University of Central Lancashire UCLan

Central Lancashire Online Knowledge (CLoK)

Title	Psychophysiological models of hypovigilance detection: A scoping review
Type	Article
URL	https://clok.uclan.ac.uk/47469/
DOI	https://doi.org/10.1111/psyp.14370
Date	2023
Citation	Marois, Alexandre, Kopf, Maëlle, Fortin, Michelle, Huot-Lavoie, Maxime, Martel, Alexandre, Boyd, J. Gordon, Gagnon, Jean-François and Archambault, Patrick M. (2023) Psychophysiological models of hypovigilance detection: A scoping review. Psychophysiology. ISSN 0048-5772
Creators	Marois, Alexandre, Kopf, Maëlle, Fortin, Michelle, Huot-Lavoie, Maxime, Martel, Alexandre, Boyd, J. Gordon, Gagnon, Jean-François and Archambault, Patrick M.

It is advisable to refer to the publisher's version if you intend to cite from the work. https://doi.org/10.1111/psyp.14370

For information about Research at UCLan please go to http://www.uclan.ac.uk/research/

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the http://clok.uclan.ac.uk/policies/

Article

Psychophysiological models of hypovigilance detection: A scoping review

Marois, Alexandre, Kopf, Maëlle, Fortin, Michelle, Huot-Lavoie, Maxime, Martel, Alexandre, Boyd, J. Gordon, Gagnon, Jean-François and Archambault, Patrick M.

Available at https://clok.uclan.ac.uk/47469/

Marois, Alexandre, Kopf, Maëlle, Fortin, Michelle, Huot-Lavoie, Maxime, Martel, Alexandre, Boyd, J. Gordon, Gagnon, Jean-François and Archambault, Patrick M. (2023) Psychophysiological models of hypovigilance detection: A scoping review. Psychophysiology . ISSN 1469-8986

It is advisable to refer to the publisher's version if you intend to cite from the work.
http://dx.doi.org/10.1111/psyp.14370

For more information about UCLan's research in this area go to http://www.uclan.ac.uk/researchgroups/ and search for <name of research Group>.

For information about Research generally at UCLan please go to http://www.uclan.ac.uk/research/

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the policies page.

CLoK
Central Lancashire online Knowledge
www.clok.uclan.ac.uk

Received: 30 January 2023 | Revised: 31 May 2023 | Accepted: 1 June 2023

DOI: 10.1111/psyp.14370

PSYCHOPHYSIOLOGY SPR WILEY

REVIEW

Psychophysiological models of hypovigilance detection: A scoping review

Alexandre Marois[1,2] | **Maëlle Kopf**[1] | **Michelle Fortin**[3] | **Maxime Huot-Lavoie**[3] | **Alexandre Martel**[3] | **J. Gordon Boyd**[4,5] | **Jean-François Gagnon**[1] | **Patrick M. Archambault**[3,6,7]

[1]Thales Research and Technology Canada, Quebec City, Québec, Canada

[2]School of Psychology and Computer Science, University of Central Lancashire, Preston, Lancashire, United Kingdom

[3]Faculty of Medicine, Université Laval, Quebec City, Québec, Canada

[4]Department of Medicine, Queen's University, Kingston, Ontario, Canada

[5]Kingston General Hospital, Kingston, Ontario, Canada

[6]Centre de recherche intégrée pour un système apprenant en santé et services sociaux, Centre intégré de santé et de services sociaux de Chaudière-Appalaches, Lévis, Québec, Canada

[7]VITAM - Centre de recherche en santé durable, Centre intégré universitaire de santé et de services sociaux de la Capitale-Nationale, Quebec City, Québec, Canada

Correspondence
Patrick M. Archambault, Université Laval, Pavillon Ferdinand-Vandry, 1600 Avenue des Sciences-de-la-Vie, Quebec City, Québec G1V 5C3, Canada.
Email: patrick.archambault@fmed.ulaval.ca

Funding information
Mitacs, Grant/Award Number: IT16351; Université Laval; Fonds de recherche du Québec - Santé, Grant/Award Number: 283211

Abstract

Hypovigilance represents a major contributor to accidents. In operational contexts, the burden of monitoring/managing vigilance often rests on operators. Recent advances in sensing technologies allow for the development of psychophysiology-based (hypo)vigilance prediction models. Still, these models remain scarcely applied to operational situations and need better understanding. The current scoping review provides a state of knowledge regarding psychophysiological models of hypovigilance detection. Records evaluating vigilance measuring tools with gold standard comparisons and hypovigilance prediction performances were extracted from MEDLINE, PsychInfo, and Inspec. Exclusion criteria comprised aspects related to language, non-empirical papers, and sleep studies. The Quality Assessment tool for Diagnostic Accuracy Studies (QUADAS) and the Prediction model Risk Of Bias ASsessment Tool (PROBAST) were used for bias evaluation. Twenty-one records were reviewed. They were mainly characterized by participant selection and analysis biases. Papers predominantly focused on driving and employed several common psychophysiological techniques. Yet, prediction methods and gold standards varied widely. Overall, we outline the main strategies used to assess hypovigilance, their principal limitations, and we discuss applications of these models.

KEYWORDS

ECG, EEG, eye tracking, hypovigilance, scoping review, state prediction models

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2023 The Authors. *Psychophysiology* published by Wiley Periodicals LLC on behalf of Society for Psychophysiological Research.

1 | INTRODUCTION

Sleepiness is a major contributor to many accidents and hazardous situations in several domains (e.g., Lyznicki et al., 1998; Philip & Akerstedt, 2006; Tefft, 2010). Estimations point out that it is involved in at least 15%–20% of all accidents in transport operations (Akerstedt, 2000; Connor et al., 2002; Horne & Reyner, 1999). As such, mental fatigue and sleepiness can importantly compromise safety and integrity of individuals and infrastructures, especially in high-stake situations such as in complex and safety-critical environments. This can be explained by the consequences of sleepiness and mental fatigue on human performance. In fact, mental fatigue and sleepiness have important impacts on perception, attention, decision-making, and judgment, and can lead to slower reaction times, misjudgments, and inferior detection of critical elements within one's environment (e.g., Carretta & French, 2012; Gunzelmann & Gluck, 2009; Guo et al., 2016; Lopez de la O et al., 2012; see Abd-Elfattah et al., 2015, for a review).

Although fatigue and sleepiness are sometimes considered the same phenomenon, some distinctions exist. As outlined by Salvati et al. (2021), sleepiness (or drowsiness) represents an intermediate progressive state between an awakening state and sleep, which is related to altered awareness and to a desire to sleep (Mehreen et al., 2019; Slater, 2008). It is a normal transitional state but it can also be caused by sleep-related problems such as lack of sleep, poor sleep quality, or circadian rhythm disequilibrium (May & Baldwin, 2009). Fatigue rather represents a larger phenomenon. It is a consequence of either physical or mental work, and is construed as a reluctance—and a difficulty—to pursue and focus on a given task (Boksem & Tops, 2008; Brown, 1982). Vigilance is "the capability to be aware of relevant, unpredictable changes in one's environment, irrespective of whether or not such changes occur" (van Schie et al., 2021, p. 178). This scoping review aims at providing a portrait of the literature related to hypovigilance and, more particularly, on sensing methods to assess this phenomenon for operational applications. For the sake of parsimony and because the current paper is mainly interested in the observable effects of mental fatigue, we hereafter focus on the concept of hypovigilance as an integrative concept at the center of fatigue, drowsiness, and sleepiness.[1]

[1]One could argue that referring to hypovigilance to discuss a large variety of phenomena such as fatigue, drowsiness, and sleepiness may represent an important generalization. Literature on these subjects is vast and we acknowledge that distinctions indeed exist between hypovigilance-related phenomena induced, for example, by cognitive resource depletion, circadian rhythm, boredom, or sedation. Yet, it still remains unclear how all of these concepts are related to each other and to what extent they can be assessed using common methods in a real-world setting. Here, the scoping review approach allows to address this question without any a priori from a larger perspective in order to draw the lines around common observations and gaps in the literature on hypovigilance. Since the literature on the subject is broad, this allows us to cover a broader initial scope as a first step toward identifying best ways to monitor vigilance in several real-world applied situations.

It allows increasing the scope to not only discuss biological effects induced by homeostatic- and circadian-related phenomena, but also situational consequences of mental effort, monotony, and time on task.

In operational contexts, effects of hypovigilance (either induced by sleepiness or fatigued mental/physical states) can be observed via key domain-specific performance indicators. In aviation, studies have outlined that hypovigilance can lead to in-flight error-making (Aljurf et al., 2018; Gregory et al., 2010), inferior situation awareness, longer reaction times and increased distractibility (Miller & Melfi, 2006), visual and auditory perception impairment (Dehais et al., 2014; Previc et al., 2009; Russo et al., 2005), and to reduced cognitive flexibility and hand-eye coordination (O'Hagan et al., 2018). In driving studies, evidence of increase in reaction time (Guo et al., 2016; Liu et al., 2012), reduced time headway (i.e., between-vehicles duration; Fuller, 1983; Zhang et al., 2016), and increased lateral deviation errors and variability (Brookhuis & De Waard, 1993; Matthews & Desmond, 2002; Philip et al., 2003) among hypovigilant drivers were also largely reported. Hypovigilance is even associated with difficulties in takeover performance in automated driving situations (Jarosch et al., 2019; Matthews et al., 2019). The consequences of hypovigilance can also be observed in non-driving domains such as command and control operations, i.e., occupations entailing providing key information and orders for security operations such as emergency management, police or firefighting operations, and surveillance (e.g., Carretta & French, 2012). In the last few decades, the role of human operators has constantly evolved with the emergence of automation, shifting toward systems supervision and the management of malfunctions and unusual events (Parasuraman, 1986; Sheridan, 1987). Consequently, vigilance still remains a key asset for many operational domains including but not limited to military surveillance, industrial quality control, robot manufacturing, seaboard navigation, and transportation (Warm et al., 1996). Vigilance is also a key capacity that can be altered by many organic brain syndromes, such as delirium (American Psychiatric Association, 2013).

1.1 | Measuring hypovigilance

Currently, one of the key strategies in the management of hypovigilance in operational contexts is sleep (Petrilli et al., 2006). An important part of the accountability remains with the operators, which typically have to report—and manage their performance on task—when they find themselves in a hypovigilant state. Nevertheless, this phenomenon is still highly prevalent (e.g., between 68% and 91% of commercial airline pilots still experience fatigue; Aljurf et al., 2018; Jackson & Earl, 2006). To counter this problem, alternative methods must be developed to help individuals better monitor their own vigilance level and, ultimately, to reduce potential consequences for the safety and integrity of populations and infrastructures. In fact, there have been calls for the development of new strategies to better monitor vigilance, such as the European New Car Assessment Programme (EuroNCAP). In its 2025 Roadmap, EuroNCAP recommends that driver-state monitoring is a key and priority part of safety assessments (EuroNCAP, 2017). According to Schwarz and Fuchs (2018), adaptive systems are also essential in better-supporting operators in human-machine systems to mitigate high-risk user states and performance decrements. From their standpoint, the different user states useful to monitor in complex and safety-critical work environments include, among others, attention and fatigue.

Different approaches can be taken to monitor human states, and more particularly hypovigilance (see, e.g., Kerick et al., 2013; Oken et al., 2006). First, measures of task performance can be used. This approach relies on the identification of signs of hypovigilance, that is, the behavioral manifestation of a reduced ability to focus on the main task. For instance, in a driving context, missing traffic signals, tailgating, swerving and crossing lanes can be used to assess a driver's hypovigilant and distracted state (Kashevnik et al., 2021). Second, performance on a secondary task can also be used to evaluate hypovigilance levels while performing a primary task (either concurrently or in alternation, at given intervals). The Psychomotor Vigilance Task (PVT; Dinges et al., 1997; Dinges & Powell, 1985; Doran et al., 2001; Lim & Dinges, 2008) is a common way to measure behavioral alertness wherein one must react as quickly as possible to the simple presentation of a stimulus occurring at random interstimulus intervals. It is used in laboratory settings but has also real-life applications (e.g., letters attention test for diagnosing delirium; Ely, Gautam, et al., 2001). This task can be used as a unique test or added while a person is performing another task, hence providing information on how simple reaction time to the stimulus evolves as a function of time/effort on the primary task (e.g., Buckley et al., 2016; Dinges et al., 1998). Third, subjective measures have also been used in certain contexts, wherein operators report their own (self-perceived) level of drowsiness or vigilance (e.g., Dorrian et al., 2008; Luna et al., 2022). However, one of the limitations of these preceding techniques (i.e., behavioral and self-reported subjective measures) is that they are not specific to hypovigilance. Indeed, behavioral measures (e.g., performance disruption from a given task) represent the product of processing neural networks from task stimulus detection to motor reaction (e.g., Hughes & Marsh, 2017). During this process, factors such as motivation and emotional states can have an impact on behavior (Pessoa, 2009). Consequently, although they do relate to one's hypovigilance level, both behavioral and self-reported measures of hypovigilance can lack validity and specificity because of the different confounding variables that might modulate them.

A fourth strategy to measure hypovigilance concerns the collection and analysis of psychophysiological proxies (e.g., Boudaya et al., 2020; Parasuraman et al., 1998; Rush et al., 2019; Sahayadhas et al., 2015). This technique relies on measures of the physiological activity of an operator—either of the central or the peripheral nervous systems—to estimate one's level of sustained attention (vigilance) deployment. The rationale behind this approach lies in the significant implication of the locus coeruleus-norepinephrine (LC-NE) system for attention-related activities. Activity of this system has been largely related to vigilance, attention orienting, arousal, and to the sleep–wake cycle (e.g., Aston-Jones & Cohen, 2005; Bouret & Sara, 2004; Nieuwenhuis et al., 2011; Rajkowski et al., 2004; Southwick et al., 1999). NE is secreted across the brain in multiple areas including cerebral cortex, limbic structures, diencephalon, midbrain, and spinal cord (e.g., Miller & Cohen, 2001; Nieuwenhuis et al., 2005; Sara & Bouret, 2012). Its secretion from the pons-located LC in these brain structures makes synapse appositions with postsynaptic specializations on target neurons, hence generating further electric activity in the brain (Marzo et al., 2014; Papadopoulos & Parnavelas, 1990). Consequences of such specialized activity enhance the selectivity of certain neurons to specific targets and increase the signal-to-noise ratio to allow preferential processing of the stimuli presented to the system (Foote et al., 1975; Waterhouse et al., 1998). Peripheral sympathetic activity increase (and concurrent parasympathetic activity decrease) has also been reported (e.g., Elam et al., 1986; Sara & Bouret, 2012; Wang & Munoz, 2015), ensuing from the multiple efferent projections of the LC-NE system in the brain. Taken together, this means that multiple psychophysiological proxies of the (hypo)vigilant state can be collected via measures of the central nervous system and of the peripheral nervous system.

Multiple models for quantifying hypovigilance or associated concepts (e.g., drowsiness and fatigue) have been developed over the years in laboratory conditions using behavioral and/or physiological correlates of the vigilance level (e.g., Oken et al., 2006). In fact, as outlined above, it is known that a decrease in vigilance is associated with multiple physiological and behavioral manifestations and that measuring such manifestations can provide information on the level of vigilance. Drowsiness and vigilance have, for example, been assessed by measuring the PERCLOS (percentage of eyelid closure over the pupil; e.g., Lin et al., 2012; Sommer & Golz, 2010). Heart rate and respiration rate are also associated with the sleep onset period. These physiological responses can thus be integrated into predictive models to detect hypovigilance for safety purposes (EuroNCAP, 2017; Schwarz & Fuchs, 2018). Multiple reviews have been published to summarize methods for assessing hypovigilance and other related concepts from multiple perspectives (e.g., Arun et al., 2011; Bafna & Hansen, 2021; Bier et al., 2020; Duffy & Feltman, 2022; Larue et al., 2010; Mogilever et al., 2018; Mohanavelu et al., 2017; Sahayadhas et al., 2012). For instance, Bendak and Rashid (2020) provide a systematic review of the causes of fatigue observed in the aviation industry and the ways to measure it. As a result, they outlined many different objective metrics (e.g., fitness-for-duty tests, physiological monitoring, performance monitoring, flight data monitoring) and subjective measures (e.g., self-rating scales, air safety reports, and fatigue prediction). Literature on the different measures of hypovigilance, however, is scattered through different approaches (e.g., ergonomics, engineering, cognitive neuroscience) and research is thus difficult to reconcile.

