

Methods and Measures for Mental Stress Assessment in Surgery: A Systematic Review of 20 Years of Literature

Mastaneh Torkamani-Azar , Member, IEEE, Ahreum Lee , and Roman Bednarik 

I. INTRODUCTION

Abstract—Real-time mental stress monitoring from surgeons and surgical staff in operating rooms may reduce surgical injuries, improve performance and quality of medical care, and accelerate implementation of stress-management strategies. Motivated by the increase in usage of objective and subjective metrics for cognitive monitoring and by the gap in reviews of experimental design setups and data analytics, a systematic review of 71 studies on mental stress and workload measurement in surgical settings, published in 2001–2020, is presented. Almost 61% of selected papers used both objective and subjective measures, followed by 25% that only administered subjective tools - mostly consisting of validated instruments and customized surveys. An overall increase in the total number of publications on intraoperative stress assessment was observed from mid-2010 s along with a momentum in the use of both subjective and real-time objective measures. Cardiac activity, including heart-rate variability metrics, stress hormones, and eye-tracking metrics were the most frequently and electroencephalography (EEG) was the least frequently used objective measures. Around 40% of selected papers collected at least two objective measures, 41% used wearable devices, 23% performed synchronization and annotation, and 76% conducted baseline or multi-point data acquisition. Furthermore, 93% used a variety of statistical techniques, 14% applied regression models, and only one study released a public, anonymized dataset. This review of data modalities, experimental setups, and analysis techniques for intraoperative stress monitoring highlights the initiatives of surgical data science and motivates research on computational techniques for mental and surgical skills assessment and cognition-guided surgery.

Index Terms—Surgery, mental stress, intraoperative stress, workload, cognitive monitoring, physiological monitoring, surgical data science, robot-assisted surgery, microsurgery, operating room, wearable sensors, multimodal acquisition, data analytics, machine learning, heart-rate variability, eye-tracking, electroencephalography.

Manuscript received 13 July 2021; revised 1 April 2022 and 17 May 2022; accepted 5 June 2022. Date of publication 13 June 2022; date of current version 9 September 2022. This work was supported in part by the Emil Aaltonen Foundation under Grant 64898, and in part by the Academy of Finland under Grant 334658. (Corresponding author: Mastaneh Torkamani-Azar.)

The authors are with the Interactive Technologies Research Group, School of Computing, University of Eastern Finland, 80110 Joensuu, Finland (e-mail: mastaneh.torkamani@uef.fi; ahreum.lee@uef.fi; roman.bednarik@uef.fi).

Digital Object Identifier 10.1109/JBHI.2022.3182869

STRESS affects around one third of employees in the world and is prevalent among health-care professionals such as physicians, surgeons, and nurses [1], [2]. Prolonged exposure to stress can lead to burnout, substance abuse, work absence, and substantial financial loss [3]. Long working hours, time pressure, sleep deprivation, complex surgical interventions and new surgical technology, interactions with traumatizing events, noise, and interruptions, multitasking, communication, and working with or supervising junior colleagues are key factors that affect cognitive states and psychomotor performance of surgeons and staff in the operating rooms (ORs) [4]–[7]. Excessive levels of cognitive impairment in the OR materialize as increased distress, frustration, and workload [5], [6], [8]. According to descriptive models of [9], heightened levels of acute mental stress result in attention narrowing, distraction, loss in working-memory, and preservation. When occurred during surgery, elevated levels of “intraoperative” mental stress also impair decision making, judgement, teamwork, and performance of individual personnel and surgical teams [7], [10].

In their classic review of stress, appraisal, and coping, Lazarus and Folkman defined stress as a “particular relationship between the person and environment” with psychological and physiological responses affected by both external stimuli and internal or personal features [11]. In psychology, stress is defined as “a dynamic process that occurs when an individual appraises situational demands as exceeding available resources” [12]. This cognitive component is also defined as a response to demands and external stimuli that surpass the coping abilities of an organism [13]. This perspective associates stress with resource limitation, workload, and role strain [14]. Empirical research indicate that stress can increase neural activity, available working memory, and experienced cognitive load for learners and decrease task processing efficiency [15]. The conceptual overlaps and close associations of workload and stress have imposed challenges for their measurements in the past 30 years [16], [17]. Therefore, in this systematic review, we focus on real-time assessment of psychophysiological stress in the OR such that elements of cognitive workload and arousal –both tense and energetic– and physiological or autonomic arousal are covered [18].

A stressful situation can be self-evaluated and appraised as being harmful, inducing a threat, or merely acting as a challenge that provides new opportunities [9]. Stress can have negative effects on task performance; however, positive stress and successful stress adaptation can improve skills, alertness, motivation, stimulation, and sense of enjoyable task completion [6], [19]–[21]. Research and anecdotes have shown that expert

surgeons efficiently control their brain activation under stress while young trainees tend to use beta-blockers to overcome anxiety [22], [23]. Therefore, studies focusing on quality and safety as two main markers of surgical procedures address the need for improvement of technical and non-technical skills [24], [25], identification of stress factors, assessment of acute stress, and implementation of stress-coping strategies using, for example, closed-loop human-computer systems for simulating surgical situations [26], [27] and technology-assisted meditation/relaxation [28], [29].

In general, no measurement tool or technique can directly assess cognitive states such as stress or workload [30], [31]. Therefore, stress is commonly estimated by measuring “its effects” through individual and subjective perceptions and objective assessment of aroused physiologic states and responses [17], [30], [32]. While using subjective questionnaires and other self-report tools has been a common method for evaluating intraoperative levels of “self-perceived” stress [17], [24], [33], objective assessment of these states from elevated sympathetic, parasympathetic, and electrophysiological activities has been less frequently exercised due to technological complications, difficulties with correct interpretation, interference with task performance, and privacy concerns [24], [34]–[36]. We believe that spatio-temporal patterns of psychophysiological stress, recorded during different stages of surgical procedures, need to be compared with those of expert surgeons for mastering surgical skill acquisition. This real-time monitoring of acute mental stress from surgeons and surgical staff in the OR may also help to: (a) predict periods of high stress, (b) analyze effects of novel image-guided and minimally invasive surgical techniques on workload, stress, anxiety, and technical skills in real and simulated surgeries [37]–[40], (c) accelerate implementation of stress-management interventions to reduce occupational stress and improve clinical outcomes and patients’ well-being [41]–[44], (d) introduce mental skill training and multi-session monitoring of skill acquisition in trainees [45]–[47], (e) improve surgical automaticity and performance [7], [17], [48], and (f) develop intelligent surgical assistance in smart ORs [36], [49].

