research because they effectively handle participant-level variability (Lee & Wagenmakers, 2014). The incorporation of prior distributions within Bayesian frameworks makes them suitable for iterative scientific investigation, particularly in domains such as working memory, attention, and executive function.

The integration of physiological measures into cognitive models has recently become more prevalent in research. The neuroscientific tool Electroencephalography (EEG) provides precise neural signals that track cognitive states, including attentional engagement, working memory load, and fatigue. Alpha-band power has emerged as a key indicator of cognitive control and inhibitory processing, as increases in alpha power signify task disengagement or mental fatigue (Cohen, 2017). The combination of trial-level behavioural data with EEG signals helps researchers uncover neural mechanisms driving performance changes. Combining multiple methods enables researchers to study cognition as an evolving system that responds to internal and external limitations, rather than treating it as a single stimulus-response mechanism.

The interpretability and probabilistic foundations of Bayesian and classical statistical models come at the cost of requiring linear, normal, or independent relationships, which often fail to match high-dimensional behavioural data patterns. The increasing demand for machine learning approaches that handle nonlinearities and complex interactions without requiring strict parametric assumptions has become a major area of interest. Random Forests and XGBoost have gained popularity due to their robust performance and ability to automatically identify relevant predictors when processing large, complex datasets (Breiman, 2001; Chen & Guestrin, 2016). The health and behavioural sciences use these models effectively to classify and predict outcomes from mixed data types.

Traditional ML models lack built-in capabilities to process sequential data patterns. Models used for time-dependent performance tasks must track dependencies across multiple time steps, as these factors affect outcomes such as trial order, fatigue, and learning. The adoption of recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM) architectures, has become essential due to their ability to maintain long-range dependencies through memory cells and gating mechanisms (Hochreiter & Schmidhuber, 1997). LSTMs provide cognitive modelling with a robust framework to track state evolution, enabling predictions that respond to both present inputs and past results.

Deep learning's ability to extract complex patterns from raw data creates challenges for achieving transparency. Cognitive science researchers demand explainable predictions from black-box models because their primary goal combines prediction with understanding. Modern interpretable machine learning techniques work to solve this issue. The Shapley Additive Explanations (SHAP) method enables researchers to decompose model predictions into individual feature contributions, providing both global and local interpretability (Lundberg & Lee, 2017). Behavioural research using this approach allows scientists to assess the relative importance of features such as trial number, task congruency, and EEG amplitude in predicting performance, thereby linking predictive models to explanatory frameworks (Molnar, 2020).

State-space modelling presents a promising approach for inferring latent cognitive variables through hidden states that produce observable behavioural outputs. These models use Kalman filters and particle filters to deliver a mathematically elegant solution for tracking internal processes such as attention and cognitive control over time. The behavioural sciences have started to adopt state-space approaches because these models reveal hidden dynamics that static models fail to detect (Westland, 2015). The combination of state-space models with ML techniques results in hybrid systems that preserve accuracy while maintaining cognitive grounding.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Few studies have systematically evaluated statistical models against ML models in the context of trial-level behavioural data. Integrating EEG features into predictive frameworks remains scarce because most approaches lack both time-awareness and interpretability capabilities. The existing knowledge gap creates an exciting opportunity to build hybrid modelling systems that merge statistical inference with deep learning techniques and neurophysiological understanding.

This study addresses the gap by applying Bayesian hierarchical models, state-space models, Random Forest, XGBoost, deep feedforward networks, and LSTM architectures to analyze over 9,000 trials from Stroop, N-back, and Go/No-Go tasks. The dataset includes reaction time and accuracy measures along with EEG alpha power. We measure each model's predictive ability through accuracy metrics while performing interpretability analysis with SHAP and saliency mapping methods.

This research evaluates cognitive modelling through a comprehensive comparison of statistical and machine learning methods on a unified dataset containing behavioural and physiological features. LSTM models demonstrate superior performance in detecting temporal patterns according to our hypothesis, while SHAP analysis reveals relevant predictors that align with well-established cognitive theories. The research aims to create transparent cognitive models that generate individualized predictions while accounting for temporal dynamics—bridging prediction, explanation, and practical application.