1.2 | The current study

The goal of this review is to map the state of the current knowledge about the psychophysiological methods for hypovigilance detection. We aim to identify relevant literature regarding the psychophysiological responses identified as proxies for human hypovigilance from a broader perspective, regardless of the specific domain of application, in order to provide the scientific community with a better sense of possible ways for investigating/monitoring hypovigilance. To reach this goal, we performed a systematic scoping review of empirical studies found on several databases that included both diagnostic and prediction studies (with respective detection of hypovigilant state on a given dataset/context vs. prediction of hypovigilant levels with measures that could be generalizable to other datasets/contexts). We chose to conduct a scoping review because of the potentially large scope of the literature emerging from heterogeneous but interconnected disciplines such as medicine, psychology, and engineering. Also, the key concepts underpinning hypovigilance detection from psychophysiological responses remain a rapidly emerging area of study (Munn et al., 2018; Peters et al., 2015) that would benefit from a scoping review to guide future research and development.

2 | METHOD

We conducted our review using the Levac et al. scoping review methodology (Levac et al., 2010) and report our findings using the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) framework (Tricco et al., 2018; see Appendix S1 for the PRISMA-ScR checklist). We did not register our review protocol.

2.1 | Eligibility criteria

Inclusion and exclusion criteria were pre-specified for each step of the selection of sources of evidence. Inclusion criteria were: (1) studies evaluating vigilance measurement tool(s) compared to a gold standard, and (2) studies had to report data about either the accuracy, sensitivity, or specificity of their measurement tools. We had defined a priori a list of accepted gold standards prior to the screening of titles and abstracts. These gold standards were associated with a variety of concepts that are related to hypovigilance. The gold standards were: the Attention Network Test, the AVPU scale, the Fatigue Scale, the SAFTE Model, the Confusion Assessment Method, the Delirium Severity Scale, the Glasgow Coma Scale, the Intensive Care Delirium Screening Checklist, the Karolinska Sleepiness Scale, the PERCLOS, the Psychomotor Vigilance Task, the Psychomotor Vigilance Test, the Ramsay Sedation Scale, the Richmond Agitation-Sedation Scale, the Sour Seven Questionnaire, the Stanford Sleepiness Scale, the Epworth Sleepiness scale, the Maintenance of Wakefulness Test, the Confusion Performance Test, the Recognizing Acute Delirium as Part of Your Routine (RADAR) tool, and electroencephalography studies. We did not a priori determine a specific threshold for each of these gold standards because of the scoping nature of this review and because of lack of consensus in this emerging field of study. Besides, some gold standards may not necessarily possess clear thresholds for determining episodes of hypovigilance (e.g., PERCLOS measures) and this allowed us to include a larger set of studies to better scope current practices in predicting vigilance levels. Moreover, considering the scoping nature of our review, we also allowed additional new gold standards if the authors defined these clearly in the methods of their published manuscripts. For example, video recordings using the Wierwille scale (Wierwille & Ellsworth, 1994) were accepted. The scoping review methodology allows researchers to define post hoc inclusion and exclusion criteria based on new familiarity with the subject matter through reading the identified studies (Levac et al., 2010).

Exclusion criteria were also determined before the research strategy was initiated. Studies that were not in English nor French, that involved irrelevant populations (e.g., animal studies or children), editorials, letters to editor, concepts only, clinical image pieces, and non-scientific publications were excluded. Studies that did not compare a new measurement tool to an accepted gold standard were also excluded for lack of evidence. Sleep and anesthesia studies were also excluded because the subject of interest was rather the variation of vigilance in relation to a task. Studies evaluating mental workload, muscle fatigue, and use of pharmacologic psychostimulants without any measure of vigilance were also rejected. Non-peer-reviewed literature was also rejected. All duplicate publications were removed.

2.2 | Information sources and search strategies

In collaboration with two research librarians from Université Laval, we selected three databases relevant to our study: MEDLINE, PsychInfo, and Inspec, specialized in medicine, psychology, and engineering, respectively. These domains constitute the three main areas of interest for this project. We then built a research strategy with the two information specialists. Our strategy had three main axes: hypovigilance and associated concepts, gold standards for hypovigilance measurement, and potential new physiological measures of hypovigilance (see an example in Table 1). We created an exhaustive list of keywords for each of these domains. The research librarians validated our keywords and adapted our research strategy to the three selected databases.

We thoroughly searched each database for relevant articles published from the inception date of each database (MEDLINE: 1966; PsycINFO: 1967; Inspec: 1967) until April 22nd, 2021. We repeated the search strategy on November 10th, 2021 to make sure our findings were up to date. All the references figuring in the selected articles were manually checked to make sure no additional article was missed. We used the Covidence Systematic Review Software to manage all the review steps (Veritas Health Innovation, Melbourne, Australia). Table 1 presents the full electronic search strategy used for MEDLINE. The research strategies used for PsycINFO and Inspec can be found in Appendix S2.

2.3 | Selection of sources of evidence

We proceeded in a three-step manner with the help of the Covidence Systematic Review Software. First, two teams of reviewers (MHL & AMartel, and MF & MK) independently screened abstracts and titles based on inclusion and exclusion criteria (first step on April 22, 2021; second step on November 10, 2021). To ensure consistency in the application of criteria screening, training sessions were conducted for a set of approximately 100 citations before the reviewers started their independent work. The articles had to be approved by the two reviewers to be included in the next steps. If reviewers disagreed about sorting an article, they met to discuss either in person, by phone, or videoconference. If a consensus could not be reached between the two teams of reviewers, a third reviewer (PMA) made the final decision. Second, reviewers proceeded to another round of screening by applying exclusion criteria to the full texts. The remaining selected studies were then thoroughly analyzed for data extraction and risk of bias assessment.

2.4 | Data charting and collation

Data charting was independently executed by two authors (MF & MK), and reviewed by a third author (AMarois). A calibrated worksheet was set before the data extraction. The two researchers then compared their data extraction. If reviewers disagreed on quality assessment, a third reviewer made the final decision, but this was unnecessary in practice. We did not communicate with the authors to collect missing data because the aim of the study was to evaluate the accessible literature and not the raw data.

For each source, we sought the publication year, article type, and source of funding. We identified which concepts related to hypovigilance were studied in each paper (e.g., drowsiness, sleepiness, or fatigue). We then looked for a definition of the cognitive state studied when available. We extracted the following data from the included studies: (a) number of participants, (b) sex, (c) age, (d) health conditions, (e) study approach (either diagnostic or prediction model); (f) physiological measuring approach employed, (g) domain or context of study, (h) method to induce hypovigilance, (i) experimental task, (j) differences between the two experimental groups, (k) selected gold standard and its prespecified threshold(s) if available, (l) the specific sensors used to collect physiological measures, (m) the specific diagnostic/prognostic physiological measures, (n) statistical model used, and (o) a summary of the main findings. Measures of sensitivity, specificity, and accuracy were also collected. Information (a), (b), (e), (f), and (g) were first reported for the overall description of the records selected for the scoping review. Then, aspects pertaining to points (h), (i), (k), (l), (m), (n), and (o) were presented in a more specific discussion depending on the approach used in each article (i.e., diagnostic vs. prediction model).

Detailed information about the diagnostic/prediction models (model/test type, predictors source, number of classes, accuracy, sensitivity, and specificity) is reported in Appendix S3. In order to summarize each paper with one

TABLE 1 Research strategy for the MEDLINE database.

Search iteration number	MEDLINE research request	Number of records found
1	(Drowsiness* or Fatigue or Hypovigilance or "Hypo vigilance" or "Loss of alertness*" or Tiredness* or Vigilan* or Sleepiness or lassitude or Wakefulness* or Arousal* or "Sustained attention" or Delirium).ab,kf,ti	186,311
2	Arousal/ or Sleepiness/ or exp Fatigue/ or Wakefulness/ or Delirium/	85,774
3	1 or 2	225,160
4	(4AT or "Attention Network Test*" or "AVPU Scal*" or "AVPU Scor*" or (Fatigue adj2 Scal*) or "SAFTE Model*" or "Confusion Assessment Method*" or "Delirium Severity Scal*" or "Glasgow Coma Scal*" or "Intensive Care Delirium Screening Checklist*" or "Karolinska Sleepiness Scal*" or "Percentage Eye Closure*" or PERCLOS or "Psychomotor Vigilance Task*" or "Psychomotor Vigilance Test*" or "RAMSAY Sedation Scal" or "Richmond Agitation-Sedation Scal*" or "Sour Seven Questionnaire*" or "Stanford Sleepiness Scal*" or "Epworth Sleepiness scal*" or "Maintenance of Wakefulness Test*" or "Continuous Performance Test*" or "Continuous Performance Task*" or Electroencephalogram* or "Recognizing Acute Delirium as Part of Your Routine").ab,kf,ti	81,125
Gold standard MeSH—5	Glasgow Coma Scale/ or Electroencephalography/mt [Methods]	29,405
6	4 or 5	99,399
7	("Consciousness Monitor*" or Electrocardiograph* or Electrodiagnosis or Electroencephalogra* or Electromyograph* or EMG or Electrooculograph* or "Electro Oculograph*" or EOG or FMRI or FNIRS or "Galvanic Skin Response*" or "Heart rate variabilit*" or "Hemodynamic Monitoring" or "Gordon Diagnostic System*" or Kinarm* or "Functional Magnetic Resonance Imag*" or Magnetocardiograph* or Magnetoencephalograph* or "Functional Near-Infrared Spectroscop*" or "Neurologic Examination*" or "Neuromuscular Monitoring" or "Neurophysiological Monitoring" or Polysomnograph* or "Skin conductance level*" or "Skin conductance response*" or "Bispectral Index Monitor*" or (("Blood Pressure" or "Blood Glucose" or "Eye* Movement" or ((Eye* or Visual or Gaze*) adj1 track*) or "Facial Expression*" or Gait* or "Heart Rate*" or "Respiratory rate*" or "Vital Sign*") adj3 (Analysis or Determination or Monitoring or Measurement* or Procedure* or Test*))).ab,kf,ti	332,553
New tech MeSH 8	Blood Glucose Self-Monitoring/ or Blood Glucose/ or Blood Pressure Determination/ or Blood Pressure/ or exp Consciousness Monitors/ or Electrocardiography/ or Electrodiagnosis/ or Electromyography/ or Electrooculography/ or Electroencephalography/ or Exp Eye Movement Measurements/ or Exp Neurologic Examination/ or Exp Vital Signs/ or Eye Movements/ or Facial Expression/ or Gait/ or Galvanic Skin Response/ or Heart Rate Determination/ or Hemodynamic Monitoring/ or Magnetic Resonance Imaging/ or Magnetocardiography/ or Magnetoencephalography/ or Neuromuscular Monitoring/ or Neurophysiological Monitoring/ or Polysomnography/ or Respiratory rate/ or Spectroscopy, Near-Infrared/	1,543,386
9	7 or 8	1,634,829
10	6 and 9	77,985
11	3 and 10	9158
12	(exp Child/ or exp Infant/) not ((exp Adult/ or exp Adolescent/) and (exp Child/ or exp Infant/))	1,261,390
13	11 not 12	8772
14	(Animals/ NOT (Animals/ AND Humans/))	4,658,904
15	13 not 14	7408

score for each considered metric, the following rules were followed. When the classification was not binary (three classes and more), sensitivity and specificity were given for the class where hypovigilance was prominent (i.e., if the classes were "alert", "slightly drowsy" and "drowsy", the scores are given for the "drowsy" class). When the paper presented the performances of more than one model, the reported scores are those of the best-performing model, based on accuracy (or specificity if accuracy was not available). A short description and count of the other models presented are given in the "Other candidate models" column. In the case where one metric was not available and could not be inferred from the data presented in the paper, the corresponding cell was filled with "NA".

2.5 | Critical appraisal and risk of bias analysis

We identified limitations and risk of bias for each article based on the PRISMA-ScR framework (see item 12, Liberati et al., 2009). The Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS-2; Whiting et al., 2011) or the Prediction Model Risk of Bias ASsessment Tool (PROBAST; Wolff et al., 2019) were used to evaluate the quality of each article depending on the type of tool studied: a diagnostic tool vs. a prediction model for the QUADAS-2 and PROBAST, respectively. Critical appraisal was independently executed by two authors (MF & MK), and reviewed by a third (AMarois). Again, both reviewers compared their analysis and, if they disagreed, a third reviewer made the final decision. The QUADAS-2 method guided our analysis based on the following risk of bias domains: (a) patient selection, (b) index test(s), (c) reference standard(s), and (d) flow and timing. The PROBAST tool focused on the following domains: (a) participants, (b) predictors, (c) outcome, and (d) analysis.

For each included study, an overall risk of bias evaluation was added for both QUADAS-2 and PROBAST analyses. This overall calculation was inspired by the Revised Cochrane risk-of-bias tool for randomized trials (RoB2) method (see Higgins et al., 2019; Sterne et al., 2019). For the overall risk-of-bias judgment, the following rule was used: (a) overall low risk of bias was attributed to studies with low risk of bias classification for all domains, (b) "some concerns" about the overall risk of bias was attributed to studies having either one or two domains for which some concerns were found, but without high risks, and (c) overall high risk of bias was assigned to studies with some concerns found in multiple domains (three or more) and for studies with at least one domain at high risk of bias. The RoB2 Excel sheet (Higgins et al., 2019), comprised of different macros, was then used and adapted to collate and generate results to summarize the QUADAS-2 and PROBAST analyses. It allowed us to display and summarize conclusions of our risk of bias analysis.

3 | RESULTS

The scoping review conducted in the three online databases yielded a total of 13,686 records (MEDLINE = 7408; PsycINFO = 4170; Inspec = 2108). In addition, we added 231 studies from other sources (total records from all sources = 13,917). After duplicate removal (n = 2125), 11,792 records were kept for initial screening. This initial assessment removed 10,534 records, leading to 1258 records that were selected for eligibility analysis. Twenty-four manuscripts could not be retrieved, therefore resulting in 1234 records that were assessed for detailed evaluation. Finally, the detailed assessment for eligibility removed 1213 records, identifying 21 studies to be included for synthesis in the review (note that the 21 studies included for synthesis are identified by an asterisk in the reference list). Figure 1 presents the PRISMA flowchart of the study selection process.