In recent years, advances in wearable neurotechnologies that enable real-time and non-invasive measurement of physiological signals without causing discomfort for users has brought a new interest in monitoring and quantifying cognitive states from engineers, athletes, and surgeons [6], [50]. Despite these advances and the surge in applying machine learning (ML) for mental stress assessment from multimodal datasets [51], the progress in implementing ML for objective stress and workload detection in the OR has been slower and limited to a few research groups [42], [44]. Challenges in quantifying mental states in real life [52], intra- and inter-subject variabilities [53], obtaining ontologies and annotating surgical stages [36], [54], privacy and data protection rights [55], and highly dynamic and case-dependent decisions and actions-interactions of surgical staff that result in data interpretation challenges [56] are a few contributing obstacles. These challenges are being addressed by new academic programs in digital health, data science, and AI-assisted technologies¹ and by public forums on data science and digital healthcare².

¹<https://www.feinberg.northwestern.edu/sites/augmentedintelligence/centers/medical-education-data-science-digital-health.html>, <https://www.surgerycdt.com/>

²<https://grand-challenges.embs.org/2021datascience/>

Timeliness of a New Systematic Review. Among the growing pool of systematic reviews on intraoperative nontechnical skills [25] and states such as fatigue [57] and workload [17], and effects of distractions [58] and music [59] on surgical performance, we identified four major reviews related to our scope on intraoperative assessment of acute mental stress and workload. In two 2006 and 2010 systematic reviews of morbidity syndromes in minimal access surgery [31] and effects of stress on performance [30], adopted stress measures included questionnaires, electrocardiogram (ECG), heart rate (HR), heart rate variability (HRV), skin conductance level (SCL), electromyogram (EMG), electrooculogram (EOG), salivary cortisol, performance scores, and communication frequencies. Another review of 28 studies from 1998 to 2010 reported that questionnaires, completion time, economy of motion, cognitive and psychomotor error metrics, final product quality, hand efficiency, HR, HRV, and SCL were used for assessment of technical and nontechnical skills [60].

More recently, 33 studies on acute stress measurement during real operations were reviewed [32] that had used objective metrics such as thermal and electrodermal activity (EDA). Finally, a review of 84 studies that measured surgeons’ cognitive workload named NASA Task Load index (NASA-TLX), Surgery Task Load Index (SURG-TLX), and HRV as the most frequent subjective–post-hoc– and objective measures [17]. Use of eye-tracking, electroencephalography (EEG), and functional near-infrared spectroscopy (fNIRS) was also reported in this review. Thus, it was only after 2010 that modern sensor-based measures such as pupilometrics, EEG, cortical hemodynamics, and functional connectivity were adopted for quantification of intraoperative mental stress [44], [61]–[65]. This trend seems to be accelerated by advances in wireless sensors and lightweight headsets that enable acquisition of physiological signals without imposing ergonomic restrictions on the surgical staff [27]. Data analysis has also witnessed widespread changes in the past decade [56]. However, none of the earlier systematic reviews on stress assessment addressed key aspects on experimental design or data analytics tools.

Contributions of the Current Systematic Review. The earlier and highly valuable systematic reviews on intraoperative stress measurements had limited scopes and inevitable shortcomings. Motivated by gaps in a thorough review of robust experimental designs and modern data analytic methods for stress assessment from multi-modal datasets, this study aims to present a replicable systematic review of methods and measures for objective assessment of mental and psychophysiological stress in surgical settings. This effort is also motivated by the fact that our initial search resulted in finding a large range of descriptive phrases for stress, including cognitive, mental, systematic, psychologic, physiologic, task induced, and perceived or subjective stress. Interestingly, stress, strain, anxiety, workload, and mental/cognitive demand have been mentioned and assessed interchangeably in the literature [17]. This vague definition is emphasized in discussions on unclear relationships between physiologic activation, stress, and self-reported anxiety [66] and multidimensionality and interaction of stress and mental fatigue [67].

In this work, we consider the earlier definition of stress (dynamic process related to self-appraisal of demands exceeding limited resources) [12]. In addition, we perform a literature search with two main keywords –mental stress/distress and cognitive workload– to cover studies on intraoperative assessment of these states due to their conceptual overlap. We aim to

provide the following contributions: (a) A review of cognitive states and skills studied in association with mental distress/stress and workload of OR professionals in surgical settings, (b) a review of temporal patterns in adopting objective and subjective measures for intraoperative stress assessment, (c) a thorough and novel review of experimental setups such as baseline recordings, wearable systems, multi-modal datasets, processing, synchronization, and data analysis, and (d) highlighting future directions for research and dissemination of results on intraoperative stress assessments.

Research Questions. The following questions were defined to clarify the research scope [68]. What constituted the studied participants, procedures, and research topics (Sections III-A and III-B); how often subjective and objective measures were used and if they were administered separately or in conjunction with each other, what were the most frequently used features in each group, and what consensus or discord were reported in their associations with intraoperative mental stress (Section III-C); what experimental setup and recording conditions were implemented, if measurements were calibrated or compared with baseline values, if multi-modal datasets were acquired and synchronized with each other and surgical events, and if wearable sensors were utilized (Section III-D); what processing and data analysis techniques were implemented, and if data was available as open source (Section III-E).

II. MATERIALS AND METHODS

Study selection was conducted with keywords on mental stress and mental distress. Since an initial search on Google Scholar demonstrated that not all papers on mental stress assessment and surgery were retrieved by these criteria, and considering the hypotheses on associations of workload, working memory, and memory load with mental stress [17], a second search was conducted to scan papers on cognitive load or workload. This step was performed to identify the literature that measured stress as a side variable of workload during intraoperative assessments.

A. Search Strategy

The search strategy protocol followed guidelines from PRISMA statement [69] and its extension PRISMA-S [70]. The following eligibility criteria were devised to define the structure of the current systematic review.

- Topic: “Mental stress”, “mental distress”, “mental workload”, “cognitive workload”, “mental load”, and “cognitive load”.
- Location: Intraoperative assessments; comparison of stress levels after surgical sessions with pre-operative baselines.
- Surgical operating systems: Laparoscopy, endoscopy, virtual reality, microsurgery, real, open, and simulated surgery.
- Potential participants: Mentors, mentees, teams, residents, novices, fellows, trainees, nurses, anesthesiologists, etc.
- Data acquisition: Questionnaires, physiological signals, salivary samples, white blood cells (WBC), respiration, and noninvasive techniques such as the EEG, MEG, MRI, blood oxygenation, fNIRS, ocular activities with EOG,

eye-tracking, and pupillometrics (gaze, pupil, saccade, and fixation), EMG, accelerometer, head positions, ECG and HRV, skin conductance, skin resistance, etc.