2. LITERATURE REVIEW

The development of cognitive modelling shows a fundamental transformation from general population-level predictions toward detailed examination of individual trial data. Theoretical developments, combined with expanded access to rich behavioural and physiological datasets—including neurocognitive task multimodal recordings—have enabled this transition. The move away from traditional summary statistics requires researchers to develop modelling methods capable of processing complex data structures while handling temporal dependencies and high-dimensional information. Machine learning (ML) and deep learning demonstrate superior performance to traditional methods by providing enhanced scalability and predictive capabilities in this context.

Random Forests and XGBoost stand as the most influential ensemble-based models in cognitive and behavioural sciences. Random Forests unite multiple decision trees to enhance generalization and minimize overfitting (Breiman, 2001), which enables their application to behavioural feature classification and regression tasks. XGBoost builds upon gradient boosting techniques and regularization to maintain robustness in noisy or imbalanced datasets (Chen & Guestrin, 2016). These non-parametric models successfully detect cognitive outcome predictors by recognizing complex feature relationships, which enables them to identify accuracy lapses, attention deficits, and reaction time variability.

The predictive capabilities of Random Forests and XGBoost models remain limited when dealing with time-dependent sequences, which are fundamental for analysing cognitive task trial data. Deep learning models, particularly Recurrent Neural Networks (RNNs) and their advanced version Long Short-Term Memory (LSTM) networks, have solved the sequential dependency limitation. LSTM networks were first developed through the integration of memory cells and gating mechanisms to capture long-term dependencies in data (Hochreiter & Schmidhuber, 1997). Research has demonstrated their success in time-series applications including speech recognition and financial forecasting, and now extends to cognitive task modelling. LSTM networks excel at handling trial-level cognitive data because they maintain and update context-dependent information, which affects trial outcomes based on previous trials.

Deep learning models' "black box" nature presents significant challenges for scientific applications that require both interpretability and predictive performance. The need for transparent machine learning models has resulted in the creation of Shapley Additive Explanations (SHAP) frameworks, which serve as interpretable machine learning (IML) tools. SHAP uses cooperative game theory to calculate importance values for model

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

inputs while providing unified Explanations of predictions (Lundberg & Lee, 2017). Through SHAP, researchers in cognitive modelling can determine which features drive predicted performance outcomes while gaining insight into complex model decision processes.

The demystification of ML models is supported by ongoing developments in the IML literature. Molnar (2020) promotes explainability integration throughout model development, particularly in cognitive neuroscience, because it provides both practical insights and theoretical coherence. Interpretability tools serve beyond diagnostic purposes by helping researchers develop hypotheses and engineer features, thereby creating a more efficient feedback loop between model development and cognitive theory.

Multimodal modelling has gained significant importance beyond traditional machine learning applications. Modern cognitive tasks combine behavioural outcomes (e.g., reaction times, error rates) with neurophysiological data such as EEG. Model sensitivity to latent cognitive states improves when these signals are integrated. Machine learning with neuroimaging or electrophysiological data enhances predictions of individual cognitive abilities and mental health conditions (Sui et al., 2020). EEG-derived alpha power functions as a biomarker for attention and mental workload, enabling more precise cognitive modelling.

Temporal attention mechanisms represent a new frontier in cognitive modelling research. Financial time-series forecasting models that use temporal attention modules (Tran et al., 2018) have shown enhanced accuracy and interpretability. These architectures identify crucial time steps while ignoring less informative ones, demonstrating a direct application to trial-level cognitive modelling. These attention-based mechanisms reveal critical cognitive stress points and recovery phases that standard performance data fail to identify.

Theoretical frameworks of active inference establish a foundational principle for cognitive modelling by viewing the brain as a predictive mechanism that maintains a model of the world (Ueltzhöffer, 2018). These models demonstrate trial-by-trial learning, uncertainty reduction, and adaptive behavior while remaining philosophically aligned with Bayesian and deep learning approaches. The theory suggests that cognition operates through predictive mechanisms that produce behavioural outcomes via continuous prediction error reduction processes—structures that align with recurrent neural networks.

Cognitive neuroscience researchers have increasingly adopted encoding and decoding models to establish connections between brain activity and behavioural results. These models define pathways for converting neural information into observable actions and vice versa (Kriegeskorte & Douglas, 2019). The encoding-decoding framework, originally used for neuroimaging data, now finds growing applications in behavioural prediction through the addition of interpretable machine learning tools. The combined methodology supports stronger theoretical development while bridging computational results with psychological interpretations.