Among the 21 included studies, five were diagnostic studies (i.e., studies aiming at presenting a hypovigilance diagnostic tool) while the other 16 were prediction studies (i.e., research on hypovigilance prediction tools relying on artificial intelligence algorithms). Table 2 presents the generic information of each included study, depending on the main approach employed (i.e., diagnostic vs. state prediction modeling). The main cognitive state outcome varied between studies. Studies sometimes focused on sleepiness (n = 1), vigilance (n = 1), drowsiness (n = 13), alertness (n = 1), fatigue (n = 3), mental fatigue (n = 1), and somnolence (n = 1). Sample sizes varied across studies. Studies employing a diagnostic approach had a mean sample size of 18.2 participants (SD = 5.9) while those related to state prediction models had a mean of 20.7 participants (SD = 14.3). Sex and gender of participants were sometimes omitted from the prediction model studies. As depicted in Table 2, the following psychophysiological techniques were studied: electroencephalography (EEG), electrooculography (EOG), electromyography (EMG), electrocardiography (ECG), respiration rate (RR) measures, oculometry (OCM), pupillometry (PCM), photo-oculography (POG), body movement (BM) measures, and near-infrared spectroscopy (NIRS). Among the diagnostic papers, three (60%) were presented in the context of driving literature and vehicle accident mitigation while the other two (40%) employed more generic approaches. Among the prediction studies, 14 (87.5%) addressed hypovigilance from a driving perspective while two (12.5%) discussed

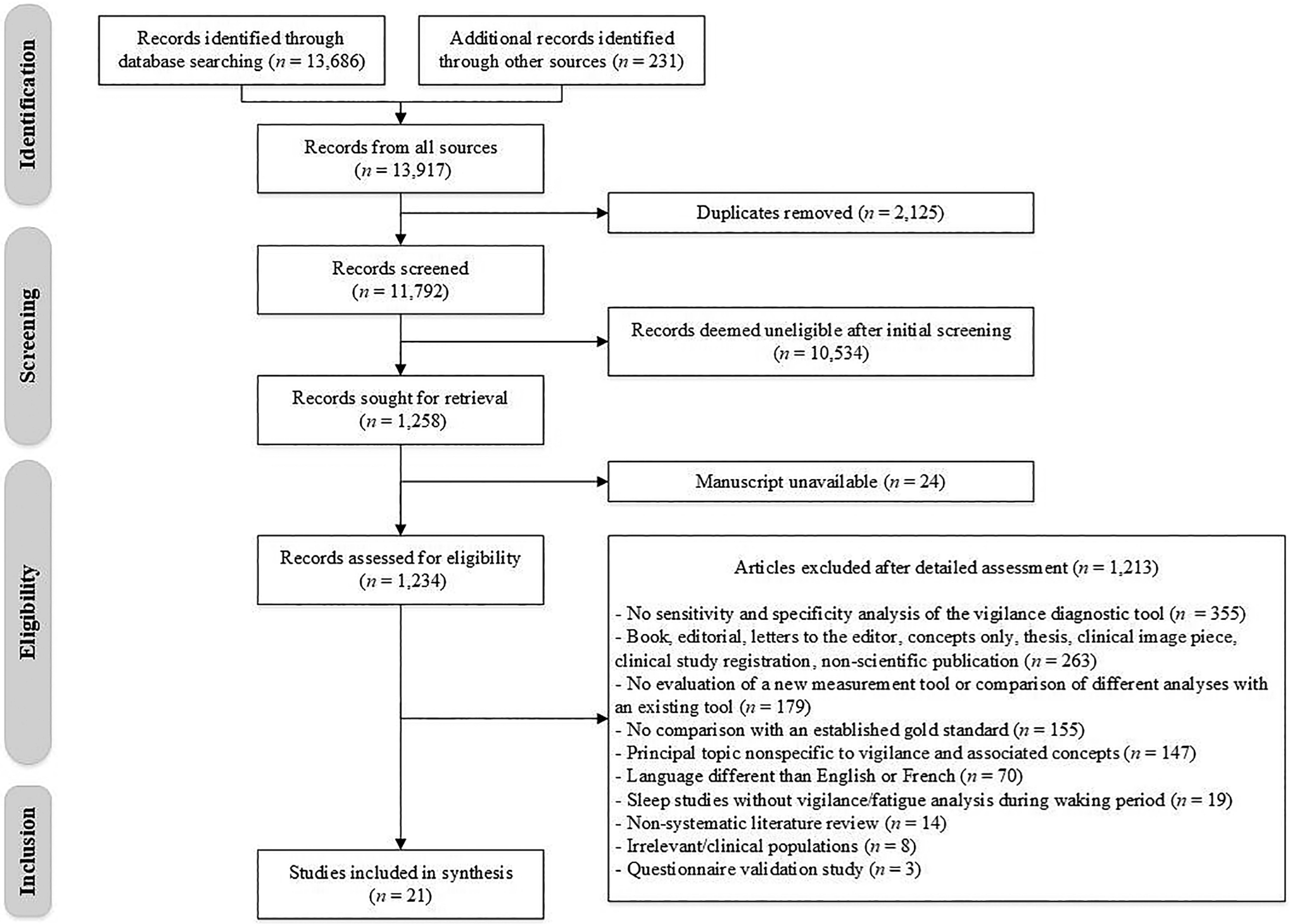

FIGURE 1 PRISMA flowchart diagram of the study selection process.

hypovigilance in a more general, domain-agnostic sense, and one (Zhang et al., 2017) studied the context of rail transport. In other words, almost all studies were related to transportation or a generic investigation of the hypovigilant state. This supports the idea that although distinctions exist between phenomena such as fatigue, somnolence, sleepiness, drowsiness, and other concepts related to hypovogilance, these concepts are applied to common real-world use cases and analyzed through a similar lens.

3.1 | Risk of bias analysis

3.1.1 | QUADAS-2 analysis

We analyzed five diagnostic studies with QUADAS-2 (Akerstedt et al., 2010; Chua et al., 2012; François et al., 2016; Maccora et al., 2018; Nguyen et al., 2017). Figure 2 depicts the overall risk of bias evaluation (panel a) as well as the detailed risk of bias analysis for each study (panel b). Overall, there were risks of bias concerns about the methods employed (with 60% considered concerning and 40% with potentially high risk for bias).

Patient selection systematically raised some or high concerns for all the studies given that it was unclear for all studies whether patients were selected consecutively or randomly. In general, insufficient information was given about patient selection, such as the exclusion criteria or case and control selection criteria. Chua et al. (2012) reported having studied only male participants, hence representing high risks of bias. Index test(s) were categorized as low risk for all studies. Reference standard(s) used raised high concerns for bias in Akerstedt et al. (2010) because the criterion of the gold standard for hypovigilance state was high (i.e., KSS score≥8, related to severe drowsiness). Other studies had a low risk of bias for reference standards. Finally, four studies out of five were considered as having some concerns about bias regarding flow and timing. Except for Nguyen et al. (2017), it was unclear for all other studies whether flow and timing aspects were correctly controlled for. For example, some papers did not present any data management reasons such as the absence of missing data management information (Chua et al., 2012),

TABLE 2 Characteristics of the different included studies depending on their approach (diagnostic, $n = 5$; state prediction modeling, $n = 16$).

Reference	Journal	Cognitive state	*N* (F/M)[a]	Physiological measure method[b]	Context of study
Diagnostic studies					
Akerstedt et al. (2010)	*Journal of Sleep Research*	Sleepiness	14 (7/7)	ECG, EEG, EMG, and EOG	Driving
Chua et al. (2012)	*Sleep*	Vigilance	24 (0/24)	EEG, ECG, and OCM	Generic
François et al. (2016)	*International Journal of Environmental Research and Public Health*	Drowsiness	24 (13/11)	POG	Generic
Maccora et al. (2018)	*Journal of Sleep Research*	Alertness	18 (8/10)	PPM	Driving
Nguyen et al. (2017)	*Scientific Reports*	Drowsiness	11 (1/10)	EEG and NIRS	Driving
Prediction model studies					
Awais et al. (2017)	*Sensors*	Drowsiness	22 (unknown)	ECG and EEG	Driving
Choi et al. (2019)	*IEEE Access*	Drowsiness	8 (4/4)	ECG, EEG, and EOG	Generic
Guo et al. (2016)	*International Journal of Environmental Research and Public Health*	Fatigue	20 (8/12)	ECG and EEG	Driving
He et al. (2016)	*IET Intelligent Transport Systems*	Drowsiness	50 (unknown)	BM, EEG, and OCM	Driving
Hu and Zheng (2009)	*Expert Systems with Applications*	Drowsiness/sleepiness	5 (3/2)	EOG	Driving
Kudinger et al. (2020)	*Sensors*	Drowsiness	30 (14/16)	ECG	Driving
Leng et al. (2015)	*IEEE Sensors Journal*	Drowsiness	20 (5/15)	EDA and PPG	Driving
Li and Chung (2015)	*Sensors*	Drowsiness	6 (unknown)	BM and EEG	Driving
Li et al. (2015)	*IEEE Sensors Journal*	Drowsiness	20 (8/12)	EEG	Driving
Lopez de la O et al. (2012)	*Procedia—Social and Behavioral Sciences*	Somnolence, drowsiness, and fatigue	23 (2/21)	BR	Driving
Mehreen et al. (2019)	*IEEE Sensors Journal*	Drowsiness	50 (20/30)	BM, EEG, and EOG	Driving
Mu et al. (2017)	*International Journal of Pattern Recognition and Artificial Intelligence*	Fatigue	15 (7/8)	EEG	Driving
Salvati et al. (2021)	*Entropy*	Drowsiness	3 (0/3)	ECG	Driving
Vicente et al. (2011)	*Computing in Cardiology*	Drowsiness	21 (unknown)	ECG	Driving
Yamada and Kobayashi (2018)	*Artificial Intelligence in Medicine*	Mental fatigue	31 (10/21)	OCM and PPM	Generic
Zhang et al. (2017)	*Sensors*	Fatigue and vigilance	10 (3/7)	EEG	(Train) driving

Abbreviations: BM, body movement measures; BR, breathing rate measures; ECG, electrocardiography; EDA, electrodermal activity; EEG, electroencephalography; EMG, electromyography; EOG, electrooculography; NIRS, near-infrared spectroscopy; OCM, oculometry; POG, photo-oculography; PPG, photoplethysmography; PPM, pupillometry.

[a]From the information available, all participants of these studies self-identified as either male or female, hence the absence of a third category for other genders. The total *N* represents the number of subjects included for analysis.

[b]The methods indicated here represent the measures tested in the paper (which was compared with a physiological or non-physiological gold standard method).

others only included a subset of participants in their analysis (Maccora et al., 2018), and others had small sample sizes (Akerstedt et al., 2010). Overall, Akerstedt et al. (2010) and Chua et al. (2012) had high risks of bias while the other studies were considered to have mild concerns (François et al., 2016; Maccora et al., 2018; Nguyen et al., 2017).

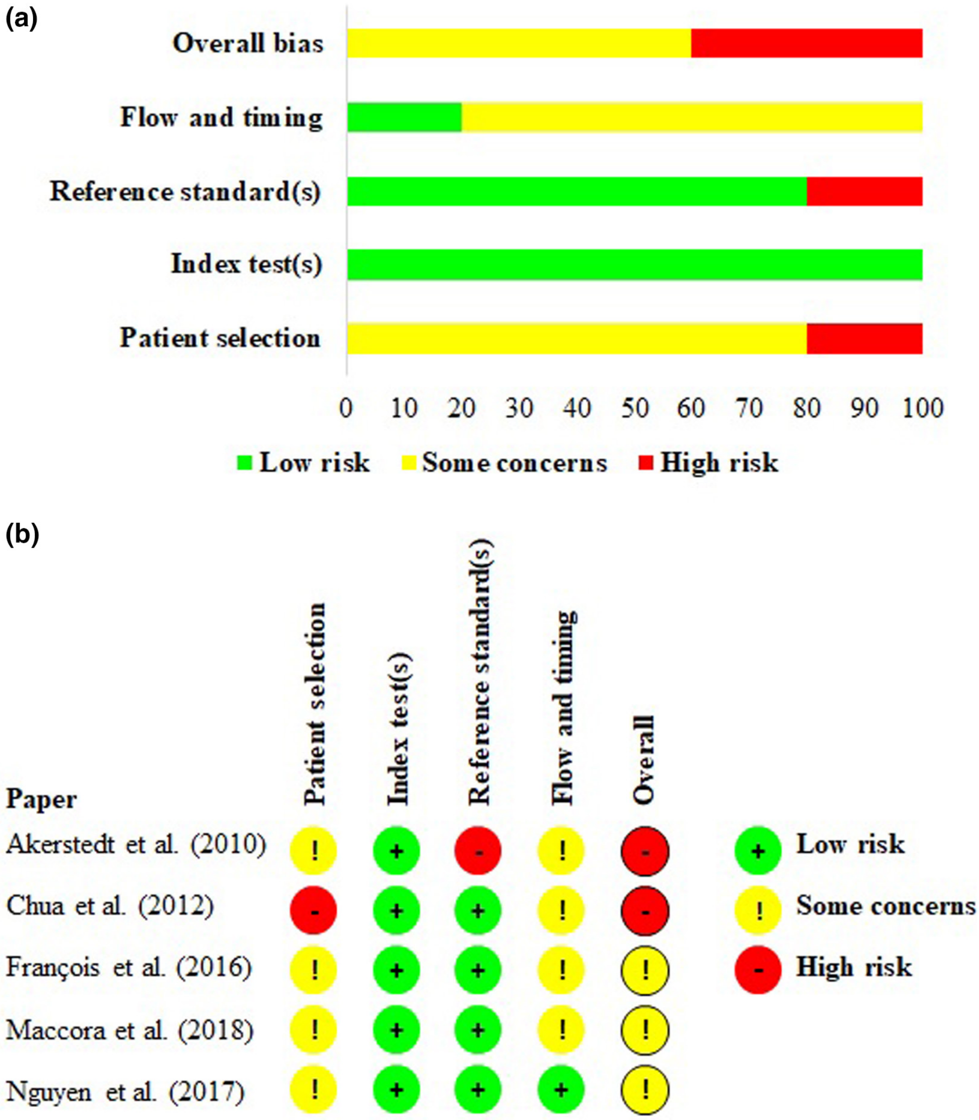

FIGURE 2 QUADAS-2 bias analysis for the diagnostic studies (Panel a: Global overview; Panel b: Detailed analysis).

3.1.2 | PROBAST analysis

We analyzed the risk of bias for 16 prediction studies (Awais et al., 2017; Choi et al., 2019; Guo et al., 2016; He et al., 2016; Hu & Zheng, 2009; Kudinger et al., 2020; Leng et al., 2015; Li et al., 2015; Li & Chung, 2015; Lopez de la O et al., 2012; Mehreen et al., 2019; Mu et al., 2017; Salvati et al., 2021; Vicente et al., 2011; Yamada & Kobayashi, 2018; Zhang et al., 2017) using PROBAST. Figure 3 displays the overall risk of bias evaluation for all included studies (panel a) as well as the detailed analysis for each of the 16 studies (panel b). Overall, the 16 studies had some concerns about bias or high risks of bias because of the methods used (with 43.8% considered with high risks vs. 56.2% with some concerns).