- Publication date: 1 January 2001 to 31 December 2020.
- Exclusion criteria: Opinion and commentary articles, posters and abstracts, non-English articles, clinical trials, and studies on stress assessment in patients and their caregivers were out of the scope of this review. Cohort or cross-sectional surveys from healthcare workers about general well-being, work styles, job satisfaction, occupational stress, burnout, coping strategies, perceived and moral stress [71]–[74], and papers on design and validation of new scales or training programs without near real-time assessment of psychological data [75] were excluded. Studies on physical and ergonomic stress, pain, or strain instead of mental stress, and those on mental stress of staff who monitored the patients but were not involved in surgical tasks or simulations [76] were also removed. However, papers with measurements collected from surgeons during office days versus on-call and operating days [77], [78], and papers on baseline and post-intervention measurements of surgical performance and mental stress were included [21], [46], [79], [80].

B. Information Sources and Search Terms

Cochrane Library, Ovid MEDLINE, PubMed, and Scopus were accessed through university subscriptions in fall 2020 and January 2021 to obtain the latest publications. When possible, search results were limited to publications in English. Database-specific search terms and limitations are accessible from our Online Repository. Search results were exported to spread sheets for further analysis.

C. Selection Process

Titles and abstracts of search results were screened by two authors in line with the inclusion and exclusion criteria in Section II-A. If an article was not publicly available, its authors were contacted on ResearchGate. Articles were discarded from this review if they were not accessible until February 2021.

D. Study Characteristics

The SPIDER (Sample, Phenomenon of Interest, Design, Evaluation, Research type) mnemonic [81] was used to define and limit key dimensions and study characteristics to be analyzed in this systematic review. Using this tool, the following pieces of information were extracted from abstracts and full texts. “Sample” consisted of participants’ features such as surgical specialties, expertise, and roles. The hypotheses on stress/workload factors and their associations with technical and nontechnical skills and cognitive states were considered as “Phenomena of Interest”. Dimension included surgical procedures, experiment conditions (e.g., silence or environmental noise, time pressure or self-paced operations) and their comparisons. “Evaluation” consisted of subjective assessment tools and their types (e.g. validated questionnaires or custom surveys), objective and physiological measures and their sub-dimensions such as computed features,

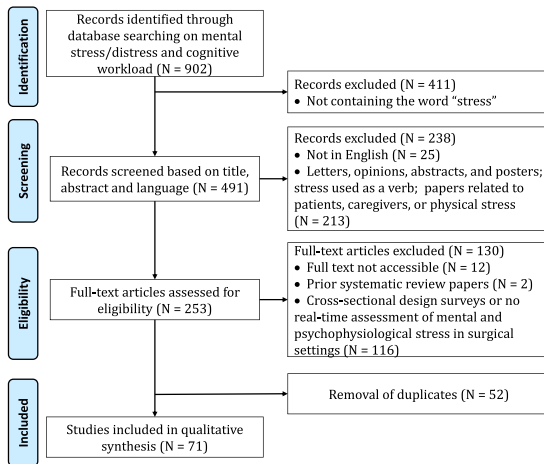


Fig. 1. Flowchart of the systematic review design and study selection for mental stress assessment according to PRISMA guidelines. N denotes the number of articles.

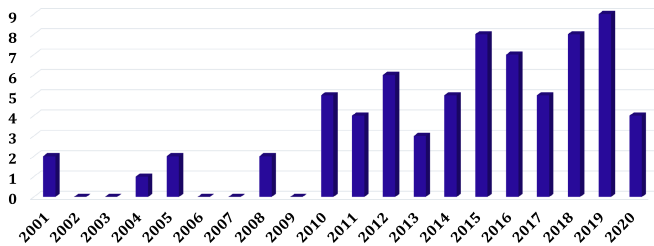


Fig. 2. Temporal distribution of systematically selected papers on intraoperative mental stress assessment, 2001–2020. The years 2019, 2015, and 2018 saw the highest number of distinct annual publications that passed the selection criteria.

their associations with stress, use of wearable systems, multi-modal datasets, synchronization practices, baseline recording sessions, and data analysis techniques. Online links were also investigated to locate supplementary, open-access materials. Finally, studies were chronologically sorted to extract information on temporal evolution of subjective and objective measures.

III. RESULTS

A total of 421 and 481 citations were retrieved after searching databases on mental stress/distress and cognitive workload, respectively. The PRISMA flow diagram of the search strategy is presented in Fig. 1. Applying the exclusion criteria and removing duplicate articles resulted in the systematic selection of 71 papers for a qualitative review of mental stress assessment in real and simulated surgery. Temporal distribution of these papers is presented in Fig. 2. The Supplementary Table presents all study characteristics, extracted with SPIDER mnemonic, that will be discussed in the rest of this section.

A. Surgical Procedures

Table I shows conducted surgical procedures for 71 selected papers. Distributions are not mutually exclusive: 41 studies involved a single surgical procedure while 20 studies compared performance and stress using two different procedures, and one included data from open, laparoscopic, and converted

TABLE I
SURGICAL PROCEDURES USED IN THE SELECTED STUDIES

Surgical Procedure	Studies	Number (%)
Laparoscopic surgery	[21], [24], [27], [29], [38]–[40], [44], [45], [61], [63], [64], [82], [84], [86]–[94], [98]–[113]	39 (54.93%)
Simulated surgery and physical models	[10], [26], [28], [45], [46], [66], [80], [94], [98], [99], [106], [110], [114]–[119]	18 (25.35%)
Open or conventional surgery	[4], [27], [77], [82], [86], [96]–[98], [102], [112], [120]–[126]	17 (23.94%)
Robot-assisted surgery	[26], [38], [40], [77], [86]–[91], [102], [126]	12 (16.90%)
Virtual reality and games	[24], [26]–[28], [106], [118], [127], [128]	8 (11.27%)
Microsurgery	[62], [65]	2 (2.82%)

TABLE II
PARTICIPANTS EXPERTISE LEVELS IN THE SELECTED STUDIES

Expertise Level	Studies	Number (%)
Expert, primary, and attending surgeons	[4], [6], [24], [38]–[40], [62], [77], [80], [82]–[84], [86], [89]–[92], [97]–[100], [103], [105], [106], [111], [118], [122]–[124], [126], [127]	32 (45.07%)
Surgical interns, trainees, and residents	[6], [24], [26], [27], [38], [40], [44], [46], [64], [66], [66], [77], [78], [80], [84], [90], [92], [94], [96], [105], [106], [108], [109], [114], [115], [117], [118], [122], [124], [127]	30 (42.25%)
Senior trainees and fellows; less experienced surgeons	[4], [6], [24], [44], [78], [80], [92], [98], [99], [106], [109], [118], [121], [122], [124], [125], [127]	17 (23.94%)
Technology novices	[10], [21], [24], [27], [29], [45], [61], [62], [65], [80], [94], [104], [107], [110], [113], [128]	16 (22.53%)
Medical students and interns	[21], [61], [65], [87], [88], [94], [101], [107], [108], [110], [111], [113], [118], [119], [127], [128]	14 (19.72%)
Surgical teams	[46], [66], [92], [124], [126]	5 (7.04%)