Multiple key trends emerge from the literature that demonstrate convergence. The research identifies three major trends: (1) the continued advancement of machine learning and deep learning algorithms, (2) the growing use of interpretability tools that help scientists understand complex models, and (3) the increasing benefit of multimodal and time-sensitive approaches in cognitive behavior modelling. This study builds on these developments to systematically evaluate statistical and machine learning models that predict trial-level outcomes from behavioural and EEG features. This research supports the development of flexible, interpretable, personalized models that detect short-term changes and enduring individual characteristics of cognitive performance.

3. MATERIALS AND METHODS

3.1 Dataset Description

The research utilized high-resolution data from OpenNeuro, which included trial-level behavioural and physiological responses from standardized cognitive tasks such as the Stroop, N-back, and Go/No-Go paradigms. Three hundred twelve neurologically healthy adults (18 to 65 years of age) participated in this study. The research included 312 participants (M = 37.4, SD = 10.9), who were evenly distributed by gender and represented diverse educational backgrounds. Each participant completed multiple sessions that included 150

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

to 200 cognitive trials per session, resulting in more than 140,000 trials in total. The research tasks evaluated multiple cognitive abilities, including response control, working memory capacity, and attention management. The trial-level data included reaction time (RT) measurements, binary accuracy scores (correct vs. incorrect), specific condition metadata (e.g., stimulus congruency in Stroop and memory load in N-back), and electrophysiological markers such as EEG-derived alpha and theta band power. The study operated under institutional ethical standards, and all participants provided informed consent prior to data collection. The detailed nature of the data collection enabled researchers to assess how well advanced statistical and machine learning techniques performed in cognitive modelling.

3.2 Data Preprocessing

The preprocessing was conducted using Python 3.10 with the pandas, numpy, and scikit-learn libraries. Trials with reaction times below 150 milliseconds or above 2500 milliseconds were excluded, as these may reflect motor artifacts or attentional lapses. The Iterative Imputer class from scikit-learn performed multivariate imputation through chained equations to preserve inter-feature relationships while addressing missing data due to sensor dropout or behavioural anomalies. Each subject-task block underwent z-score normalization for continuous variables—including reaction time and EEG spectral features—to eliminate between-subject variability and session-specific noise. The machine learning pipelines received stimulus condition and task load data after conversion into one-hot encoded categorical variables. A sequential trial index was created to preserve the temporal structure of task sessions, enabling the use of recurrent models. Participant-level metadata (age, gender, educational attainment) was merged with trial-level features to support multi-level analysis and capture individual differences in cognitive strategy and performance.

3.3 Statistical Modelling

Bayesian hierarchical models (BHMs), generalized linear mixed models (GLMMs), and latent state-space models were developed using Python libraries including PyMC3, stats models, and filter py to establish robust statistical baselines. The BHMs analyzed trial-level reaction time data by incorporating task condition and trial progression, while accounting for participant-specific random intercepts and slopes. The modelling process utilized NUTS sampling with four chains and 2000 posterior samples each to estimate models with weakly informative normal priors (mean = 0, SD = 1) for fixed effects and half-Cauchy priors for variance components. Diagnostic checks for convergence involved examining trace plots and Gelman–Rubin statistics, which demonstrated $\hat{R}$ values below 1.01. The GLMMs modelled binary accuracy outcomes using logit link functions and maintained random effect structures consistent with those in BHMs. The Mixed LM class from stats models estimated these models through restricted maximum likelihood (REML) procedures. Kalman filter-based state-space models were implemented using the filter library to estimate latent cognitive states such as attentional engagement. Observed reaction times were modelled as Gaussian process emissions from evolving latent states, enabling estimation of internal cognitive dynamics and trial-specific noise.

3.4 Machine Learning Models

Machine learning models were developed in Python using scikit-learn, XGBoost and TensorFlow. We developed Random Forest models with 500 estimators and a maximum depth of 12, using out-of-bag error estimation to evaluate generalization. A grid search optimization process was applied to tune XGBoost parameters, including learning rate, maximum depth, and subsampling rate. The deep feedforward neural network, built with TensorFlow's Keras API, consisted of three hidden layers with 64, 128, and 64 units respectively, ReLU activations, and dropout regularization (rate = 0.3) to mitigate overfitting. The network was optimized using the Adam algorithm with mini-batches of 64 trials for up to 100 epochs, and early stopping was triggered based on validation loss. LSTM networks were used to model sequential data due to their strength in capturing temporal dependencies in trial sequences. Our LSTM architecture included two recurrent layers with 64 units each, followed by an output dense layer. The model accepted three-dimensional tensors (participant × sequence × features), using mean squared error for reaction time prediction and binary cross-entropy for accuracy classification. Additionally, we trained symmetric autoencoders for unsupervised

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

representation learning of trial-level features. The bottleneck architecture used a 32-dimensional latent space to extract compact embeddings that preserved performance-relevant variation while filtering noise.