The nature and selection of participants raised some risk of bias concerns in 14 (87.5%) studies, except for Choi et al. (2019) at low risk of bias for the participant domain, and for Salvati et al. (2021) at high risk of bias. The main limitations observed were related to the lack of details regarding the participant's population and their risk for bias (e.g., night shift workers or drivers). In the case of Salvati et al. (2021), all participants were males, which can represent an important bias for the generalization of physiological prediction models. The predictors domain yielded a low risk for bias in every study except for Salvati et al. (2021). In this study, predictors were not defined a priori, but rather post hoc as determined by variations in PERCLOS. Most studies were at low risk of bias for the outcome domain (n = 11, 68.8%), but 4 (25%) still raised some concerns (Hu & Zheng, 2009; Mu et al., 2017; Salvati et al., 2021; Yamada & Kobayashi, 2018) and one was at high risk of bias (6.2%; Choi et al., 2019). Some concerns about risk of bias were due to lack of details about

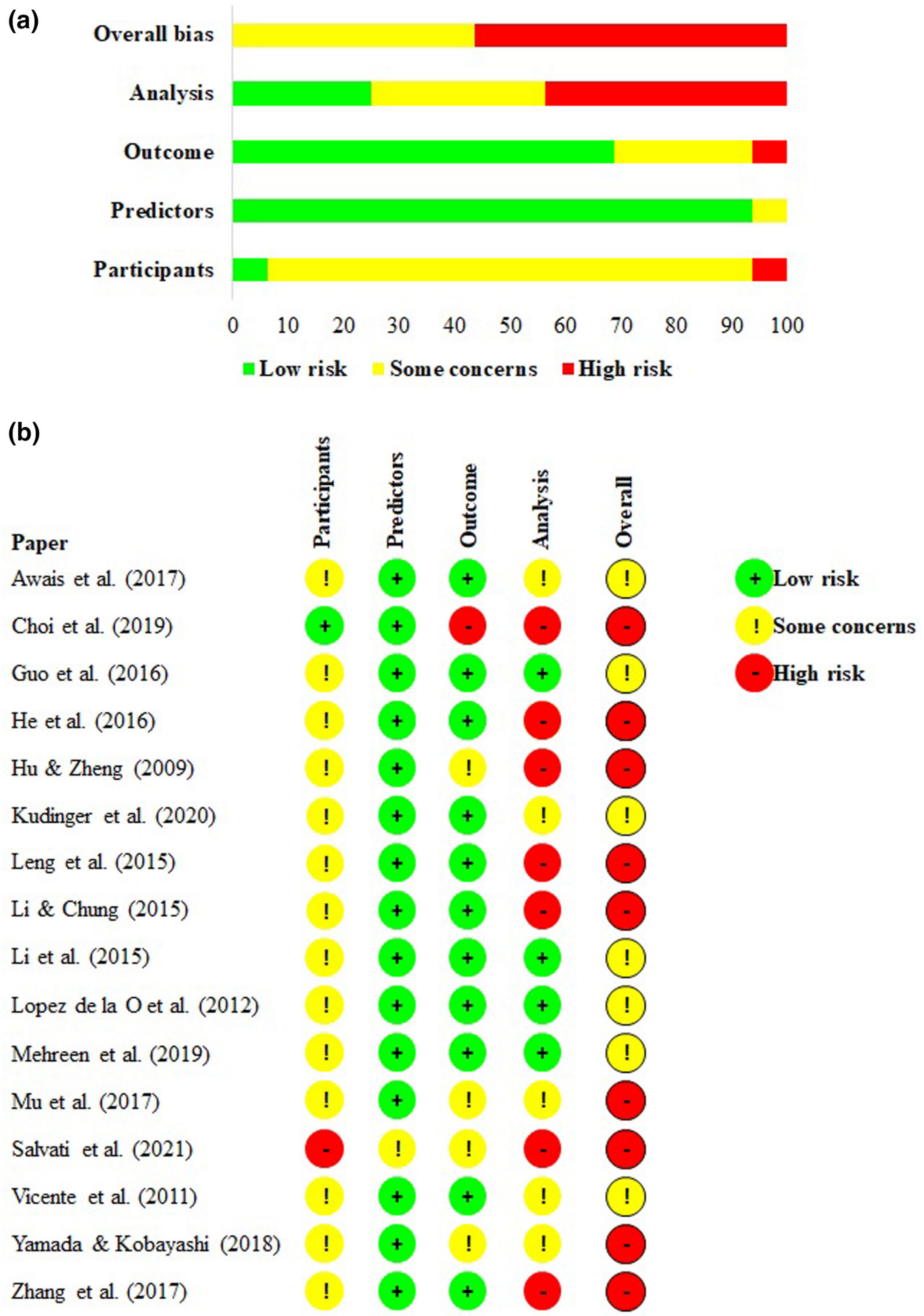

FIGURE 3 PROBAST bias analysis for the prediction studies (Panel a: Global overview; Panel b: Detailed analysis).

outcome determination and one study was at high risk of bias (Choi et al., 2019) because the study outcome (drowsiness) was determined post hoc based on the model used.

More concerns were found for the Analysis domain with 43.7% of the papers ascribed to the high-risk category, 31.3% associated with some concerns, and a minority of 25% deemed to be at low risk of bias. In the case of papers with "some concerns," risks for bias were due to the following reasons: (a) only a limited number of participants and/or a limited number of data points were included for analysis (Awais et al., 2017; Mu et al., 2017; Yamada & Kobayashi, 2018), (b) risks for overfitting were important due to the testing approach (i.e., not using leave-one-out approach; Kudinger et al., 2020), and (c) no performance measure was reported on the test set (Vicente et al., 2011). For the "high risk" papers, combinations of the preceding reasons explained this classification (Li & Chung, 2015), coupled sometimes with a lack of information on the analysis strategy, the absence of a clearly-defined test set, or because the sample size was very low (Choi et al., 2019; He et al., 2016; Hu & Zheng, 2009; Leng et al., 2015; Salvati et al., 2021; Zhang et al., 2017). Overall, we categorized nine papers to be at high risk of bias (Choi et al., 2019; He et al., 2016; Hu & Zheng, 2009; Leng et al., 2015; Li & Chung, 2015; Mu et al., 2017; Salvati et al., 2021; Yamada & Kobayashi, 2018; Zhang et al., 2017) while all the others were considered having "some concerns" for bias.

3.2 | Synthesis of included studies

3.2.1 | Diagnostic studies

Table 3 presents the summary of the five diagnostic studies, including their main findings. Hypovigilant states were induced by various methods in the five studies. Four of the studies relied on sleep deprivation/prolonged wakefulness while the other one used a monotonous driving task. Most of the studies relied on the same hypovigilant state induction technique (i.e., fatigue-induced hypovigilance). This improves our comparison of studies, establishing common bases for hypovigilance and, potentially, similar levels. However, the outstanding study that relied on the monotonous driving task differs from the other four studies, because it may have produced a lower level of hypovigilance. One could indeed expect that being importantly sleep deprived (e.g., be awake for at least 28 h or having slept less than 4 h) may cause different types and ranges of biobehavioral manifestations. Two of the studies used a driving simulator as the focal task. Two others used a constant routine task, i.e., a sequence of various daily tasks to perform, and one study used a PVT repeated at standard testing intervals over 2 days. Here, the variability in focal tasks can also induce differences in the ways that performance on a task may be modulated by the hypovigilance interventions. Still, task performance did not represent a key outcome for the study nor the diagnostic model, so the impact of such a difference among the studies reviewed may be relatively small. Of interest, however, is the fact that only three of the studies were carried out in real-life or simulation contexts close to real life (i.e., during a simulation or during constant routine tasks) that would be useful in operational situations.

Gold standards also varied importantly, both in their nature and, across similar tools, with respect to the thresholds for defining alert vs. hypovigilance states. The Karolinska Sleepiness Scale (KSS) and PVT were individually used in Akerstedt et al. (2010) and Chua et al. (2012), respectively, while the other studies proposed combinations of gold standards (e.g., EEG + slow eye movements + PVT; Maccora et al., 2018). Thresholds were determined by the research teams and varied importantly (e.g., rater's subjective visual inspection of EEG signal vs. standardized analysis of the EEG signal using Rechtschaffen and Kales' [1968] Karolinska Drowsiness Test [KDS] classifications vs. subjective evaluation of the variations in PERCLOS, EEG power bands and heart rate). This divergence in the gold standard and thresholds chosen for hypovigilance diagnosis complicates comparisons between the different measuring tools. More precisely, this causes variability in the classification of the main outcome (e.g., hypovigilant vs. vigilant state) across the included studies. The drawback of this variability is that some participants may have been assigned as hypovigilant from the perspective of a given gold standard while, from another, participants would be considered vigilant. This variability affects the external validity of the models (i.e., the capacity to generalize among new sets of individuals).

The different measurement tools used for hypovigilance diagnostics were: ECG, EOG, EEG, EMG, HRV, PERCLOS, POG, PUI, and NIRS. One of the studies explored only the variation of the pupillographic sleepiness test (Maccora et al., 2018), while the others proposed combined measures (e.g., EOG + EEG + EMG + ECG, EEG + PERCLOS + HRV frequency metrics + ECG power density). Many of the possible physiological measures reflecting hypovigilance are characterized by important between-individual variation due to difficulty to capture specific information on the state of the user, of interference from confounding variables, and more. Hence, combining multiple physiological measures seems appropriate to enhance sensitivity and specificity of diagnostic tools. However, the determined threshold varied for the same measure and was frequently decided empirically. This may have led to bias.

Different measures seemed to correlate with the level of hypovigilance, including blink duration, blink amplitude, peak closing velocity, and variability in lateral gaze position. PVT, ECG power density, EEG power density, NIRS oxyhemoglobin, POG, and PUI also had a good correlation. Four out of five studies included oculographic measures, whether it was pupillometric measures, percentage of eye closure, blink duration, lateral deviation of gaze, etc. This is probably because oculographic measures are relatively simple and cost-effective compared to EEG, which necessitate a skilled individual to install electrodes and can sometimes be invasive and/or uncomfortable. The important range of physiological measures that can be related to hypovigilance stresses the relevance of adopting a validated approach to detect this state. It also raises the potential of not only varying measures among a single technique (e.g., different spectral power bands of EEG),

TABLE 3 Summary of the domain of application, interventions, tasks, measures, and main findings of the diagnostic studies.

Reference	Hypovigilance intervention	Task	Gold standard	Hypovigilant threshold	Diagnostic sensors	Specific diagnostic measures	Main findings
Akerstedt et al. (2010)	4-h night sleep deprivation	Driving simulation	KDS	Rechtschaffen and Kales' (1968) KDS classifications	EEG (unknown)	Best predictors among EOG, EEG, EMG, and ECG measures	Prediction of KSS using blink duration, blink amplitude/peak closing velocity, and variability in lateral position
Chua et al. (2012)	40-h awaken period	Constant routine	PVT	Lapses of >0.5 s on the PVT	Comet Portable EEG (Astro-Med, Inc.), ISCAN eye-tracker (ISCAN, Inc.)	EEG power bands; PERCLOS; VLF, LF, and HF HRV; normalized LF and HF power; ECG power density	PVT correlation with EEG power density, ECG power density, and PERCLOS
François et al. (2016)	At least 28 h of sleep deprivation	PVT over two days	KDS and PVT	Rechtschaffen and Kales' (1968) KDS classifications + Lapses of >0.5 s on the PVT	Prototype of Drowsimeter R100, (Phasya)	POG model: Blinks duration, PERCLOS, % of microsleep	Coherence of POG measures with KSS, PVT, and PSG
Maccora et al. (2018)	40-h awaken period	Constant routine	EEG, SEM, and PVT	Visual inspection of EEG signal for microsleeps (intrusion of delta or theta activity >3 s) + Visual inspection of slow eye movements in EOG data + Lapses of >0.5 s on the PVT	F2D2 portable pupillographic system test (MTech Pupilknowlogy)	PST	PUI of the PST increased with time awake, similar pattern to EEG, SEM, and PVT
Nguyen et al. (2017)	Time on task (up to 10 min after signs of drowsiness)	Driving simulation	Blinking rate, PERCLOS, HR, and α and β EEG band power	Increase in blink rate with >2 s of PERCLOS, reduced HR, higher α band power, and decrease in β band power	Biosemi Active Two	Best predictors among EEG and NIRS features	Main differences in NIRS O2Hb and EEG β band power in frontal lobe

Abbreviations: EEG, electroencephalography; HF, high frequency; HRV, heart rate variability; KDS, Karolinska drowsiness scale; KSS, Karolinska sleepiness scale; LF, low frequency; O_2Hb, oxyhemoglobin; PERCLOS, percentage of eyelid closure; POG, photo-oculography; PSG, polysomnography; PST, pupillographic sleepiness test; PUI, pupillary unrest index; PVT, psychomotor vigilance task; SEM, slow eye movements; VLF, very low frequency; α, alpha band; β, Beta band.

but also multiplying the sensors included in a diagnostic model (e.g., combining measures of EEG and ECG).

3.2.2 | Prediction model studies

Table 4 presents the summary of the 16 prediction studies. Several techniques to induce a hypovigilant state are found across the different studies. These techniques comprised long time on task in a monotonous context, sleep deprivation or prolonged wakefulness, manipulation of the time of the day where testing occurred, recruitment of sleep-deprived participants (i.e., night shift workers after their shifts), and performing a cognitively-demanding task. This variability in the techniques chosen for inducing hypovigilance could have exerted different hypovigilance levels and, consequently, different outcomes for its prediction. Some physiological measures may be more or less sensitive than others and so prediction could have been enhanced or worsened for certain physiological responses if a different technique was used. Important variability can also characterize the hypovigilant vs. aroused participants across all studies. Records characterized by less severe hypovigilance-inducing techniques (e.g., time on task on a driving simulation) may incorrectly categorize alert individuals as hypovigilant compared to studies employing more severe manipulations (e.g., with subjects sleep deprived for 26 h). This limitation can, however, be mitigated by having more than two hypovigilance levels (e.g., fully awake vs. drowsiness vs. fatigue; cf. Lopez de la O et al., 2012). Here, having a third category may allow more precision in the categories and more important homogeneity in the cases ascribed to each state. In turn, the reduced variability can lead to better state prediction.

The ongoing task during which hypovigilance was measured varied to a lesser extent, the majority focusing on driving (10 studies employed a driving simulation and four a real driving task). Other studies relied on a series of recurring and continued routine tasks, on monotonous single-object tracking, and an alternation between video watching and cognitive tasks. The fact that most of the studies discussed and evaluated hypovigilance under a transportation/driving perspective speaks to the importance of such a cognitive limitation for this specific context. This also means that most of the studies aimed at developing a hypovigilance prediction model that would be applicable to/deployable in real-life settings such as in a car or on a train. In that regard, most of the sensors used to measure the physiological responses and in turn provide data to the hypovigilance state prediction model were mobile (commercial-off-the-shelf or homemade) sensing technologies.

The gold standard used varied between several physiological and behavioral outcomes. Physiological outcomes comprised measures of facial features, EOG, eye movements, ECG, body movements, and mostly, EEG spectral bands. The thresholds to label vigilance levels from these metrics often changed across studies, even when a common physiological signal was analyzed (e.g., PERCLOS evaluation in Li et al. [2014] and Lopez de la O et al. [2012]; or EEG power bands assessment based either on Rechtschaffen and Kales' [1968] KDS classifications or not in He et al. [2016], Hu and Zheng [2009] and Lopez de la O et al. [2012]). Behavioral outcomes included results on self-rated scales (e.g., Borg's CR-10 scale, Karolinska Sleepiness Scale, Li's Subjective Fatigue Scale, Stanford Sleepiness Scale, or homemade mental fatigue, physical fatigue, sleepiness, and motivation numerical scales), performance on a task to measure fatigue (e.g., PVT, and reaction time on a simple task), and performance on the focal task (steering wheel adjustments on the driving simulation). Sometimes information on the thresholds used to label vigilance was absent (Mu et al., 2017; Vicente et al., 2011; Yamada & Kobayashi, 2018) and, in other situations, label derived only from experimental manipulations (Mehreen et al., 2019; Zhang et al., 2017). Here, the diversity in gold standard measures and thresholds compromises between-studies comparisons. In fact, having different gold standard measures necessarily leads to having different thresholds for determining the hypovigilance state of a user. For example, some studies employed the KSS and used several threshold points for identifying different hypovigilance levels (e.g., KSS classes 0–4: Alertness; KSS classes 5–8: Hypovigilance; KSS classes 9–12: Drowsiness; Salvati et al., 2021). Yet, these categories can be difficult to compare with physiological-based thresholds, e.g., on measures of PERCLOS (e.g., Li et al., 2015; Lopez de la O et al., 2012) or variations in the PVT performance (e.g., Choi et al., 2019).

Common information on the measure of hypovigilance can be deduced from the main findings of the studies concerned with prediction models. Mainly, the studies relied on EEG-related measures (50% of the studies) and on ECG features (43.8% of the studies) to predict the hypovigilance level of participants. EEG features mainly reported spectral power bands (α, β, γ, δ, θ, and φ) and power density. Regarding the ECG features, frequency bands of the HRV were mainly used, but also some time-domain features such as raw HR, HRV, or RR intervals. Some papers were also interested in predicting a hypovigilant state with measures of body movement, including aspects related to the adjustments of the body and to head movements/nodding. These latter aspects can be processed and interpreted through many outcomes, as shown by the 21 features of head movement

TABLE 4 Summary of the domain of application, interventions, tasks, measures, and main findings of the prediction models studies.