surgeries [82]. Five studies did not mention the procedure used in stress-monitoring sessions [6], [78], [83]–[85]. Out of 39 studies that administered laparoscopy, 11 compared laparoscopy with other procedures or focused on different laparoscopic techniques, one paper compared laparoscopy with conventional surgery in terms of physiological correlates of mental strain [86], and seven papers analyzed differences between standard and robot-assisted laparoscopy in terms of cognitive stress, physical strain, workload, performance, attention, coping, and gaze control [38], [40], [87]–[91]. One paper compared induced stress and workload during single-incision and conventional laparoscopy [39], another administered sessions with one-port and four-port laparoscopy to quantify postural constraints, perceived workload, task difficulty, and stress [92], and one analyzed performance differences in relatively easy versus difficult laparoscopy due to different visual scanning and multi-tasking demands [93]. Simulated surgeries were performed with mannequins, physical models for mesh and knot tying tasks, or surgical simulators. Two papers addressed telerobotic operations [94], [95], and three used biofeedback devices for stress management [83], [96], [97].

B. Study Participants

Table II shows the distribution of the majority of participant groups in terms of expertise levels. Nursing and non-medical

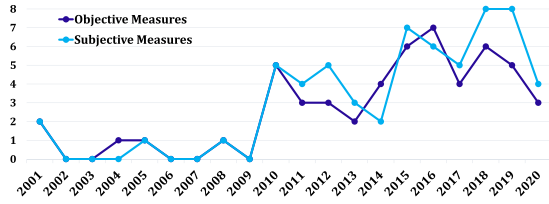


Fig. 3. Temporal distribution of selected studies in terms of using objective and subjective measures.

students [85], [93], [128], surgical technicians [66], [92], [118], and professionals from anesthesiology [66], [116], [126], emergency medicine [114], medicine [28], [83], nursing [66], [112], [118], [126], and primary care [83] were other included participants.

Intraoperative stress across surgical roles and expertise.

Seventeen papers (23.94%) compared surgical performance and stress markers between individuals from two expertise levels –novice versus expert or junior versus senior residents and surgeons [38], [62], [63], [66], [78], [86], [90], [96], [105], [109], [117], [121], or among surgical professionals with diverse professional backgrounds [6], [44], [64], [84], [118]. Six studies compared stress and technical performance when participants switched their roles between primary and assistant surgeons [4], [77], [86], [92], [108], [122], and two papers measured stress, anxiety, and workload from teams of surgeons, nurses, technicians, and assistants [66], [92].

C. Objective and Subjective Measures

Stress perception induces a variety of cognitive, emotional, and behavioral responses [99] that can affect surgical outcomes depending on personality types, emotional intelligence, trait anxiety, and stress-coping capabilities [30], [66]. However, due to the inherently subjective nature of self-rated questionnaires, assessment of brain functions and cognitive skills using objective metrics is a valuable asset when surgical skills do not demonstrate large inter-subject variations [44]. Furthermore, applying both measurement categories may enhance the assessment of underlying cognitive states [17]. Fig. 3 demonstrates the number of selected studies that used objective and subjective metrics for intraoperative assessment of mental stress. Besides the generally upward trend in both modalities, a total of 25 and 31 papers used objective and subjective measures during the 2016–2020 period.

An in-depth analysis showed that 18 papers (25.35%) had *only* administered subjective questionnaires and interviews, 10 papers (14.08%) *only* collected objective, physiological data, and 43 papers (60.56%) used both objective and subjective measures to assess the stress levels for at least one time point during the surgery. To demonstrate the variety of these measures, Fig. 4 presents the timeline of all questionnaires and physiological modalities used in the selected papers.

In what follows, we present a categorization of utilized subjective measures in Section III-C1, identify the most frequently used instruments, and investigate the cognitive states and skills that were subjectively studied in the selected studies. We then categorize all modalities of objective measures and present findings, according to the Evaluation dimension of the SPIDER mnemonic, in Section III-C2.

TABLE III
SUBJECTIVE MEASURES USED IN INTRAOPERATIVE STRESS ASSESSMENT

Subjective Measures	Studies	Number (%)
Validated instruments only	[6], [10], [21], [24], [26], [27], [29], [38]–[40], [44], [46], [62]–[66], [83], [85], [87], [88], [90], [93], [95], [104]–[107], [110], [111], [118], [119], [122], [126]–[128]	36 (50.70%)
Validated instruments & customized surveys	[45], [61], [82], [89], [92], [94], [100]–[103], [109], [113], [114], [117]	14 (19.72%)
No subjective instruments	[4], [77], [78], [84], [91], [96], [99], [115], [123], [125]	10 (14.08%)
Customized surveys only	[97], [98], [108], [112], [116]	5 (7.04%)
Observations & validated instruments	[28], [121], [124]	3 (4.23%)
Observations only	[86], [120]	2 (2.82%)
Observations, validated, & customized surveys	[80]	1 (1.41%)

1) Subjective Measures for Intraoperative Stress Assessment:

The selected studies administered a variety of questionnaires for self assessments by participants or evaluations by researchers and independent observers. Several papers also recorded the demographics and prior experience with administered surgical instruments and technologies. Subjective measures used for intraoperative assessments can be divided to three categories: (a) Instruments such as NASA-TLX and STAI listed in Fig. 4 that were validated, occasionally adapted to different languages, and presented to participants before or after each surgical session, (b) customized or newly developed surveys combining several dimensions and metrics such as those in [94], [98], and (c) observations and real-time reports by researchers or independent observers, and post-session evaluation of recorded videos for assessment of surgical progress, stress, and anxiety.

As shown in Table III, 36 papers administered between one and five validated instruments, and 14 papers used one or two validated instruments in addition to customized surveys. Only six studies utilized observational assessments and four of them paired these reports with validated instruments or customized surveys. Among 54 papers (76.06%) that used at least one validated measurement tool, the study by Stefanidis *et al.* [29] stood out for having moderated five validated questionnaires. Overall, NASA-TLX [129] ($n = 19$), full STAI ($n = 10$), SURG-TLX [130] ($n = 9$), and six-item STAI [131] ($n = 8$) were the most frequently used validated tools.