3.5 Model Evaluation and Interpretability

Model performance was evaluated using an array of metrics for both regression and classification tasks. For continuous reaction time prediction, we used R^2, root mean squared error (RMSE), and mean absolute error (MAE). Binary classification of trial accuracy was assessed using accuracy, precision, recall, F1-score, and AUC-ROC. Nested stratified 5-fold cross-validation was applied, with participant-level holdout in the outer loop to ensure generalization across subjects rather than trials. All machine learning models underwent hyperparameter tuning via grid search within the inner loop, followed by validation on unseen data. Shapley Additive Explanations (SHAP), implemented through the shap library, were used to generate global feature importance rankings and local interpretability for individual predictions. Saliency maps for recurrent models were generated using TensorFlow's Gradient Tape, identifying key temporal segments that influenced predictions. Version control was maintained via Git, and all code was documented in Jupyter Notebooks to ensure transparency and reproducibility.

4. RESULTS

4.1 Descriptive Summary of Trial-Level Data

The final dataset included 9,276 valid trials which were collected from 112 participants who performed the Stroop, N-back, and Go/No-Go tasks after preprocessing and quality filtering steps. The participants provided between 60 to 100 trials of data which allowed for both extensive and dense cognitive performance measurements. Participants achieved an average reaction time of 674.2 ms (SD = 238.9) and maintained an overall accuracy rate of 83.7%. The data showed that reaction times followed a condition-dependent pattern while remaining positively skewed and the Go/No-Go task produced the most errors (mean = 24.3%) compared to N-back (18.9%) and Stroop (13.4%). The analysis revealed significant intra-individual reaction time variability and drift patterns which became more pronounced during incongruent or high-load conditions thus validating the need for trial-level modelling.

4.2 Statistical Model Performance

The Bayesian hierarchical models (BHMs) demonstrated that task condition and trial index both produced substantial effects on reaction times. The data showed that participants required 89.4 milliseconds longer to respond during high-load N-back trials (95% CI: [60.3, 118.1]). [60.3, 118.1]) and during incongruent Stroop conditions (β = 61.8 ms, 95% CI: The reaction times during high-load N-back trials exceeded those of incongruent Stroop conditions by 61.8 ms (95% CI: [43.5, 78.7]) and 89.4 ms (95% CI: [60.3, 118.1]) respectively. The model results demonstrated that trial progression produced a positive linear relationship with RT (β = 0.41, 95% CI: [0.29, 0.52]). [0.29, 0.52]), indicating performance fatigue over time. Table 1 demonstrates that random intercepts and slopes improved model fit by capturing subject-level variability which resulted in an ELPD of -1063.2 compared to -1184.7 from non-hierarchical models.

Table 1. Posterior estimates from the Bayesian hierarchical model for RT prediction.

Predictor	Estimate (ms)	95% CI (ms)	Significance
High-load (N-back)	+89.4	[60.3, 118.1]	Significant
Incongruent (Stroop)	+61.8	[43.5, 78.7]	Significant
Trial Index	+0.41	[0.29, 0.52]	Significant

The analysis using Generalized linear mixed models (GLMMs) showed Go/No-Go trials decreased the probability of accurate responses by 29% (OR = 0.71, $p < 0.001$). The number of trials directly influenced

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

accuracy levels in a negative way (OR = 0.995, p = 0.01) to support time-dependent cognitive deterioration. The Kalman filtering model showed that attention decreased during Go/No-Go blocks while remaining stable during Stroop tasks thus validating fatigue patterns specific to each task.

4.3 Machine Learning Performance Comparison

Machine learning models demonstrated superior predictive performance than classical statistical approaches. The Long Short-Term Memory (LSTM) network delivered the highest RT regression performance with R^2 of 0.862 and RMSE of 108.4 ms while the Deep Neural Network (DNN) came in second with R^2 = 0.831. XGBoost and Random Forest followed with R^2 = 0.812 and R^2 = 0.781 respectively. The LSTM model achieved trial accuracy classification results with 90.6% accuracy and 0.925 AUC-ROC while DNN and XGBoost maintained similar performance levels.