Reference	Hypovigilance intervention	Task	Gold standard	Hypovigilant threshold	Prognostic sensors	Specific prognostic measures	Main findings
Awais et al. (2017)	Time on task/ monotony	Driving simulation	KSS, facial features, blink duration and rate, head movements	Rater's subjective evaluation from video recordings (eye blink duration, facial expressions, facial tone, eye blinking rate, and movements)	Enobio-20 channel (Neuroelectrics)	Time domain of EEG signal; δ, θ, α, β, and γ EEG band power; VLF, LF, and HF HRV	Significant predictors of the SVM: Energy and entropy measures of EEG (parietal and occipital); Absolute δ, θ, and α band power (central, parietal, and occipital); Relative α band power (occipital and parietal)
Choi et al. (2019)	<4-h night sleep deprivation	Constant routine	PVT and EOG	Lapses on the PVT (criterion undefined) + R100 EOG labeling from François et al. (2014)	Wet-electrode EEG (Beehive Horizon, Gras Technologies) and cap-type dry-electrode EEG (Ybrain Inc.)	With wired and wireless EEG: power bands; multitaper power spectral density; ECG measures; PVT	Similar prediction with wired and wireless EEG features and ECG, prediction made with XGBoost with PVT-based labeling
Guo et al. (2016)	Time on task/ monotony	Driving simulation	SSS	Level 3 on the SSS	EEG (unknown)	δ, α, β, and combined α-θ EEG band power; Overall EEG waves; HR and HRV	RT mainly related to overall EEG spectra, β power, and α/β ratio (gray correlations). SSS prediction with EEG features and RT from GA-enhanced SVM
He et al. (2016)	Time of day testing	Driving simulation	EEG δ, θ, α, and β bands	Mean power spectra ratio based on δ, θ, α, and β bands	Northern Digital Polaris Optical Tracking Systems (NDI HPS) and CCD camera	PERCLOS, and nodding frequency and angle	Features coupled with time of day and time on task can predict EEG-identified alert vs. drowsy state with ANN
Hu and Zheng (2009)	Night shift workers after their shifts	Driving simulation	KDS and KSS	Rechtschaffen and Kales' (1968) KDS classifications and KSS analysis (sleepy: KSS $\geq 7 + 15 < \text{KDS} < 25$; very sleepy: KSS $\geq 8 + \text{KDS} < 35$)	EEG (unknown)	Eyelid movement parameters (11)	Sleepiness predicted with a SVM model based in eyelid movements, better for greater sleepiness levels
Kudinger et al. (2020)	Time on task/ monotony	Driving simulation	Eye behavior and facial expressions	Rater's subjective evaluation from video recordings (eye blink closure, eyes rolling, behavior, facial expressions) + Eyelid closure time (drowsy: $1 \leq s < 2$; very drowsy: $2 \leq s < 4$; extremely drowsy: $s \geq 4$)	Empatica E4 wristband (Empatica Inc.)	Several ECG features related to the time domain, frequency domain, and non-linear domain with feature selection	Best prediction reached with user-dependent model comprised of max RR, min RR, max HR, and min HR (best accuracy with KNN)

(Continues)

TABLE 4 (Continued)

Reference	Hypovigilance intervention	Task	Gold standard	Hypovigilant threshold	Prognostic sensors	Specific prognostic measures	Main findings
Leng et al. (2015)	Time on task/monotony	Driving simulation	KSS	KSS scores related to drowsiness levels (scores 1–2: level 1; scores 3–4: level 2; scores 5–6: level 3; scores 7–8: level 4; score 9: level 5)	Homemade wrist band (Arduino Lilypad) and built-in motion sensor	EDA-extrapolated stress level, HR, PRV, RR, and number of adjustments of body	Accurate prediction of KSS with an ensemble of physiological measures in an SVM
Li and Chung (2015)	Time on task/monotony	Driving simulation	Wierwille video-based scale	Wierwille and Ellsworth's (1994) video-based criteria based on eye blinks duration and frequency, glances durations and frequency, glazed-eye looks, irregular movements, amplitude of body movements, and yawning	Homemade wireless EEG headset and gyroscope	Head movement power, and θ, α, and β EEG band power	SVM model comprised of physiological features capable of predicting subjective ratings from video scale
Li et al. (2015)	Time on task/monotony	Driving simulation	PERCLOS and number of steering wheel adjustments (NOA)	Drowsy: PERCLOS $\geq 12\% +$ NOA ≤ 9; Early-warning: $8\% <$ PERCLOS $< 12\% + 9 <$ NOA < 26; Alert: PERCLOS $< 8\% +$ NOA > 26	Wearable EEG with dry and wet electrodes (unknown)	θ, α, and β EEG band power	Early warning and fully warning SVM-based posterior probabilistic models provided accurate predictions of drowsiness level
Lopez de la O et al. (2012)	Time on task/monotony	Real driving	EEG, PERCLOS, and video-based scale	Attentive: high activity and rapid reactions + EEG θ ratio $< 1.92 +$ PERCLOS < 0.24 and low/fast blinking; Fatigued: slower reactions, yawns and big movements + EEG patterns of α waves and $1.92 > \theta$ ratio $< 8.22 + 0.24 >$ PERCLOS < 0.45; Drowsy: Fall of attention, driving errors and no facial expressivity + Loss of EEG α patterns and θ ratio $> 8.22 +$ PERCLOS > 0.45	Biomedical monitor (Bitmed eXim Pro, BitMed)	BR, conveyed into a model (mean TEDD)	The mean TEDD could correctly predict drowsiness phase, better for "fully awake" and "drowsiness" than "fatigue"
Mehreen et al. (2019)	Time on task/monotony after >8-h wakefulness	Single object tracking	EEG and KSS	Label based on experimental manipulation (drowsy: monotonous task following >18 hours of being awake; fresh: after 6–8 hours of sleep and without signs of drowsiness)	MUSE 2016 headband	δ, θ, α, β, γ, and φ EEG band power features (15); eye blink features (7); head movement features (21)	Most accurate prediction reached with backward-reduced SVM with 6 EEG features, 3 blink features, and 12 head movement features
Mu et al. (2017)	Time on task/monotony	Driving simulation	LSFS and BCR-10	No detail provided on the thresholds	Neuroscan 32 (Compumedics Neuroscan)	Fuzzy entropy of the power bands of 27 electrodes	Prediction with SVM model comprised of FP1 and FP2 electrodes (chosen from Fisher distance feature selection)
Salvati et al. (2021)	Time on task/monotony	Real driving	KSS and PERCLOS	KSS scores related to drowsiness (alertness: 0–4; hypovigilance: 5–8); drowsiness: 9–12) + PERCLOS: ratio of manifestations of drowsiness (70% of partial closing of the eyelids, blinking, yawning) on three levels (alertness: 0–0.33; hypovigilance: 0.34–0.66; drowsiness: 0.67–1)	Within-seat microphone sensor	ULF components of HRV	Model distribution of ULF components of HRV slightly related to PERCLOS and KSS, error-prone with transition phases

TABLE 4 (Continued)

Reference	Hypovigilance intervention	Task	Gold standard	Hypovigilant threshold	Prognostic sensors	Specific prognostic measures	Main findings
Vicente et al. (2011)	Sleep deprivation (7 h to 26 h) or 6-h driving	Simulated and real driving	Experts annotation on EEG, PERCLOS, and driving errors from videos	No detail provided on the thresholds	2-lead ECG (unknown)	Features of and LF and HF of HRV; BR features	LDA capable of discriminating awaken vs. drowsy state improved with merged data from simulation and real driving
Yamada and Kobayashi (2018)	Cognitively-demanding tasks	Video and cognitive task alternation	Mental fatigue, physical fatigue, sleepiness, and motivation numerical scales	No detail provided on the thresholds	EMR ACTUS eye tracking (nac Image Technology Inc.)	OCM features (9); Blink features (7); PPM features (6); Gaze allocation features (72); Eye direction features (50); and saliency features (28)	SVM with feature selection capable of prediction state of user according to two classes
Zhang et al. (2017)	Sleep deprivation and time of day testing	Real driving (train)	Comparison between sleep-deprived/night testing and non-sleep-deprived/day testing	Label based on experimental manipulation (drowsiness: test time between 4 and 6 a.m.; alertness: test time between 9 and 11 a.m.)	Homemade wireless EEG system	EEG power spectrum density	Possibility to classify alert vs. drowsy states with SVM with θ, α, and β band power and different time windows

Abbreviations: ANN, artificial neural network; BCR-10, Borg's CR-10 scale; BR, breathing rate; EDA, electrodermal activity; EEG, electroencephalography; EOG, electrooculography; GA, genetic algorithm; HF, high frequency; HR, heart rate; HRV, heart rate variability; KDS, Karolinska drowsiness test; KNN, *K*-nearest neighbor; KSS, Karolinska sleepiness scale; LDA, linear discriminant analysis; LF, low frequency; LSFS, Li's subjective fatigue scale; NOA, number of steering wheel adjustments; OCM, oculometry; PERCLOS, percentage of eyelid closure; PPM, pupillometry; PRV, pulse rate variability; PVT, psychomotor vigilance task; RR, RR interval; RT, reaction time; SSS, stanford sleepiness scale; SVM, support vector machine; TEDD, thoracic effort derived drowsiness; ULF, ultra low frequency; VLF, very low frequency; XGBoost, extreme gradient boosting; α, alpha band; β, beta band; γ, gamma band; δ, delta band; θ, theta band; φ, phi band.

collected by Mehreen et al. (2019). Some authors also relied on eye movement/behavior features leading to multiple types of measures including pupillometric data, eye fixations and saccades, and blink data. Again, all these types of measures could be processed into several outcomes of time and frequency domains (see, e.g., Hu & Zheng, 2009; Mehreen et al., 2019; Yamada & Kobayashi, 2018). Measures related to breathing were, however, scarcely used (only in two studies). Taken together, these results outline that hypovigilance can be successfully predicted using a wide range of physiological measure techniques and features.

The data originating from these outcomes can be processed using different machine learning algorithms. Techniques such as support vector machines (SVM), artificial neural network, genetic algorithms, decision trees, *K*-nearest neighbor, linear discriminant analysis, and extreme gradient boosting were used for the prediction of the hypovigilant state using ensembles of psychophysiological and behavioral features. While most of the studies used SVM models, the variability in modeling techniques is consistent with the variability already reported between studies for the selection of the gold standard, the hypovigilance-inducing techniques and the predictors. The nature of the technique may vary, among other things, depending on the type of predictor included in the models, but also according to the number of outcomes to predict (i.e., hypovigilance classes). Appendix S3 provides more details on the different models used in the 16 hypovigilance prognostic studies, and on the performance they reached with their sample.

4 | DISCUSSION

The goal of this scoping review was to map the current knowledge about the psychophysiological methods to detect human hypovigilance and to highlight strengths and gaps in the literature. The selection process and analysis of the 21 studies selected for the current scoping review highlight important trends for the scientific community interested in the detection of impaired cognitive states such as hypovigilance. First, the large number of papers assessed for eligibility (1234) confirms that this topic is indeed of interest for many researchers and that synthesis efforts such as our scoping review are needed to better understand the current state of the literature. The important diversity in journal scopes from which the papers were selected (including neuroscience, behavioral sciences, sleep research, transport systems, and sensors journals) reflects the overall interest of many different scientific communities. Interestingly, the detection of hypovigilance does not only apply to the medical or psychological domains, but also to applied sciences such as transport and engineering journals. The selection criteria chosen for the current scoping review were purposely focused on the cognitive aspect of hypovigilance detection. As a result, an important number of papers focused on the engineering side were excluded: most of them did not necessarily include an established gold standard (155 out of 1213) or focused on signal processing technologies. Although the technologies presented in these papers are necessary to develop robust systems in real situations, they were not the objective of our research and did not meet the inclusion criteria. The automobile industry is at the heart of the research for hypovigilance detection. Not all of the selected papers relied on a driving-related task, but they almost all aimed to be applied to the transport industry. As a result, most of the experiments conducted, either inside or outside the lab, investigated embedded or at least portable systems with low invasiveness (wearables such as wrist bands, contactless cameras for eye tracking, or sensors integrated in the driver's seat).

Throughout the selected papers, the physiological measures used to detect hypovigilance were relatively consistent. Indeed, out of the 21 studies considered, almost all papers relied on at least one of the following signals: ECG/PPG, EEG, EOG, and eye tracking. This conclusion is interesting given the small diversity observed in the specific phenomena assessed in these papers (vigilance vs. drowsiness vs. fatigue and so forth). This outlines that hypovigilance-related measures found in these studies may be underpinned by common mechanisms even if, from a semantic point of view, studies may have referred to this concept in different ways. Interestingly, other measures were also used, including body temperature, breathing rate, NIRS, body movement-based data, and, sometimes, behavioral measures. The combination of techniques may be motivated by the idea that physiological monitoring devices (e.g., heart rate monitors) are subject to several artifacts such as movement noise (Kranjec et al., 2014). Therefore, combinations allow for the possibility to collect state information when data from a given sensor or a group of sensors may comprise too much noise. Considering that hypovigilance measures of the central nervous system seem important, a great challenge is to transfer the usually bulky and sensitive sensors out of the laboratory (e.g., Awais et al., 2017; Choi et al., 2019; Li et al., 2015), but also to pinpoint proper cerebral indices of the (hypo)vigilant state.

The important diversity of gold standards (and sometimes thresholds) observed across studies is also noteworthy. Although common assessment measures were found across studies, gold standards were not used in the same way. The KSS was often used, but could be interpreted differently using, e.g., different number of categories.

Measures focused on observable behaviors or on physiological signals (e.g., PERCLOS, EEG power bands, heart rate variations, or body movements) that were analyzed differently. Sometimes, these measures relied on standardized/a priori-defined techniques (e.g., Rechtschaffen and Kales' [1968] KDS classifications, Wierwille and Ellsworth's [1994] video-based criteria, or PERCLOS percentage categories). Yet, in other situations, raters provided subjective evaluations based on their own observations, and the criteria they relied on were not explicitly discussed. This outlines that literature on hypovigilance is highly scattered and that, although common techniques can be pinpointed, between-studies comparisons are difficult to perform. Nevertheless, this information can still be of high use to help defining better ways to predict hypovigilance and guide future studies to compare different diagnostic tools and thresholds. Our results will be helpful in guiding standardized approaches to define proper ground truth labels to use to develop new prediction models. These approaches should ultimately all rely on common gold standards and thresholds to ensure that prediction models all rely on a common view of hypovigilance and to make between-model comparisons feasible.

Petersen and Posner (2012) suggest that the brain has three distinct attentional networks: alerting, orienting, and executive control. The alerting system is deemed to condition the general level of arousal and is influenced among other things by subcortical activity of the locus coeruleus (LC) (Foote et al., 1991). The LC generates norepineprhine (NE) and spreads it through the brain, in particular in the right thalamic, frontal, and parietal regions. Many papers used EEG or NIRS to measure activity in these cortical regions as downstream cortical indicators of the LC-NE system activity. Some focused on a generic approach with electrodes in every region of the brain, whereas others reduced the number and locations of electrodes (e.g., the frontal and temporal lobes or over the occipital lobe). Overall, the best cerebral locations from which to collect brain activity do not seem to have reached consensus in the studies we included. The neural pathways associated with hypovigilance still seem under-investigated (Li & Chung, 2022) and rarely corresponded with attention-related brain areas. Moreover, the placement of the electrodes was rarely justified. This could explain why several—rather than a single—regions of the brain were used to detect hypovigilance.