Subjectively assessed cognitive states and skills. Overall, 25 reviewed papers collected subjective responses for mental and psychological stress and strain [6], [10], [24], [26], [28], [29], [65], [66], [80], [83], [85], [87], [88], [90], [91], [93], [98], [101], [102], [105], [108], [110], [112], [114], [117], and 11 studies extended their assessments to physical, postural, and ergonomic stress and strain [24], [27], [39], [80], [89], [94], [100]–[102], [105], [108]. Furthermore, 26 articles assessed perceived workload, cognitive load, and task demands [10], [21], [26], [29], [38]–[40], [46], [61]–[65], [87]–[89], [92], [93], [103], [106], [113], [118], [119], [121], [122], [124]. Four papers analyzed teamwork and communication flow among team members [80], [108], [112], [120], and three examined stress coping profiles [80], [85], [87]. Furthermore, two studies collected data about trauma history [28], [114], another two about mental and general well-being [100], [105], and one study about previous experience with virtual reality (VR) [28].

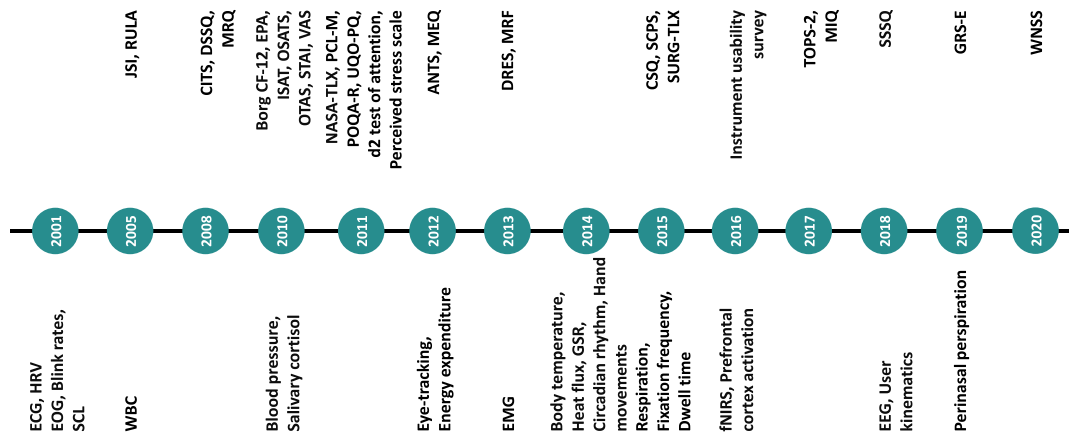


Fig. 4. Evolution of (top row) subjective and (bottom row) objective measures for intraoperative stress assessment, including the year in which a particular modality appeared first in the present sample. JSI: Job strain index; RULA: Rapid upper limb assessment; CITS: Coping inventory of task stress; DSSQ: Dundee stress state quest.; MRQ: Multiple resources quest.; EPA: End product assessment; ISAT: Imperial stress assessment tool; OSATS: Objective structured assessment of technical skill; OTAS: Observational teamwork assessment for surgery; STAI: State trait anxiety inventory; VAS: Visual analog scale; PCL-M: PTSD Checklist-military version; POQA-R: Personal and organizational quality assessment-revised; UQO-PQ: Université du Québec en outaouais presence quest.; ANTS: Anaesthetists' non-technical skills; MEQ: Musical experience quest.; DRES: Demand resource evaluation score; MRF: Mental readiness form; CSQ: Clinical stress quest.; SCPS: Stress-coping patterns scale; TOPS: Test of performance strategies; MIQ: Mental imagery quest.; SSSQ: Short stress state quest.; GRS-E: Global rating scale of endovascular performance; WNSS: Weinstein's noise sensitivity scale.

Self-perceived or graded surgical performance [29], [46], [63], [65], [80], [109], [116], [118], [126], satisfaction and confidence in performance [100], [101], motor functions [107], nontechnical performance [114], [116], concentration and selective attention [29], [98], [103], mental imagery [29], noise sensitivity [82], music reactions [61], [104], [113], fear and perceptions of mistakes [128], and frustration [100] were other states and skills for which subjective measures were collected.

2) Objective Measures for Intraoperative Stress Assessment: Table IV presents 12 identified categories for objective measurements, names of physiological features, and lists of corresponding studies. The majority of reviewed papers had reported heart activity, salivary stress hormones, ocular activity, blood pressure, and electrodermal activity. It should be mentioned that energy expenditure and heat flux, kinematics, respiration rate, and cortical activation were not covered by earlier reviews on assessment of acute intraoperative stress [31], [32], [132], nor by systematic reviews on impacts of mental stress and nontechnical skills on surgical performance [30], [60]. However, a review on intraoperative workload measurement covered these modalities and pointed to the interchangeability of mental stress and cognitive workload [17]. In what follows, we present a few observations on these objective measures with stress/workload markers exclusively along the boundary defined in Section II-D, and leave details on statistically significant observations to the “Significant Objective Measures” column in the Supplementary Table.

a) Cardiac activity: Cardiac activity, either calculated offline from recorded ECG signals or measured in real time with wearable devices, was the most frequently used physiological modality among selected studies ($n = 40$). Generally, an increase in average HR indicates elevated sympathetic activity and physical expenditure, and decrease in HRV corresponds to smaller changes in time interval between consecutive heart beats, more arousal and strain, and less cardiac relaxation [4]. HRV is controlled by the autonomic nervous system [125] and is calculated through temporal features (SDNN, RMSSD,

pNN50) and spectral features (LF: sympathetic activity, HF: parasympathetic activity, LF/HF) [133]. An increase in LF/HF shows imbalance in the autonomic tone and increase in mental stress [125]. The majority of reviewed studies computed or read measured heart rates, followed by spectral and temporal feature analysis for HRV. However, as summarized in Column K of the Supplementary Table, findings are divided in observing significant or consistent changes in HR and HRV parameters under perceived or designed stressful conditions and significant associations with subjective ratings, and have to be carefully interpreted considering the timeline of reading measurements [107], calculation of absolute values or relative changes from baselines [4], [6], [80], [125], different or low sample sizes [10], [108], and the type of metric–temporal vs. spectral– and body positions during HRV recording [134].

b) Stress-related hormones levels: Cortisol level was the most highly analyzed feature in 11 studies that collected saliva samples. Associations of salivary cortisol with stressful situations were either contradictory [80], [111] or did not result in statistically significant differences among control and intervention groups [21], [83], [117]. The highest cortisol levels were observed 20–30 minutes after task completion in a high-stress scenario [114].

c) Ocular activity: Eye activities were recorded using surface EOG [98], eye-tracking glasses [38], [63], [106], [107], [115], [121], or eye trackers connected to microscope oculars [62]. Gaze and fixation metrics were computed from image frames using commercial or custom scripts [38], [107], [115], [121]. While significant associations were reported between less frequent blinking with higher self-perceived workload and frustration, lack of baseline data was a challenge for comparisons [106]. Changes in pupil dilation from the baseline were the largest under noise distractions than silent or music conditions, in line with higher subjective workload levels. Furthermore, experienced surgeons showed smaller changes in pupil dilation than less experienced operators [63]. Finally, average and SD of pupil size changes during suturing successfully classified surgical