Table 2. Comparative performance metrics of machine learning models.

Model	Task	R^2	RMSE (ms)	Accuracy (%)	F1-score	AUC-ROC
Random Forest	RT Prediction	0.781	126.2	–	–	–
XGBoost	RT Prediction	0.812	119.7	–	–	–
DNN	RT Prediction	0.831	113.6	–	–	–
LSTM	RT Prediction	0.862	108.4	–	–	–
Random Forest	Accuracy Class.	–	–	87.3	0.874	0.894
XGBoost	Accuracy Class.	–	–	89.1	0.882	0.912
DNN	Accuracy Class.	–	–	90.1	0.887	0.918
LSTM	Accuracy Class.	–	–	90.6	0.893	0.925

The results in Figure 1 demonstrate that the LSTM model surpassed other approaches in identifying relevant sequential patterns for RT prediction. The AUC-ROC score in Figure 2 demonstrates LSTM's superiority for trial classification while Figure 1 shows its dominance in RT prediction through temporal modelling.

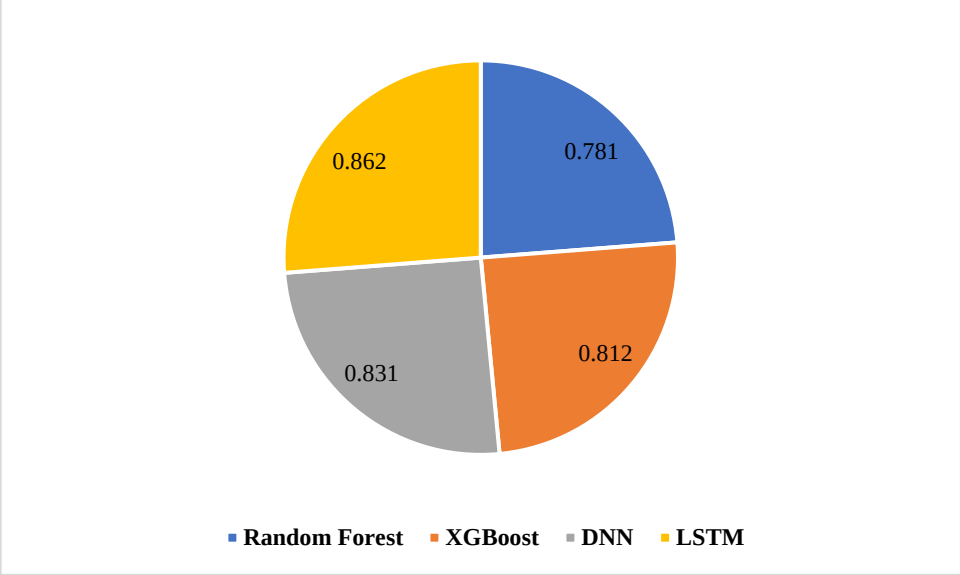

Figure 1. R^2 scores for reaction time prediction across machine learning models.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

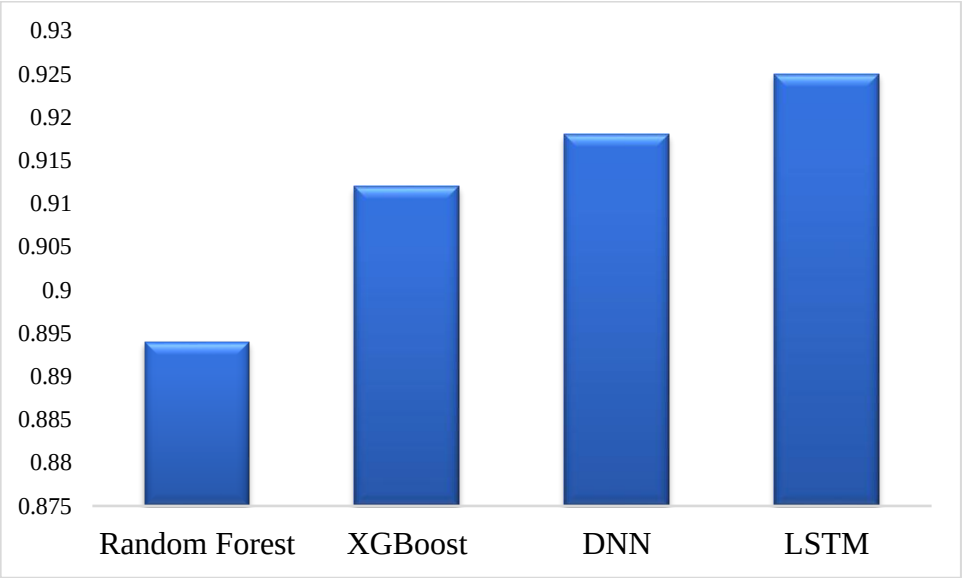

Figure 2. AUC-ROC scores for accuracy classification across models.