The approaches used to process and aggregate data are manifold, although the use of the spectral domain to process EEG and ECG is dominant. The level of details provided by each paper varies greatly, and it is not always stated: (a) how the data was processed, (b) which features were actually used as predictors for the detection of hypovigilance; and (c) what thresholds have been used specifically for labeling the vigilance level. Among others, the lack of transparency increased the concerns for some papers during the bias assessment, and more precisely the Outcome domain with PROBAST. Moreover, since the majority of the models investigated used machine learning techniques, the “black box” effect remains important, as the models and the between-variables relationships can be either difficult or impossible to fully interpret (e.g., Lipton, 2018). More precisely, the effect of each predictor on the target metric and their interactions were not necessarily explained. Unlike statistical analyses, the direction and values of one physiological parameter cannot be directly associated with specific variations of hypovigilance, which affects the interpretability of the models. Work focusing on predicting hypovigilance states with large varieties of psychophysiological features should integrate techniques to understand such a black box effect. Machine learning techniques exist to increase the explainability of models (i.e., methods of explainable artificial intelligence [XAI]; e.g., Antoniadi et al., 2021; Gunning et al., 2019; Tjoa & Guan, 2021), and efforts should be deployed to make use of them to better understand the mechanisms underlying hypovigilance detection.

In terms of algorithms, the selected papers reflect the recent advances of machine learning and its potential for human-centered applications (many prediction models were based on machine learning). The use of deep learning was not found to be dominant. Different techniques were utilized using algorithms that are well-established in the machine learning community for supervised learning such as Random Forest, XGBoost, LDA, and SVM. Interestingly, all of the proposed models were classifiers, discriminating between two and sometimes three classes (increasing levels of hypovigilance). None of the papers seem to have considered regression to infer vigilance levels (prediction of a continuous output such as an interpolated KSS score). At this stage, it is unclear whether using regressors is not efficient, or if it has not yet been considered sufficiently. This approach, if proven efficient, could be a way to introduce more granularity in the predictions. Moreover, a continuous prediction might make more sense than simply classification given that hypovigilance is not a binary state and progressively grows as time on task/difficulty increases (Robertson & O'Connell, 2010). Such an approach would however require defining and operationalizing a continuous ground truth equivalent, i.e. a measure representative of the normal level of vigilance over a certain time window, to ensure the constant validity of the new continuous physiological models.

Modeling a cognitive state based on psychophysiological data requires training a model that is sensitive enough to take into account intra-individual variability. Moreover, in order to be used in a large variety of applied situations, prediction models should ideally run in real time, and follow a "one size fits all" approach. This suggests that models must be robust enough to provide relevant predictions on different individuals, even if no prior information is available on specific individuals. Several methods in the training, validation, and testing phases of a machine-learning model can be used to quantify its generalization capacity. It is the case of the "leave-one-participant-out" cross-validation approach (de Rooij & Weeda, 2020), with which the validation phase happens on unknown participants' data. Similarly, performances of models on the test set should be evaluated on independent, isolated individuals. These methods usually lead to models that generalize better, but might show lower cross-validation performances (Drew et al., 2014; Suresh & Guttag, 2019; see also Gronau & Wagenmakers, 2019, for considerations of "leave-one-participant-out" cross-validation models). In the current state of the literature, the methods of training and evaluation of the models are manifold and heterogeneous. More generally, the differences in evaluations and hypotheses resulted in certain papers having higher bias estimations than others, more particularly in the Analysis domain of the PROBAST. Performances as reported by the authors are given in Appendix S3. However, they should be interpreted with caution, rather than used for comparison between two systems. Indeed, the variability of the techniques used to train and evaluate the models, as well as the discrepancies observed between papers during the bias assessment, would not lead to a fair and objective comparison.

4.1 | Practical implications

The aim of this scoping review was to describe the various tools available to detect and predict hypovigilance. As it was previously stated, this is of great importance in the transport industry, but also in aerospace, command and control, and other such complex and dynamic domains. Many attention-demanding tasks (such as traffic control, supervising military operations, vehicle driving, or piloting) with critical outcomes could eventually be assisted by a device designed to detect hypovigilance, with the aim of preventing hazardous events (see, e.g., Bendak & Rashid, 2020; Bier et al., 2020; Duffy & Feltman, 2022; Mogilever et al., 2018).

The ability to monitor the physiology of individuals to infer their mental states is already seen as highly valuable in a variety of contexts such as the monitoring of soldiers in military operations (Friedl, 2018; Salvan et al., 2022) and different kinds of adaptive systems have been developed for such purpose (Blackhurst et al., 2012; Marois et al., 2020; Parnandi et al., 2013; Zhao et al., 2020). Consequently, the potential uses of hypovigilance detection technologies are extensive. Industries like automobile and aerospace are evidently involved in this research field to prevent accidents, as inattention is a key human factor that can be monitored and supported to prevent casualties. Moreover, isolated, confined, and extreme environments (often referred to as ICE; Mogilever et al., 2018; Palinkas, 2003) could also benefit from such technologies, since they are known to induce mental health challenges with attention-related symptoms evolving into vigilance challenges (e.g., depressive states, anger, and anxiety; see, e.g., Haney, 2003; Palinkas et al., 2004). For all those cases, the information extracted from the literature reviewed in the current paper can represent a great asset from both researchers' and decision makers' point of view. The different physiological techniques identified (with their advantages and drawbacks) as well as the prediction/modeling approaches raised could contribute to the development and integration of such systems for real-life applications.

Another interesting field of research is the detection of hypovigilance in hospitalized patients. Artificial intelligence opens wide possibilities in the medical field, where multiple clinicians' decisions could be supported by machine learning (e.g., radiologic automated analysis). One of the main challenges in medicine is identifying patients at risk for and with actual delirium, especially for hypoactive-type delirium characterized by reduced vigilance (e.g., Gual et al., 2019; Hosker & Ward, 2017; Inouye, 1994). Delirium is defined in the DSM-5 as a state of "disturbance in attention (i.e., reduced ability to direct, focus, sustain, and shift attention) and awareness (reduced orientation to the environment)" (American Psychiatric Association, 2013, p. 596). Clinical diagnostic criteria are well-defined and helped develop a clinical tool used at the bedside by clinicians to diagnose delirium, called the Confusion Assessment Method (CAM). Its application in the intensive care unit (ICU) is possible through the CAM-ICU (Ely, Margolin, et al., 2001). The diagnosis requires both acute onset and fluctuating course, and the presence of either disorganized thinking or altered level of consciousness. While the CAM-ICU administration can take less than 1 min (Guenther et al., 2010), it needs to be carried out frequently while a patient is hospitalized. Consequently, efforts must be invested to integrate this tool into the patient's follow-up workflow and into the routine of the busy and overburdened ICU personnel. Developing automated tools that would help identify hypovigilant situations for the diagnostic of delirium and/or identify patients more at risk could be a way to increase delirium detection in understaffed ICUs. Such tools would be useful given that ICU delirium is associated with worse patient-oriented outcomes, including increased ICU/hospital length of stay, more frequent mortality, and worse cognitive outcomes among ICU survivors (see, e.g., Ely, Margolin, et al., 2001; Fiest et al., 2021; Salluh et al., 2015).

Of all the studies screened, only one concerning delirium met all the inclusion criteria in the first steps of inclusion assessment, but it was later removed. This study by Oh et al. (2018) was not included in the scoping review because it focused only on the hypovigilance experienced by ICU patients diagnosed with delirium and did not fit well with the scope of the other selected papers. Other studies are currently in progress to evaluate EEG variation analysis to identify delirium in ICU patients (e.g., Ducharme-Crevier, 2021). One could also suggest that automated measures with machine learning could open doors to diagnose many medical conditions, for example, sepsis and psychosis. This could represent a great asset for health systems, given that human factors and lack of time represent important practical limitations (Goodie & Crooks, 2004; Weinger & Slagle, 2002).

4.2 | Limitations

While the current review provides a comprehensive portrait of the literature on hypovigilance detection and prediction models, it still possesses some caveats that must be addressed. First, the imposition of a finite list of gold standards might have reduced the number of papers selected for review. Although this list was flexible through the source selection step, it still excluded potentially relevant papers that presented other (unique or sets of) psychophysiological proxies related to hypovigilance. While imposing the presence of a gold standard can help to ensure better validity of the models presented, some models that we missed could still be highly relevant. Yet, to prevent reducing, even more, our capacity to include papers in the review, we did not include preimposed thresholds for these gold standards. Second, the fact that all the studies selected raised concerns for bias—and sometimes high concerns—reduces the scope of interpretation and, potentially, the generalizability of the conclusions reached by these studies. Indeed, the results discussed herein might only be applicable to certain groups of persons, or specific to given contexts, tasks, or vigilance level states. This might be especially true for studies characterized by training/test approach limitations. As indicated earlier, optimal generalizability should subtend a "one size fits all" approach as much as possible, but this was not necessarily achieved by the studies selected for review. Third, we did not attempt to distinguish the different subconditions under the general term hypovigilance (e.g., such as fatigue vs. sleepiness) that may have different physiological manifestations. These different states may potentially need different diagnostic or prediction models given that their mechanisms of origin may vary (e.g., circadian rhythm vs. cognitive resource depletion vs. homeostasis). Having considered all the models together to investigate for potential methods to measure hypovigilance was relevant for the context of this scoping review, which aimed at defining the general state of the literature regarding hypovigilance and outlining existing gaps. Still, before providing more specific suggestions as to the best ways to measure hypovigilance induced by, for example, fatigue, sedation, and cognitive overload, a more granular analysis of the literature is needed. Finally, more detailed information about the performances of the models would have been useful to collect. Indeed, understanding whether the models found here can outperform gold standard prediction and diagnostic models could represent a key tool for researchers and developers interested in applying the techniques reviewed in real-life settings. However, because of the heterogeneity in the studies, this information was not always available and/or comprehensively collected. Parts of this information can be found in Appendix S3, but it must be regarded with caution given the lack of consensus about defining hypovigilance and the heterogeneity in the choice of performance metrics and ways to measure them.

5 | CONCLUSION

Hypovigilance is considered an important cause of many accidents and hazardous situations in several fields. Therefore, improved hypovigilance detection capacities could help to facilitate how it is managed and, in turn, to increase safety and security of people and infrastructures. In the current scoping review, we identified the main techniques used to assess hypovigilance using sensor-based models. As outlined, the choice of sensors to infer hypovigilance was relatively similar between all papers. Indeed, many focused on the central nervous system via EEG (or NIRS) and/or the peripheral nervous system with eye-tracking technologies and/or ECG/PPG-based measures. Among the selected papers, a majority relied on a prediction approach and used machine learning, rather than a diagnostic approach. Although the training and feature computing methods remained unclear in most of the papers, some common methods such as the use of SVM for model training were highlighted. However, certain gaps remain, in particular concerning the different training and performance evaluation methods used. For example, some models were trained using a leave-one-out approach, whereas other models were trained for each participant individually. Overall, the ability to infer hypovigilance (possibly in real time) with a reduced invasiveness has great potential in many contexts from military to medical, and the current state of the literature on this topic is likely to show important progress in the upcoming years.

AUTHOR CONTRIBUTIONS

Alexandre Marois: Data curation; methodology; supervision; visualization; writing – original draft; writing – review and editing. **Maëlle Kopf:** Conceptualization; data curation; formal analysis; investigation; methodology; resources; software; writing – original draft; writing – review and editing. **Michelle Fortin:** Data curation; formal analysis; investigation; methodology; resources; software; validation; writing – original draft; writing – review and editing. **Maxime Huot-Lavoie:** Conceptualization; investigation; methodology; resources; writing – review and editing. **Alexandre Martel:** Conceptualization; investigation; methodology. **J. Gordon Boyd:** Conceptualization; writing – review and editing. **Jean-François Gagnon:** Conceptualization; supervision; writing – review and editing. **Patrick Archambault:** Conceptualization; funding acquisition; supervision; writing – review and editing.

ACKNOWLEDGMENTS

We are grateful to William Witteman and Raymond Klein for their valuable help and feedback during data collation and manuscript preparation.

FUNDING INFORMATION

This research was supported by a Mitacs grant attributed to Maxime Lavoie-Huot and a Research Bursary awarded to Alexandre Martel by the Faculty of Medicine at Université Laval. Patrick Archambault holds a Fonds de recherche du Québec – Santé Clinical Scholar Award.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any personal or financial relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Alexandre Marois https://orcid.org/0000-0002-4127-4134
Maxime Huot-Lavoie https://orcid.org/0000-0001-7378-9776
Patrick M. Archambault https://orcid.org/0000-0002-5090-6439

REFERENCES

Abd-Elfattah, H. M., Adbelazeim, F., & Elshennawy, S. (2015). Physical and cognitive consequences of fatigue: A review. *Journal of Advanced Research*, *6*, 351–358. https://doi.org/10.1016/j.jare.2015.01.011

Akerstedt, T. (2000). Consensus statement: Fatigue and accidents in transport operations. *Journal of Sleep Research*, *9*, 395. https://doi.org/10.1046/j.1365-2869.2000.00228.x

*Akerstedt, T., Ingre, M., Kecklund, G., Anund, A., Sandberg, D., Wahde, M., Philip, P., & Kronberg, P. (2010). Reaction of sleepiness indicators to partial sleep deprivation, time of day and time on task in a driving simulator—The DROWSI project. *Journal of Sleep Research*, *19*, 298–309. https://doi.org/10.1111/j.1365-2869.2009.00796.x

Aljurf, T. M., Olaish, A. H., & BaHammam, A. S. (2018). Assessment of sleepiness, fatigue, and depression among Gulf Cooperation Council commercial airline pilots. *Sleep & Breathing*, *22*, 411–419. https://doi.org/10.1007/s11325-017-1565-7

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

Antoniadi, A. M., Du, Y., Guendouz, Y., Wei, L., Mazo, C., Becker, B. A., & Mooney, C. (2021). Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: A systematic review. *Applied Sciences*, *11*, 5088. https://doi.org/10.3390/app11115088

Arun, S., Murugappan, M., & Sundaraj, K. (2011). Hypovigilance warning system: A review on driver alerting techniques. *IEEE Control and System Graduate Research Colloquium*, *2011*, 65–69. https://doi.org/10.1109/ICSGRC.2011.5991831

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Reviews of Neuroscience*, *28*, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

*Awais, M., Badruddin, N., & Drieberg, M. (2017). A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability. *Sensors*, *17*, 1991. https://doi.org/10.3390/s17091991

Bafna, T., & Hansen, J. P. (2021). Mental fatigue measurement using eye metrics: A systematic literature review. *Psychophysiology*, *58*, e13828. https://doi.org/10.1111/psyp.13828

Bendak, S., & Rashid, H. S. J. (2020). Fatigue in aviation: A systematic review of the literature. *International Journal of Industrial Ergonomics*, *76*, 102928. https://doi.org/10.1016/j.ergon.2020.102928

Bier, L., Wolf, P., Hilsenbek, H., & Abendroth, B. (2020). How to measure monotony-related fatigue? A systematic review of fatigue measurement methods for use on driving tests. *Theoretical Issues in Ergonomics Science*, *21*, 22–55. https://doi.org/10.1080/1463922X.2018.1529204

Blackhurst, J. L., Gresham, J. S., & Stone, M. O. (2012). The quantified warrior: How DoD should lead human performance augmentation. *Armed Forces Journal*. http://armedforcesjournal.com/the-quantified-warrior/

Boksem, M. A., & Tops, M. (2008). Mental fatigue: Costs and benefits. *Brain Research Reviews*, *59*, 125–139. https://doi.org/10.1016/j.brainresrev.2008.07.001

Boudaya, A., Bouaziz, B., Chaabene, S., Chaari, L., Ammar, A., & Hökelmann, A. (2020). EEG-based hypo-vigilance detection using convolutional neural network. In M. Jmaiel, M. Mokhtari, B. Abdulrazak, H. Aloulou, & S. Kallel (Eds.), *The impact of digital technologies on public health in developed and developing countries. ICOST 2020. Lecture Notes in Computer Science* (Vol. *12157*, pp. 69–78). Springer. https://doi.org/10.1007/978-3-030-51517-1_6