TABLE IV
CATEGORIES FOR OBJECTIVE MEASURES USED IN INTRAOPERATIVE STRESS ASSESSMENT

Objective Measures	Studies	Number (%)
1. Cardiac activity		40 (56.34%)
HR features (mean, median, min, and max)	[4], [6], [10], [21], [24], [29], [39], [40], [44], [45], [61], [64], [77], [78], [80], [83], [84], [86], [89], [91], [95]–[97], [100]–[102], [110], [111], [113], [114], [116]–[118], [122], [123], [125]	36
HRV spectral features (LF, HF, LF/HF)	[4], [6], [21], [77], [80], [84], [86], [91], [96], [97], [100], [118], [125]	13
HRV temporal features (SDNN, RMSSD, pNN50)	[4], [6], [38], [61], [80], [100], [115], [124], [125]	9
ECG analysis, impedance cardiography	[103], [107]	2
2. Stress-related hormones		11 (15.49%)
Salivary cortisol	[21], [24], [39], [80], [83], [92], [102], [103], [111], [114], [117]	11
Alpha-amylase, testosterone, DHEA	[21], [92], [103]	3
3. Ocular activity		8 (11.27%)
Eye blink frequency and duration [EOG and video]	[98], [106]	2
Gaze metrics, fixation, dwell time [video and IR]	[38], [107], [115], [121]	4
Pupil dilation (mean and SD), rate of change, entropy	[62], [63], [121]	3
4. Blood-related metrics		7 (9.86%)
Blood pressure, mean arterial pressure	[83], [101], [107], [111], [113], [123]	
White blood cell (WBC) count	[78]	
5. Electrodermal activity		6 (8.45%)
Galvanic skin response (GSR)	[66], [95], [108]	3
Skin conductance level (tonic response, mean, range)	[95], [98], [111]	3
Skin resistance, electro-sympathicography	[99]	1
6. EMG activity		4 (5.63%)
Masseter tone	[96], [97]	2
Maximum EMG activity, root mean square	[89]	1
Mean power spectral density (MNP)	[95]	1
7. fNIRS		4 (5.63%)
PFC activity ([HbO ₂] and [HHb], channel activation)	[21], [40], [44], [64]	
PFC asymmetry (laterality index)	[21]	
8. Respiration rate	[96], [97], [122]	3 (4.23%)
9. Energy expenditure and heat flux	[4], [108]	2 (2.82%)
10. Body temperature		2 (2.82%)
Facial temperate [IR thermal imaging camera]	[108]	1
Perinasal perspiration signal	[65]	1
11. Body movements		2 (2.82%)
Hand movements, video-based	[115]	1
Arm kinematics (angular velocity, linear acceleration, jerk) [IMUs]	[95]	1
12. EEG activity		1 (1.41%)
Custom Engagement and Workload indices	[95]	1

expertise while fatigue, caffeine, and illumination changes were possible reasons for no classification improvement from blink rate features [62].

d) Blood-related metrics: Among six studies that measured systolic and diastolic blood pressure (BP) or mean arterial pressure (MAP), four studies did not report any significant differences between control and intervention groups or in changes from baseline (pre-simulation) to post-task [83], [101], [111], [113]. In [107], regression analysis showed that during the baseline phase, evaluating the task as challenging was associated with higher cardiac output and lower peripheral resistance – computed from outputs of a wearable device and measured MAP. A significantly higher BP was seen during real neurosurgeries than in resting states [123], and increase in WBC after on-call duties with respect to pre-call days was inversely proportional with the level of surgical training [78].

e) Electrodermal activity: Intraoperative sympathetic arousal or nervous response have been measured in different ways in the literature. More recent papers used wearable devices to measure GSR with an armband [66], [95]. In a study with an ergonomic station, experienced surgeons showed less variability in SCL levels across the tasks, and SCL correlated with the reported increase in mental stress [98]. Another study reported higher GSR for surgeons of varying expertise after a crisis simulation and a negative correlation between state anxiety and maximal GSR during debriefing [66]. Finally, SCL analysis was discarded due to its large changes when participants held laparoscopic tools [111].

f) EMG activity: In [89], surface electrodes recorded muscular activities during robotic and standard laparoscopy. The EMG-based root mean square (RMS), standardized with respect to a voluntary maximal contraction (MVC) period, was significantly larger for all muscles and both sides during the standard laparoscopy than in robot-assisted surgery. In [95], the EMG-based spectral mean power (MNP) was the largest during joint space teleoperation scenario and interpreted as a high muscle fatigue [95]. In [96], critical surgery events were logged and a higher HR and significantly larger masseter tone, measured from EMG signals by a biofeedback device, were detected in high workload segments. However, in endoscopic simulations of [97], the masseter tone and respiratory rate were only used for detection of artifacts and tense situations.

g) Functional near-infrared spectroscopy (fNIRS): Four studies quantified prefrontal activation (PFC) through changes in cortical oxygenation levels using non-portable fNIRS and characterized temporal and motor demands for different surgical technologies under time pressure and normal conditions. Both cortical hemispheres were almost identically activated during a multi-week laparoscopic and motor learning program [21]; but no significant correlation was found between these and other physiological markers of stress. An improved and careful operation with robotic laparoscopy under time pressure resulted in larger responses in the right ventrolateral PFC – associated with improved sustained attention and distraction suppression [40]. In a similar experiment, junior and intermediate residents showed more bilateral PFC activation during self-paced tasks while

senior residents had consistent activation patterns in both conditions [44].

h) Respiration metrics: The breathing rate, measured via a chest belt, was not statistically different between primary and assistant surgeons and not correlated with surgery duration in [122]. Respiration events were unprocessed in [127], but used for interpreting LF and HF HRV bands [96] and locating moving and talking artifacts [97]. One study concluded that changes in breathing frequency were too slow for real-time data acquisition [96], and another pointed out that measuring respiration through expansion and during movement may induce artifacts to the recorded data [122].

i) Energy expenditure (EE) and heat flux: HRV-based EE was larger in primary surgical fellows than in assisting senior surgeons in [4]. Wearable armbands with a heat flux sensor measured the heat convection subset of total thermal energy in [108]. The standard deviation of heat flux during simulation training was correlated with subjective stress levels.

j) Body temperature: Facial temperature (FT) was measured alone or in addition to EDA [65], [108]. When using an infrared thermal imaging camera inside an OR with controlled temperature, significantly lower mean frontal head temperature was reported for participants with a poor performance in a simulation laparoscopic training [108]. Instantaneous perinasal perspiration signal, calculated from facial thermal imagery data as a measure of sympathetic arousal response, was not linearly correlated with task completion time or proficiency scores in a multi-week training program [65].

k) Body movements and kinematics: One group measured hand and gaze movements during open surgery training using an eye tracker, and calculated the total hand movement time based knot-tying videos [115]. Arm kinematics were collected via body inertial measurement units (IMU) during simulated robotic operations [95]. Performance metrics from 3D kinematics were improved in specific scenarios although respective physiological features from EEG, HR, and SCL were elevated.

l) EEG activity: Through a commercial package and validated framework, EEG activity recorded with a wearable headset was used to calculate probability-based engagement and workload metrics from power spectral density (PSD). These indices were higher in experimental scenarios that also demonstrated the highest mental and physical demands and effort from NASA-TLX, EMG, HRV, and SCL features.