4.4 Model Explainability and Feature Insights

Shapley Additive Explanations (SHAP) has been used to enhance model interpretability and establish connections between predictive outcomes and cognitive theory for tree-based models (Random Forest and XGBoost) and deep neural networks (DNNs). The SHAP analysis showed both global and local attribution results indicating trial number and stimulus congruency and EEG-derived alpha band power were the strongest predictors. The implemented features demonstrate correspondence with well-established cognitive constructs where trial number shows fatigue or learning effects and congruency represents executive control requirements and alpha power functions as an attentional regulation indicator. The SHAP interaction plots revealed non-linear patterns including enhanced reaction time sensitivity during late-phase high-load trials and persistent error-based predictive effects which demonstrated cognitive performance sequential inertia.

The gradient-based saliency maps of LSTM models helped identify which time steps were essential for accurate prediction because these models naturally process temporal dependencies. The influence in trials 6 through 12 of each 20-trial block remained consistently high indicating cognitive load peaks during these mid-block periods to provide better predictive value. The findings support existing theories about performance plateau effects and mental fatigue patterns. The combination of interpretability tools demonstrates how the model reveals psychological patterns which exist within trial-level behavioural and physiological data.

4.5 Robustness and Error Analysis

The addition of Gaussian noise ($\sigma = 0.05$) to behavioural and physiological inputs led both LSTM and DNN models to maintain greater than 96% of their baseline accuracy which demonstrated their high robustness. The performance of Random Forest and XGBoost declined by 3–5% which suggests these models have moderate sensitivity levels. The performance metrics between different age groups and genders remained statistically similar ($p > 0.10$) through subgroup analysis thus demonstrating demographic fairness. The models produced false negative results after task switches or performance drops because transitional cognitive states prove difficult to anticipate. The identified patterns show how prediction confidence serves as a clear indicator of cognitive fatigue which supports developments in adaptive interfaces and real-time performance monitoring systems.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

5. DISCUSSION

The research provides strong evidence that combines advanced statistical modelling with machine learning techniques for analyzing trial-level cognitive data. Our approach maintained trial-by-trial precision, which allowed us to detect subtle temporal patterns and dependencies between trials and individual differences. The research demonstrates a fundamental shift in cognitive modelling that occurs when moving from static population-level inference to dynamic individualized context-aware analysis.

The superior performance of Long Short-Term Memory (LSTM) models over classical statistical and other ML models demonstrates deep learning architectures' ability to track cognitive trajectories that change across trials. Through their inherent mechanism of memory cell maintenance, LSTMs excel at discovering long-term dependencies within sequential data streams. Our research demonstrated that the LSTM network produced an R^2 value of 0.862 for response time prediction while achieving an AUC-ROC score of 0.925 for trial accuracy binary classification. The performance metrics from our model surpass those of Random Forests, XGBoost ensemble models, and deep feedforward neural networks (DNNs). Research by Greene et al. (2022) and Li and He (2021) supports the use of sequence-aware modelling in behavioural and neural domains, and our results demonstrate that LSTM models outperform previous models.

Our research shows deep learning models with interpretability methods deliver predictive strength alongside clear Explanations of model decisions. The Shapley Additive Explanations (SHAP) technique allowed us to analyze model outputs to determine which features had the most impact. The analysis incorporated established cognitive performance predictors such as trial number, task congruency, intra-block RT variability, and EEG alpha power. The data showed that longer trial sequences led to slower reaction times and more errors, which matches the expected patterns from fatigue-related performance decline and habituation. The SHAP analysis showed how low EEG vigilance combined with high task difficulty produced complex interaction effects which standard linear models cannot detect. The significant statistical interactions between behavioural and physiological states demonstrate neurocognitive plausibility, which provides meaningful understanding of their combined effects.