Bouret, S., & Sara, S. J. (2004). Reward expectation, orientation of attention and locus coeruleus-medial frontal cortex interplay during learning. *European Journal of Neuroscience*, *20*, 791–802. https://doi.org/10.1111/j.1460-9568.2004.03526.x

Brookhuis, K. A., & de Waard, D. (1993). The use of psychophysiology to assess driver status. *Ergonomics*, *36*, 1099–1110. https://doi.org/10.1080/00140139308967981

Brown, I. D. (1982). Driving fatigue. *Endeavour*, *6*, 83–90. https://doi.org/10.1016/0160-9327(82)90109-0

Buckley, R. J., Helton, W. S., Innes, C. R. H., Dalrymple-Alford, J. C., & Jones, R. D. (2016). Attention lapses and behavioural microsleeps during tracking, psychomotor vigilance, and dual tasks. *Consciousness and Cognition*, *45*¸, 174–183. https://doi.org/10.1016/j.concog.2016.09.002

Carretta, T. R., & French, G. A. (2012). Combating vigilance decrements in a sustained attention task: Lack of support for the utility of a cognitive intervention secondary task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*, 1446–1450. https://doi.org/10.1177/1071181312561407

*Choi, H.-S., Min, S., Kim, S., Bae, H., Yoon, J.-E., Hwang, I., Oh, D., Yun, C.-H., & Yoon, S. (2019). Learning-based instantaneous drowsiness detection using wired and wireless electroencephalography. *IEEE Access*, *7*, 146390–146402. https://doi.org/10.1109/ACCESS.2019.2946053

*Chua, E. C.-P., Tan, W.-Q., Yeo, S.-C., Lau, P., Lee, I., Mien, I. H., Puvanendran, K., & Gooley, J. J. (2012). Heart rate variability can be used to estimate sleepiness-related decrements in psychomotor vigilance during total sleep deprivation. *Sleep*, *35*, 325–334. https://doi.org/10.5665/sleep.1688

Connor, J., Norton, R., Ameratunga, S., Robinson, E., Civil, I., Dunn, R., Bailey, J., & Jackson, R. (2002). Driver sleepiness and risk of serious injury to car occupants: Population based case control study. *BMJ*, *324*, 1125. https://doi.org/10.1136/bmj.324.7346.1125

de Rooij, M., & Weeda, W. (2020). Cross-validation: A method every psychologist should know. *Advances in Methods and Practices in Psychological Science*, *3*, 248–263. https://doi.org/10.1177/2515245919898466

Dehais, F., Causse, M., Vachon, F., Régis, N., Menant, E., & Tremblay, S. (2014). Failure to detect critical auditory alerts in the cockpit: Evidence for inattentional deafness. *Human Factors*, *56*, 631–644. https://doi.org/10.1177/0018720813510735

Dinges, D. F., Mallis, M. M., Maislin, G., & Powell, J. W. (1998). *Evaluation of techniques for ocular measurement as an index of fatigue and as the basis for alertness management* (Rep. DOT HS 808 762). National Highway Traffic Safety Administration.

Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., Aptowicz, C., & Pack, A. I. (1997). Cumulative sleepiness, mood disturbance and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. *Sleep*, *20*(4), 267–277.

Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers*, *17*, 652–655. https://doi.org/10.3758/BF03200977

Doran, S. M., van Dongen, H. P. A., & Dinges, D. F. (2001). Sustained attention performance during sleep deprivation: Evidence of state instability. *Archives Italiennes de Biologie*, *139*, 253–267. https://doi.org/10.4449/aib.v139i3.503

Dorrian, J., Lamond, N., Kozuchowski, K., & Dawson, D. (2008). The driver vigilance telemetric control system (DVTCS): Investigating sensitivity to experimentally induced sleep loss and fatigue. *Behavior Research Methods*, *40*, 1016–1025. https://doi.org/10.3758/BRM.40.4.1016

Drew, B. J., Harris, P., Zègre-Hemsey, J. K., Mammone, T., Schindler, D., Salas-Boni, R., Bai, Y., Tinoco, A., Ding, Q., & Hu, X. (2014). Insights into the problem of alarm fatigue with physiologic monitor devices: A comprehensive observational study of consecutive intensive care unit patients. *PLoS One*, *9*, e110274. https://doi.org/10.1371/journal.pone.0110274

Ducharme-Crevier, L. (2021). *Pediatric delirium screening in the PICU via EEG (PEDEEGO)*. U.S. National Library of Medicine, Clinical Trials Study Design Repository. No. NCT04846023. https://clinicaltrials.gov/ct2/show/NCT04846023

Duffy, M., & Feltman, K. A. (2022). *A systematic literature review of operator state detection using physiological measures* (USAARL-TECH-TR–2023-11; p. 33). U.S. Army Aeromedical Research Laboratory.

Elam, M., Svensson, T. H., & Thorén, P. (1986). Locus coeruleus neurons and sympathetic nerves: Activation by cutaneous sensory afferents. *Brain Research*, *366*, 254–261. https://doi.org/10.1016/0006-8993(86)91302-8

Ely, E. W., Gautam, S., Margolin, R., Francis, J., May, L., Speroff, T., Truman, B., Dittus, R., Bernard, R., & Inouye, S. K. (2001). The impact of delirium in the intensive care unit on hospital length of stay. *Intensive Care Medicine*, *27*, 1892–1900. https://doi.org/10.1007/s00134-001-1132-2

Ely, E. W., Margolin, R., Francis, J., May, L., Truyman, B., Dittus, R., Speroff, T., Gautam, S., Bernard, G. R., & Inouye, S. K. (2001). Evaluation of delirium in critically ill patients: Validation of the confusion assessment method for the intensive care unit (CAM-ICU). *Critical Care Medicine*, *29*, 1370–1379. https://doi.org/10.1097/00003246-200107000-00012

EuroNCAP. (2017). *EuroNCAP 2025 Roadmap Technical Report*. EuroNCAP.

Fiest, K. M., Soo, A., Lee, C. H., Niven, D. J., Ely, E. W., Doig, C. J., & Stelfox, H. T. (2021). Long-term outcomes in ICU patients with delirium: A population-based cohort study. *American Journal of Respiratory and Critical Care Medicine*, *204*, 412–420. https://doi.org/10.1164/rccm.202002-0320OC

Foote, S. L., Berridge, C. W., Adams, L. M., & Pineda, J. M. (1991). Electrophysiological evidence for the involvement of the locus coeruleus in alerting, orienting, and attending. *Progress in Brain Research*, *88*, 521–532. https://doi.org/10.1016/S0079-6123(08)63831-5

Foote, S. L., Freedman, R., & Oliver, A. P. (1975). Effects of putative neurotransmitters on neuronal activity in monkey auditory cortex. *Brain Research*, *86*, 229–242. https://doi.org/10.1016/0006-8993(75)90699-x

*François, C., Hoyoux, T., Langohr, T., Wertz, J., & Verly, J. G. (2016). Tests of a new drowsiness characterization and monitoring system based in ocular parameters. *International Journal of Environmental Research and Public Health*, *13*, 174. https://doi.org/10.3390/ijerph13020174

François, C., Wertz, J., Kirkove, M., & Verly, J. G. (2014). Evaluation of the performance of an experimental somnolence quantification system in terms of reaction times and lapses. In *Proceedings of the 36th Annual International Conference of IEEE Engineering in Medicine and Biology Society* (pp. 5820–5823). Chicago, USA. https://doi.org/10.1109/EMBC.2014.6944951

Friedl, K. E. (2018). Military applications of soldier physiological monitoring. *Journal of Science and Medicine in Sport*, *21*, 1147–1153. https://doi.org/10.1016/j.jsams.2018.06.004

Fuller, R. G. C. (1983). *Effects of prolonged driving on time headway adopted by HGV drivers* (Research Note 83–33). Research Institute for the Behavioral and Social Sciences, U.S. Army.

Goodie, A. S., & Crooks, C. L. (2004). Time-pressure effects on performance in a base-rate task. *The Journal of General Psychology*, *131*, 18–28. https://doi.org/10.3200/GENP.131.1.18-28

Gregory, K. B., Winn, W., Johnson, K., & Rosekind, M. R. (2010). Pilot fatigue survey: Exploring fatigue factors in air medical operations. *Air Medical Journal*, *29*, 309–319. https://doi.org/10.1016/j.amj.2010.07.002

Gronau, Q. F., & Wagenmakers, E.-J. (2019). Limitations of Bayesian leave-one-out cross-validation for model selection. *Computational Brain & Behavior*, *2*, 1–11. https://doi.org/10.1007/s42113-018-0011-7

Gual, N., Garcia-Salmones, M., & Perez, L. M. (2019). Diagnosing delirium in patients with dementia, a great challenge [Diagnóstico del delirium en pacientes con demencia, un gran reto]. *Medicina Clínica (English Edition)*, *153*, 284–289. https://doi.org/10.1016/j.medcle.2019.05.004

Guenther, U., Popp, J., Koecher, L., Muders, T., Wrigge, H., Ely, E. W., & Putensen, C. (2010). Validity and reliability of the CAM-ICU flowsheet to diagnose delirium in surgical ICU patients. *Journal of Critical Care*, *25*, 144–151. https://doi.org/10.1016/j.jcrc.2009.08.005

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI-Explainable artificial intelligence. *Science Robotics*, *4*, eaay7120. https://doi.org/10.1126/scirobotics.aay7120

Gunzelmann, G., & Gluck, K. A. (2009). An integrative approach to understanding and predicting the consequences of fatigue on cognitive performance. *Cognitive Technology*, *14*(1), 14–25.

*Guo, M., Li, S., Wang, L., Chai, M., Chen, F., & Wei, Y. (2016). Research on the relationship between reaction ability and mental state for online assessment of driving fatigue. *International Journal of Environmental Research and Public Health*, *13*, 1174. https://doi.org/10.3390/ijerph13121174

Haney, C. (2003). Mental health issues in long-term solitary and “supermax” confinement. *Crime & Delinquency*, *49*, 124–156. https://doi.org/10.1177/0011128702239239

*He, Q., Li, W., Fan, X., & Fei, Z. (2016). Evaluation of driver fatigue with multi-indicators based on artificial neural network. *IET Intelligent Transport Systems*, *10*, 555–561. https://doi.org/10.1049/iet-its.2015.0021

Higgins, J. P. T., Savocic, J., Page, M. J., & Sterne, J. A. C. (2019). *Revised Cochrane risk-of-bias tool for randomized trials (RoB 2)*. https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool/current-version-of-rob-2

Horne, J. A., & Reyner, L. A. (1999). Vehicle accidents related to sleep: A review. *Occupational and Environmental Medicine*, *56*, 289–294. https://doi.org/10.1136/oem.56.5.289

Hosker, C., & Ward, D. (2017). Hypoactive delirium. *BMJ*, *357*, j2047. https://doi.org/10.1136/bmj.j2047

*Hu, S., & Zheng, G. (2009). Driver drowsiness detection with eyelid related parameters by support vector machine. *Expert Systems with Applications*, *36*, 7651–7658. https://doi.org/10.1016/j.eswa.2008.09.030

Hughes, R. W., & Marsh, J. E. (2017). The functional determinants of short-term memory: Evidence from perceptual-motor interference in verbal serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 537–551. https://doi.org/10.1037/xlm0000325

Inouye, S. K. (1994). The dilemma of delirium: Clinical and research controversies regarding diagnosis and evaluation of delirium in hospitalized elderly medical patients. *The American Journal of Medicine*, *97*, 278–288. https://doi.org/10.1016/0002-9343(94)90011-6

Jackson, C. A., & Earl, L. (2006). Prevalence of fatigue among commercial pilots. *Occupational Medicine*, *56*, 263–268. https://doi.org/10.1093/occmed/kql021

Jarosch, O., Bellem, H., & Bengler, K. (2019). Effects of task-induced fatigue in prolonged conditional automate driving. *Human Factors*, *61*, 1186–1199. https://doi.org/10.1177/0018720818816226

Kashevnik, A., Shchedrin, R., Kaiser, C., & Stocker, A. (2021). Driver distraction detection methods: A literature review and framework. *IEEE Access*, *9*, 60063–60076. https://doi.org/10.1109/ACCESS.2021.3073599

Kerick, S., Metcalfe, J., Fend, T., Ries, A., & McDowell, K. (2013). *Review of fatigue management technologies for enhanced military vehicle safety and performance*. Report No. ARL-TR-6571. Army Research Laboratory.

Kranjec, J., Begus, S., Drnovsek, J., & Gersak, G. (2014). Novel methods for noncontact heart rate measurement: A feasibility study. *IEEE Transactions on Instrumentation and Measurement*, *63*, 838–847. https://doi.org/10.1109/TIM.2013.2287118

*Kudinger, T., Sofra, N., & Riener, A. (2020). Assessment of the potential of wrist-worn wearable sensors for driver drowsiness detection. *Sensors*, *20*, 1029. https://doi.org/10.3390/s20041029

Larue, G. S., Rakotonirainy, A., & Petit, A. N. (2010). Predicting driver's hypovigilance on monotonous roads: Literature review. In *1st International Conference on Driver Distraction and Inattention*. Gothenburg, Sweden.

*Leng, L. B., Giin, L. B., & Chung, W.-Y. (2015). Wearable driver drowsiness detection system based on biomedical and motion sensors. *IEEE Sensors*, *2015*, 1–4. https://doi.org/10.1109/ICSENS.2015.7370355

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science*, *5*, 69. https://doi.org/10.1186/1748-5908-5-69

Li, G., & Chung, W.-Y. (2014). Estimation of eye closure degree using EEG sensors and its application in driver drowsiness detection. *Sensors*, *14*, 17491–17515. https://doi.org/10.3390/s140917491

*Li, G., & Chung, W. Y. (2015). A context-aware EEG headset system for early detection of driver drowsiness. *Sensors*, *15*, 20873–20893. https://doi.org/10.3390/s150820873

Li, G., & Chung, W. Y. (2022). Electroencephalogram-based approaches for driver drowsiness detection and management: A review. *Sensors*, *22*, 1100. https://doi.org/10.3390/s22031100

*Li, G., Lee, B.-L., & Chung, W.-Y. (2015). Smartwatch-based wearable EEG system for driver drowsiness detection. *IEEE*

Sensors Journal, *15*, 7169–7180. https://doi.org/10.1109/JSEN.2015.2473679

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ*, *339*, b2700. https://doi.org/10.1136/bmj.b2700

Lim, J., & Dinges, D. F. (2008). Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences*, *1129*, 305–322. https://doi.org/10.1196/annals.1417.002

Lin, S. T., Tan, Y. Y., Chua, P. Y., Tey, L. K., & Ang, C. H. (2012). PERCLOS threshold for drowsiness detection during real driving. *Journal of Vision*, *12*, 546. https://doi.org/10.1167/12.9.546

Lipton, Z. C. (2018). The mythos of model interpretability. *Queue*, *16*, 31–57. https://doi.org/10.1145/3236386.3241340

Liu, Z., Shao, Y., & Wang, P. (2012). *Road traffic safety engineering*. Higher Education Press.