D. Experiment Design and Data Acquisition Practices

In this section, we review the selected studies in terms of experimental design aspects and data acquisition techniques.

1) Wearable Systems and Wireless Data Transfer: Wearable systems combine sensor suites with data analytics, enable continuous acquisition of physiological and biomechanical data, and provide flexibility and mobility without interfering with daily activities of operators of critical tasks [50], [135]. In surgical settings, wearable sensors with continuous measurement capabilities can augment psychometric instruments to detect cognitive load variations and classify its sub-types [34]. When used for stress and fatigue monitoring, these systems can help with introducing stress coping and intervention techniques in surgery [6], [67]. Twenty-nine reviewed papers (40.85%) used at least one device with wearable or wireless technologies for data acquisition and transmission. Thirteen studies measured cardiac

activities via sensors mounted in belt straps, batches, or body-worn heart monitoring systems [6], [21], [39], [40], [61], [77], [78], [100], [103], [116], [122], [124], [125], and 12 others used a combination of elastic straps and wristbands for displaying the processed information [4], [10], [24], [29], [45], [80], [102], [110], [114], [115], [117], [118]. Other uses include transmitting surface-recorded EMG via Bluetooth or WiFi modules [89], [95], recording EEG with a wireless headset [95], using wearable eye trackers with a belt-worn recorder [63], and recording HRV and BP with battery-operated biofeedback devices [83].

2) Multi-Modal Datasets: Data fusion, inspired by processing data from a variety of sensory sources in living beings, is a common practice in clinical decision making and disease prognosis [136], [137]. Relying on synchronous recordings of different signal sources, multi-modal analysis can improve detection accuracy by enhancing predictions of weaker modalities [135], [136]. Among reviewed papers, 24 studies (33.80%) recorded and analyzed data from one, 16 papers (22.54%) from two, 10 papers (14.08%) published between 2010 and 2016 from three, one 2016 paper from four [111], and one 2018 paper from five different physiological modalities [95], pointing to the rise in using multiple information sources to account for changes in intraoperative stress and workload³.

3) Synchronization: Robust multi-modal fusion is affected by challenges at the data, feature, and model design levels, including misaligned or non-synchronized streams [137], [138]. Synchronization of multiple sources such as physiological signals, motion data, speech, and operating scene videos with each other and with ongoing surgical phases is a critical process in data acquisition and processing [67]. Among collected studies, 16 papers (22.53%) clearly mentioned marking and annotating surgical phases and major time-locked events [6], [39], [62], [66], [77], [82], [86], [91], [96]–[99], [111], [121], [123], [125] and one study synchronized physiological data with an endoscopic video stream [97].

4) Baseline and Multi-Point Data Acquisition: Humans demonstrate large inter-subject variability in their sympathetic, parasympathetic, and cortical activations [5], [52], [62], [66], [77], [139], and it is beneficial to interpret intraoperative measurements with respect to individual baseline values to reliably identify the effects of stress and workload demands [106]. For example, surface EMG amplitudes can be normalized with respect to maximum amplitudes from voluntary contractions [89], [95]. Cardiac metrics, hormones, GSR values, and PFC activations are normalized with respect to pre-operative measurements (see, e.g., [6], [21], [66], [103]). In the current selection, 38 studies (53.52%) reported relative changes in their measurements with respect to baseline values from resting, sitting, or standing positions [4], [6], [10], [21], [24], [29], [40], [44], [62]–[66], [80], [83], [86], [89], [91], [92], [95]–[98], [107], [110], [111], [113], [114], [117], [123], [125], working during office days [77], [78], [116], performing cognitive and vigilance tasks [95], practicing a peg transfer task [61], or attending eye-tracker calibration sessions [38], [62], [63], [121]. Furthermore, 35 papers (49.30%) acquired objective or subjective metrics continuously or at multiple discrete events such as before and after a procedure or intervention to investigate temporal variations [10], [21], [24], [26], [39], [40], [44], [45], [61], [62], [66], [77], [78],

³One study collected data from five different physiological modalities but only reported its subjective metrics [127].

[80], [82]–[84], [88], [92], [96], [97], [100], [105], [107], [108], [110], [111], [114], [116], [117], [119], [122], [123], [125], [127]. Absolute values or relative changes were computed to assess the effects of procedures, phases, interventions, induced and simulated trauma, and behavioral or cognitive distractions.

E. Data Processing and Analysis Techniques

In this section, we review the techniques used for data pre-processing and analysis in the selected papers.

1) *Preprocessing Techniques*: Artifacts from biological sources, physical movements, electrode drifts, measurement errors caused by experimenters or environmental conditions (e.g. illumination changes), interference by tools and recording sensors affect underlying probabilistic distributions of recorded time series [140] and validity of computed features [141], [142]. Conducting artifact removal and signal correction is challenging in intraoperative settings where the highly mobile personnel are engaged in device operation and inter-personal interactions [32], [107]. Seven papers (9.86%) mentioned artifact removal and correction [6], [40], [62], [91], [95], [97], [125]. HRV was inspected for motion artifacts and arrhythmias [6], [91]. Artifacts from ectopic beats were removed from raw ECG and low-frequency trends were removed with a smoothness-priors algorithm [125]. EMG masseter tone and respiratory frequency were used to mark motion and speaking artifacts [97]. High-frequency noise, motion artifacts, and cardiac contamination were removed from haemoglobin data [40]. Low-pass filters were applied to GSR [95] and pupil dilation time series after blink removal [62].

2) *Data Analysis*: Sixty-six papers (92.96%) used at least one statistical analysis technique to report measured subjective or objective metrics or analyze effects of experimental sessions, surgical technologies, roles, and expertise on intraoperative performance and stress metrics. These techniques included descriptive statistics, hypothesis testing, correlation and effect sizes, parametric and non-parametric testing, single and multivariate analysis of variance (ANOVA), and Wilcoxon signed-rank and rank-sum tests for within- and between-group comparisons. Details are provided in the Supplementary Table.