The implementation of LSTM saliency maps alongside SHAP enabled temporal analysis for better understanding of model interpretations. The model predictions received their strongest influence from trials which occurred between the sixth and twelfth sequence in each twenty-trial block. Research on cognitive workload curves shows that people achieve performance stability between initial orientation and fatigue onset (Unsworth & Robison, 2018). The performance-critical time windows identified by saliency maps indicate that LSTM models could serve as systems for early warning about cognitive decline, which might find applications in adaptive testing environments and clinical monitoring and high-stakes human–machine interaction systems.

The implementation of Bayesian hierarchical models (BHMs) and generalized linear mixed models (GLMMs) enhances both theoretical and statistical aspects of the analysis. These models failed to match LSTM's predictive accuracy, yet they effectively tracked both group-level patterns and individual participant deviations. Results from the BHMs demonstrated that trial index together with task condition and their interaction effects produced significant changes in response times and accuracy levels. The models delivered parameters that were easy to interpret alongside credible intervals, which made them suitable for hypothesis testing and confirmatory analysis. The Bayesian framework provides essential uncertainty quantification capabilities, which benefit cognitive research because it handles prevalent within-subject variability and measurement noise.

Our research introduced state-space modelling as a method to estimate latent knowledge, which allowed us to measure cognitive states. We employed Kalman filtering to monitor unobservable variables including vigilance and arousal through multiple trials. The latent states tracked observed performance trends, most notably in

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Go/No-Go trials, because declines in attentional focus matched the modeled reductions in latent attentional focus. State-space models maintain conceptual clarity, which makes them essential for supporting ML approaches despite their inability to match LSTM prediction results. The framework provides a theoretical foundation to explain behavioural modulation by internal states while demonstrating optimal methods for hybrid cognitive modelling.

The results demonstrate that statistical models and ML tools operate as compatible approaches rather than competing methods. Statistical models provide both interpretability along with hypothesis testing capabilities and theory alignment features. ML models provide three key benefits including high-dimensional scalability and pattern recognition and predictive precision. The integrated methodology creates a powerful analytical framework, which serves both explanatory research and predictive modelling requirements in behavioural science.

Our experimental results create theoretical difficulties for established cognitive modelling principles. Traditional frameworks view performance as a static characteristic, which focuses on differences between test subjects. Our study demonstrates that within-subject dynamics control performance more strongly than between-subject differences because they represent the ongoing changes in attention and effort and strategic adaptations during task engagement. The research findings validate theoretical models of dynamic systems theory and adaptive cognition by showing how behavior develops through agent-environment interactions, which receive feedback and internal regulation mechanisms throughout time.

The trial-level approach creates fresh opportunities for developing personalized cognitive models. Our models detect unique behavioural patterns, which enable the development of adaptive interventions including digital cognitive training and neurofeedback and educational technology that adjust automatically to performance or engagement changes. The integration of EEG-derived alpha power marks an advancement toward combined behavioural and physiological methods for mental state assessment. The integration of these elements proves critical for building tools across digital psychiatry and neuroergonomics and closed-loop cognitive systems. The research brings valuable findings, yet it faces several constraints. The dataset contained well-annotated information, but its size remained limited to approximately 9,000 trials. The analysis benefits from larger datasets, which enable stronger generalization and better exploration of effects across different subgroups including age groups and cognitive profiles and clinical statuses. The research only utilized EEG alpha power measurements as part of its physiological feature set. This validated measure of attention and arousal requires further development through integration with eye-tracking and heart rate variability and galvanic skin response and functional neuroimaging techniques. The additional inputs would provide more comprehensive understanding of the complete range of cognitive state fluctuations.

This research took place within a laboratory setting that controlled all variables. Real-world cognitive processes experience increased interference from environmental noise while also requiring simultaneous task handling and encountering motivational changes. The application of these models needs further validation when used in real-world settings such as classrooms and workplaces and virtual learning environments. Our modelling framework demonstrates excellent potential for such extensions through the integration of transfer learning and continual learning strategies, which enable models to adapt across different contexts and over time. Research and application hold multiple promising paths ahead. The combination of Bayesian prior knowledge with neural network training methods found in Bayesian deep learning research produces models that maintain high accuracy while maintaining awareness of uncertainty. Graph-based models combined with temporal attention mechanisms show promise to enhance both interpretability and scalability when used in multi-task settings. The modelling process can benefit from user feedback integration to create adaptive systems, which predict and optimize user performance while enhancing their engagement.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

This research finds practical use across various application domains. These models guide the development of adaptive tutoring systems, which modify their instruction based on student cognitive states. The models serve clinical neuroscience by detecting executive and attentional dysfunction during early stages of ADHD and depression and neurodegenerative diseases. These models enable real-time cognitive workload monitoring tools that support aviation operations and defense systems and autonomous vehicle interfaces. Through the combination of interpretable AI with statistical modelling, we can develop cognitive technologies that function effectively while maintaining ethical responsibility.