*Lopez de la O, J., Ibanez, N. R., Gonzalez, M. N., Vicente, J. M. B., Garcia Gonzalez, M. A., Castro, J. R., & Fernandez-Chimeno, M. (2012). Development of a system to test somnolence detectors with drowsy drivers. *Procedia—Social and Behavioral Sciences*, *48*, 2058–2070. https://doi.org/10.1016/j.sbspro.2012.06.1179

Luna, F. G., Barttfeld, P., Martín-Arévalo, E., & Lupiáñez, J. (2022). Cognitive load mitigates the executive but not the arousal vigilance decrement. *Consciousness and Cognition*, *98*, 103263. https://doi.org/10.1016/j.concog.2021.103263

Lyznicki, J. M., Doege, T. C., Davis, R. M., & Williams, M. A. (1998). Sleepiness, driving, and motor vehicle crashes. *JAMA*, *279*, 1908–1913. https://doi.org/10.1001/jama.279.23.1908

*Maccora, J., Manousakis, J. E., & Anderson, C. (2018). Pupillary instability as an accurate, objective marker of alertness failure and performance impairment. *Journal of Sleep Research*, *28*, e12739. https://doi.org/10.1111/jsr.12739

Marois, A., Lafond, D., Williot, A., Vachon, F., & Tremblay, S. (2020). Real-time gaze-aware cognitive support system for security surveillance. *Proceedings of the Human Factors and Ergonomics Society 2020 Annual Meeting*, *64*, 1145–1149. https://doi.org/10.1177/1071181320641274

Marzo, A., Totah, N. K., Neves, R. M., Logothetis, N. K., & Eschenko, O. (2014). Unilateral electrical stimulation of rat locus coeruleus elicits bilateral response of norepinephrine neurons and sustained activation of medial prefrontal cortex. *Journal of Neurophysiology*, *111*, 2570–2588. https://doi.org/10.1152/jn.00920.2013

Matthews, G., & Desmond, P. A. (2002). Task-induced fatigue states and simulated driving performance. *The Quarterly Journal of Experimental Psychology: Section A*, *55*, 659–686. https://doi.org/10.1080/02724980143000505

Matthews, G., Neubauer, C., Saxby, D. J., Wohleber, R. W., & Lin, J. (2019). Dangerous intersections? A review of studies of fatigue and distraction in the automated vehicle. *Accident; Analysis and Prevention*, *126*, 85–94. https://doi.org/10.1016/j.aap.2018.04.004

May, J. F., & Baldwin, C. L. (2009). Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transportation Research Part F: Traffic Psychology and Behaviour*, *12*, 218–224. https://doi.org/10.1016/j.trf.2008.11.005

*Mehreen, A., Anwar, S. M., Haseeb, M., Majid, M., & Ullah, M. O. (2019). A hybrid scheme for drowsiness detection using wearable sensors. *IEEE Sensors Journal*, *19*, 5119–5126. https://doi.org/10.1109/JSEN.2019.2904222

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Miller, J. C., & Melfi, M. L. (2006). *Causes and effects of fatigue in experienced military aircrew*. Report No. AFRL-HE-BR-TR-2006-0071. Air Force Research Laboratory.

Mogilever, N. B., Zuccarelli, L., Burles, F., Iaria, G., Strapazzon, G., Bessone, L., & Coffey, E. B. J. (2018). Expedition cognition: A review and prospective of subterranean neuroscience with spaceflight applications. *Frontiers in Human Neuroscience*, *12*, 407. https://doi.org/10.3389/fnhum.2018.00407

Mohanavelu, K., Lamshe, R., Poonguzhali, S., Adalarasu, K., & Jagannath, M. (2017). Assessment of human fatigue during physical performance using physiological signals: A review. *Biomedical & Pharmacology Journal*, *10*, 1887–1896. https://doi.org/10.13005/bpj/1308

*Mu, Z., Hu, J., & Yin, J. (2017). Driving fatigue detecting based on EEG signals of forehead area. *International Journal of Pattern Recognition and Artificial Intelligence*, *31*, 1750011. https://doi.org/10.1142/S0218001417500112

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*, 143. https://doi.org/10.1186/s12874-018-0611-x

*Nguyen, T., Ahn, S., Jang, H., Jun, S. C., & Kim, J. G. (2017). Utilization of a combined EEG/NIRS system to predict driver drowsiness. *Scientific Reports*, *7*, 43933. https://doi.org/10.1038/srep43933

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, *131*, 510–532. https://doi.org/10.1037/0033-2909.131.4.510

Nieuwenhuis, S., De Geus, E. J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology*, *48*, 162–175. https://doi.org/10.1111/j.1469-8986.2010.01057.x

Oh, J., Cho, D., Park, J., Na, S. H., Kim, J., Heo, J., Shin, C. S., Kim, J. J., Park, J. Y., & Lee, B. (2018). Prediction and early detection of delirium in the intensive care unit by using heart rate variability and machine learning. *Physiological Measurement*, *39*, 035004. https://doi.org/10.1088/1361-6579/aaab07

O'Hagan, A. D., Issartel, J., McGinley, E., & Warrington, G. (2018). A pilot study exploring the effects of sleep deprivation on analogue measures of pilot competencies. *Aerospace Medicine and Human Performance*, *89*, 609–615. https://doi.org/10.3357/AMHP.5056.2018

Oken, B. S., Salinsky, M. C., & Elsas, S. M. (2006). Vigilance, alertness, or sustained attention: Physiological basis and measurement. *Clinical Neurophysiology*, *117*, 1885–1901. https://doi.org/10.1016/j.clinph.2006.01.017

Palinkas, L. A. (2003). The psychology of isolated and confined environments: Understanding human behavior in Antarctica. *American Psychologist*, *58*, 353–363. https://doi.org/10.1037/0003-066x.58.5.353

Palinkas, L. A., Johnson, J. C., & Boster, J. S. (2004). Social support and depressed mood in isolated and confined environments. *Acta Astronautica*, *54*, 639–647. https://doi.org/10.1016/s0094-5765(03)00236-4

Papadopoulos, G. C., & Parnavelas, J. G. (1990). Distribution and synaptic organization of dopaminergic axons in the lateral geniculate nucleus of the rat. *The Journal of Comparative Neurology*, *294*, 356–361. https://doi.org/10.1002/cne.902940305

Parasuraman, R. (1986). Vigilance, monitoring, and search. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance*. Vol. *2*, Cognitive processes and performance (pp. 1–39). John Wiley & Sons.

Parasuraman, R., Warm, J. S., & See, J. E. (1998). Brain systems of vigilance. In R. Parasuraman (Ed.), *The attentive brain* (pp. 221–256). The MIT Press.

Parnandi, A., Son, Y., & Gutierrez-Osuna, R. (2013). A control-theoretic approach to adaptive physiological games. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (pp. 7–12). Geneva, Switzerland. https://doi.org/10.1109/ACII.2013.8

Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends in Cognitive Sciences*, *13*, 160–166. https://doi.org/10.1016/j.tics.2009.01.006

Peters, M. D. J., Godfrey, C. M., Khalil, H., McInerney, P., Parker, D., & Soares, C. B. (2015). Guidance for conducting systematic scoping review. *International Journal of Evidence-Based Healthcare*, *13*, 141–146. https://doi.org/10.1097/XEB.0000000000000050

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*, 73–89. https://doi.org/10.1146/annurev-neuro-062111-150525

Petrilli, R. M., Roach, G. D., Dawson, D., & Lamond, N. (2006). The sleep, subjective fatigue, and sustained attention of commercial airline pilots during an international pattern. *Chronobiology International*, *23*, 1357–1362. https://doi.org/10.1080/07420520601085925

Philip, P., & Akerstedt, T. (2006). Transport and industrial safety, how are they affected by sleepiness and sleep restriction? *Sleep Medicine Reviews*, *10*, 347–356. https://doi.org/10.1016/j.smrv.2006.04.002

Philip, P., Taillard, J., Klein, E., Sagaspe, P., Charles, A., Davies, W. L., Guilleminault, C., & Bioulac, B. (2003). Effect of fatigue on performance measured by a driving simulator in automobile drivers. *Journal of Psychosomatic Research*, *55*, 197–200. https://doi.org/10.1016/S0022-3999(02)00496-8

Previc, F. H., Lopez, N., Ercoline, W. R., Daluz, C. M., Workman, A. J., Evans, R. H., & Dillon, N. A. (2009). The effects of sleep deprivation on flight performance, instrument scanning, and physiological arousal in pilots. *The International Journal of Aviation Psychology*, *19*, 326–346. https://doi.org/10.1080/10508410903187562

Rajkowski, J., Majczynski, H., Clayton, E., & Aston-Jones, G. (2004). Activation of monkey locus coeruleus neurons varies with difficulty and behavioral performance in a target detection task. *Journal of Neurophysiology*, *92*, 361–371. https://doi.org/10.1152/jn.00673.2003

Rechtschaffen, A., & Kales, A. (1968). *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. Public Health Service, U.S. Government Printing Office.

Robertson, I. H., & O'Connell, R. (2010). Vigilant attention. In A. C. Nobre & J. T. Coull (Eds.), *Attention and time* (pp. 79–88). Oxford University Press.

Rush, B., Celi, L. A., & Stone, D. J. (2019). Applying machine learning to continuously monitored physiological data. *Journal of Clinical Monitoring and Computing*, *33*, 887–893. https://doi.org/10.1007/s10877-018-0219-z

Russo, M. B., Stetz, M. C., & Thomas, M. L. (2005). Monitoring and predicting cognitive state and performance via physiological correlates of neuronal signals. *Aviation, Space, and Environmental Medicine*, *76*(7 Suppl), C59–C63.

Sahayadhas, A., Sundaraj, K., & Murugappan, M. (2012). Detecting driver drowsiness based on sensors: A review. *Sensors*, *12*, 16937–16953. https://doi.org/10.3390/s121216937

Sahayadhas, A., Sundaraj, K., Murugappan, M., & Palaniappan, R. (2015). Physiological signal based detection of driver hypovigilance using higher order spectra. *Expert Systems with Applications*, *42*, 8669–8677. https://doi.org/10.1016/j.eswa.2015.07.021

Salluh, J. I. F., Wang, H., Schneider, E. B., Nagaraja, N., Yenokyan, G., Damluji, A., Serafim, R. B., & Stevens, R. D. (2015). Outcome of delirium in critically ill patients: Systematic review and meta-analysis. *BMJ*, *350*, h2538. https://doi.org/10.1136/bmj.h2538

Salvan, L., Marois, A., Kopf, M., & Gagnon, J.-F. (2022). Sensors-enabled human state monitoring system for tactical settings. In *2022 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)* (pp. 55–61). Salerno, Italy. https://doi.org/10.1109/CogSIMA54611.2022.9903277

*Salvati, L., d'Amore, M., Fiorentino, A., Pellegrino, A., Sena, P., & Villecco, F. (2021). On-road detection of driver fatigue and drowsiness during medium-distance journeys. *Entropy*, *23*, 135. https://doi.org/10.3390/e23020135

Sara, S. J., & Bouret, S. (2012). Orienting and reorienting: The locus coeruleus mediates cognition through arousal. *Neuron*, *76*, 130–141. https://doi.org/10.1016/j.neuron.2012.09.011

Schwarz, J., & Fuchs, S. (2018). Validating a "real-time assessment of multidimensional user state" (RASMUS) for adaptive human-computer interaction. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 704–709). Miyazaki, Japan. https://doi.org/10.1109/SMC.2018.00128

Sheridan, T. (1987). Supervisory control. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 1244–1268). John Wiley & Sons.

Slater, J. D. (2008). A definition of drowsiness: One purpose for sleep? *Medical Hypotheses*, *71*, 641–644. https://doi.org/10.1016/j.mehy.2008.05.035

Sommer, D., & Golz, M. (2010). Evaluation of PERCLOS based current fatigue monitoring technologies. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology* (pp. 4456–4459). Buenos Aires, Argentina. https://doi.org/10.1109/IEMBS.2010.5625960

Southwick, S. M., Bremner, J. D., Rasmusson, A., Morgan, C. A., III, Arnsten, A., & Charney, D. S. (1999). Role of norepinephrine in the pathophysiology and treatment of posttraumatic stress disorder. *Biological Psychiatry*, *46*, 1192–1204. https://doi.org/10.1016/S0006-3223(99)00219-X

Sterne, J. A., Savovic, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H. Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., ... Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, *366*, l4898. https://doi.org/10.1136/bmj.l4898

Suresh, H., & Guttag, J. V. (2019). A framework for understanding unintended consequences of machine learning. *arXiv preprint*. https://doi.org/10.48550/arXiv.1901.10002

Tefft, B. C. (2010). *The prevalence and impact of drowsiness driving*. AAA Foundation for Traffic Safety.

Tjoa, E., & Guan, C. (2021). A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, *32*, 4793–4813. https://doi.org/10.1109/TNNLS.2020.3027314

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., ... Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, *169*, 467–473. https://doi.org/10.7326/M18-0850

van Schie, M. K. M., Lammers, G. J., Fronczek, R., Middelkoop, H. A. M., & van Dijk, J. G. (2021). Vigilance: Discussion of related concepts and proposal for a definition. *Sleep Medicine*, *83*, 175–181. https://doi.org/10.1016/j.sleep.2021.04.038

*Vicente, J., Laguna, P., Bartra, A., & Bailón, R. (2011). Detection of driver's drowsiness by means of HRV analysis. *Computing in Cardiology*, *38*, 89–92.

Wang, C.-A., & Munoz, D. P. (2015). A circuit for pupil orienting responses: Implications for cognitive modulation of pupil size. *Current Opinion in Neurobiology*, *33*, 134–140. https://doi.org/10.1016/j.conb.2015.03.018

Warm, J. S., Dember, W. N., & Hancock, P. A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 183–200). Lawrence Erlbaum Associates.

Waterhouse, B. D., Moises, H. C., & Woodward, D. J. (1998). Phasic activation of the locus coeruleus enhances responses of primary sensory cortical neurons to peripheral receptive field stimulation. *Brain Research*, *790*, 33–44. https://doi.org/10.1016/S0006-8993(98)00117-6

Weinger, M. B., & Slagle, J. (2002). Human factors research in anesthesia patient safety: Techniques to elucidate factors affecting clinical task performance and decision making. *Journal of the American Medical Informatics Association*, *9*, S58–S63. https://doi.org/10.1197/jamia.M1229

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M., Sterne, J. A., Bossuyt, P. M., & the QUADAS-2 Group. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, *155*, 529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

Wierwille, W. W., & Ellsworth, L. A. (1994). Evaluation of driver drowsiness by trained raters. *Accident Analysis & Prevention*, *26*, 571–581. https://doi.org/10.1016/0001-4575(94)90019-1

Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., Mallett, S., & the PROBAST Group. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine*, *170*, 51–58. https://doi.org/10.7326/M18-1376

*Yamada, Y., & Kobayashi, M. (2018). Detecting mental fatigue from eye-tracking data gathered while watching video: Evaluation in younger and older adults. *Artificial Intelligence in Medicine*, *91*, 39–48. https://doi.org/10.1016/j.artmed.2018.06.005

Zhang, H., Wu, C., Yan, X., & Qiu, T. Z. (2016). The effect of fatigue driving on car following behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, *43*, 80–89. https://doi.org/10.1016/j.trf.2016.06.017

*Zhang, X., Li, J., Liu, Y., Zhang, Z., Wang, Z., Luo, D., Zhou, X., Zhu, M., Salman, W., Hu, G., & Wang, C. (2017). Design of a fatigue detection system for high-speed trains based on driver vigilance using a wireless wearable EEG. *Sensors*, *17*, 486. https://doi.org/10.3390/s17030486

Zhao, S., Liu, J., Gong, Z., Lei, Y., OuYang, X., Chan, C. C., & Ruan, S. (2020). Wearable physiological monitoring system based on electrocardiography and electromyography for upper limb rehabilitation training. *Sensors*, *20*, 4861. https://doi.org/10.3390/s20174861

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1.

Table S1.1. Preferred reporting items for systematic reviews and meta-analyses extension for scoping reviews (PRISMA-ScR) checklist.

Appendix S2.

Table S2.1. Research strategy for the PsycINFO database.

Table S2.2. Research strategy for the Inspec database.

Appendix S3.

Table S3.1. Details on the different models generated for the diagnosis or the prediction of a hypovigilant state.

How to cite this article: Marois, A., Kopf, M., Fortin, M., Huot-Lavoie, M., Martel, A., Boyd, J. G., Gagnon, J.-F., & Archambault, P. M. (2023). Psychophysiological models of hypovigilance detection: A scoping review. *Psychophysiology*, *00*, e14370. https://doi.org/10.1111/psyp.14370