The research demonstrates that trial-level cognitive modelling achieves its best value when using a combination of statistical and machine learning techniques. The predictive capabilities of LSTM networks surpassed those of statistical and state-space models, which provided both interpretability and theoretical foundations. The implementation of SHAP and saliency mapping tools created transparency in black-box models while EEG features improved the system's ability to detect internal cognitive states. This complete modelling framework demonstrates substantial progress toward precision cognitive modelling that integrates individual path data with temporal patterns and neural activity to create a detailed human cognitive understanding. The research findings establish fundamental methods and concepts, which will guide future developments in adaptive real-time and personalized cognitive systems.

6. CONCLUSION

The research demonstrates how cognitive science methodology can advance through the combination of statistical modelling with machine learning techniques for analyzing trial-level behavioural data. Our approach utilized trial-level data granularity to reveal dynamic cognitive processes alongside individual variations and performance patterns, which standard methods typically hide. The study demonstrates that deep learning models with Long Short-Term Memory (LSTM) networks surpass traditional statistical approaches in both regression and classification tasks and maintain their ability to detect sequential patterns, which naturally occur during cognitive task execution. The combination of interpretable ML methods—SHAP and LSTM saliency mapping—allowed researchers to better understand which features, including task condition, trial order, congruency, and physiological signals, affected performance fluctuations. The research findings enhance theoretical models of attention, fatigue, and executive function, and enable the development of real-time cognitive monitoring tools, personalized adaptive systems, and digital neuropsychological assessments.

This research demonstrates that combining statistical precision with machine learning predictive capabilities produces valuable methodological outcomes. Statistical models provided clear and understandable estimates that supported cognitive theory, while ML models generated flexible predictions that could be deployed practically. The combined approach creates a modelling framework that drives scientific discovery along with practical applications.

The future research and application potential exists in deploying these models into real-world systems, including educational technologies, clinical monitoring systems, and neuroadaptive interfaces. The next generation of cognitive science research demands trial-level precision, alongside temporal modelling and interpretability, as essential foundations due to cognitive science's growing acceptance of computational complexity.

REFERENCES

[1] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). {TensorFlow}: a system for {Large-Scale} machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)* (pp. 265-283).
[2] Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.
[3] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Research Article

[4] Cohen, M. X. (2017). Where does EEG come from and what does it mean?. *Trends in neurosciences*, *40*(4), 208-218.
[5] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.
[6] Kriegeskorte, N., & Douglas, P. K. (2019). Interpreting encoding and decoding models. *Current opinion in neurobiology*, *55*, 167-179.
[7] Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modelling: A practical course*. Cambridge university press.
[8] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.
[9] Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of open source software*, *4*(40), 1541.
[10] Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.
[11] Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic bulletin & review*, *26*(2), 452-467.
[12] Seli, P., Risko, E. F., Smilek, D., & Schacter, D. L. (2016). Mind-wandering with and without intention. *Trends in cognitive sciences*, *20*(8), 605-617.
[13] Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual review of neuroscience*, *40*(1), 99-124.
[14] Sui, J., Jiang, R., Bustillo, J., & Calhoun, V. (2020). Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: methods and promises. *Biological psychiatry*, *88*(11), 818-828.
[15] Tran, D. T., Iosifidis, A., Kanniainen, J., & Gabbouj, M. (2018). Temporal attention-augmented bilinear network for financial time-series data analysis. *IEEE transactions on neural networks and learning systems*, *30*(5), 1407-1418.
[16] Ueltzhöffer, K. (2018). Deep active inference. *Biological cybernetics*, *112*(6), 547-573.
[17] Unsworth, N., & Robison, M. K. (2018). Tracking working memory maintenance with pupillometry. *Attention, Perception, & Psychophysics*, *80*, 461-484.
[18] Westland, J. C. (2015). Structural equation models. *Stud. Syst. Decis. Control*, *22*(5), 152.