Two studies applied uni- and multivariate general linear models (GLMs) [64], [80], and one used GLM-ANOVA and intra-cluster correlation for comparing different surgical procedures [92]. Hierarchical regression [107] and LOWESS linear regression [82] were other utilized models. Six papers published after 2016 applied linear mixed models (LMM), generalized linear models (GLM) and their hierarchical versions [92], generalized estimating equations (GEE) for repeated measurements and autoregressive correlational structure [21], and generalized linear mixed models (GLMM) with mixed, fixed, and random effects [44], [65], [82], [127]. Among the less frequent methods, the area under the curve (AUC) was calculated for cortisol integration [103], and sensitivity and specificity analysis were run with HR and cortisol levels as two predictors of STAI-based subjective psychological stress [102]. Common factor analysis [83], principal factors [104], and CUSUM sequential analysis [126] were run on subjective measures. From an ML perspective, only one study used support vector machine (SVM) after feature selection to predict surgical expertise levels [62].

3) *Open Data*: Among 71 reviewed studies, only one paper from 2019 included a publicly accessible anonymized dataset of

their entire participants [65]. Two studies [44], [64] published average changes in channel-wise cortical haemodynamics for different groups and four papers published administered questionnaires [83], [92], [119], [128]. One paper released a short video of conducted simulated surgeries [127]. Links of these additional data are provided in the Supplementary Table.

IV. DISCUSSION

Real-time monitoring of surgical personnel is an important component in smart ORs and an emerging topic in surgical data science [35], [36] that motivates new research on computational techniques for mental and surgical skills [143], cognition-guided surgery [144], and feedback systems [6]. Cognitive states such as stress and workload may negatively influence the quality and outcome of clinical care [145] or regulate positive cognitive flow [19]. The present study has systematically reviewed papers that assessed mental stress during real and simulated operations with a special attention to the temporal patterns and categorization of utilized subjective and objective measures. For the first time in the literature, a detailed systematic overview of experimental setups, wearable systems, uni- and multi-modal physiological data acquisition, feature and surgical events synchronization, multi-point and baseline recordings, preprocessing and artifact correction, data analysis methods, and open-access data is presented for intraoperative stress monitoring.

Our analysis shows that 60.56% of selected papers used both objective and subjective measures while 25.35% only administered subjective instruments. Validated instruments, notably NASA-TLX, SURG-TLX, and STAI, alone or combined with customized surveys and observations were used in 71.83% of selected papers to assess subjective stress, workload, task demands, etc. (Table III). Not only the number of studies on intraoperative mental stress assessment has increased since the mid-2010 s (Fig. 2), but also the administration of discrete subjective and continuous objective measurements has gained momentum (Fig. 3). This trend is accelerated by advances in wearable systems that enable unrestricted monitoring of physiological modalities and cognitive states. Concerns regarding limitations of subjective questionnaires [95] and the need for fine and real-time assessments that do not disrupt the operators' concentration and task flow support this growth [34], [52].

Cardiac activity is the most frequently used physiological modality for intraoperative stress assessment (Table IV), and has been acquired by wearable devices in 25 (31.21%) reviewed papers; this statistics far exceeds the use of wearables for other modalities. The systematic review in [17] categorizes HRV as a real-time assessment tool; however, since reductions in HRV caused by increased mental strain are detected after 3–5 minutes [6], [91], its use raises challenges for precise and real-time detection of stress and workload across different surgical phases. Same holds true for other measures of sympathetic and parasympathetic activities [146], [147] used in 45 studies (63.38%) (Table IV). For example, stress-related changes in plasma and cortisol are not synchronized with mood variations, and their correlations rarely exceed 0.2 due to different temporal system dynamics [146], [147]. Furthermore, results summarized in Column K of Supplementary Table should be interpreted with great care since a recent meta-analysis found a small effect size between ECG-computed HRV and outputs of portable devices and a significant error sensitive to temporal and spectral features

and recording positions [134]. Furthermore, sleep deprivation and caffeine consumption need to be carefully considered as confounding factors for cardiovascular and cortisol measures [148], [149].

Among exercised physiological modalities, the high temporal resolution of electrophysiological signals such as EEG, EMG, and EOG is highly desirable for real-time monitoring of physiological and cognitive processes and human factors in interactive environments. However, EEG was recorded only in two reviewed papers for stress assessment while more evidence exists about EEG correlates of stress in non-surgical contexts such as maritime training and construction [150]–[152]. Discrepancies between EEG-based stress studies such as lack of standardized protocols and variations in recorded brain regions, stressors, experiment duration, signal processing, feature extraction, and classification techniques need to be addressed by the community [153]. Furthermore, although fNIRS cortical hemodynamics have relatively slow responses around 1 Hz, their superior spatial resolution makes hybrid EEG-fNIRS solutions attractive for multi-modal recordings [51]. The introduction of lightweight and wearable hybrid systems listed in [154] has resolved earlier concerns about long setup time in surgical environments [155]. Likewise, slow task-evoked pupillary response with 0.5–1 s resolution, recorded by eye-tracking glasses or lenses attached to microsurgical equipment [143], can decode stress and workload without causing any discomfort for surgeons [62]. Recent progress in battery-operated wearables and introduction of commercial solutions for synchronization, annotation, and feature extraction allow researchers to conduct long and multi-modal data acquisition for intraoperative stress monitoring similar to experiments performed in clinical environments [156].

To the best of our knowledge, this is the first systematic review that addresses experimental design and data acquisition practices, including multi-modal data collection and synchronization, for intraoperative stress assessment. Integration of multiple behavioral, physiological, and audiovisual measurements has been encouraged in assessment of cognitive load [33], [40], intelligent decision-support and objective surgical skill assessment [35], and detecting the best time to deliver stress-reduction interventions [157]. Our analysis showed that 24 studies (33.80%) had collected one and a promising ratio of 28 papers (39.44%) collected two or more objective modalities. Multi-modal data collection necessitates the implementation of data fusion techniques to overcome different sampling rates, physiological properties, spatio-temporal resolutions, sample sizes, and misaligned or non-synchronized streams [138]. Only 16 reviewed papers (22.53%) reported using manual or automated methods to synchronize and annotate videos and physiological activities with surgical events. Surgical event annotation is a time consuming but essential task that has motivated several solutions to gather expert knowledge, annotate tools and actions, and improve reliability and model generalization [36], and should be widely exercised by the community. Finally, baseline data acquisition helps with interpreting observed effects considering inter-subject variability, and is even suggested by studies that did not incorporate baseline recordings [106]. A promising ratio of reviewed papers, 76.06%, performed baseline and multi-point collection of objective and subjective measures.

Investigation of data processing and analysis techniques constitutes another novel contribution of this systematic review. While the majority (92.96%) of selected papers used statistical

methods, only 14.08% applied advanced regression models, and only one reviewed study conducted feature selection and classification for stress assessment [62]. Applying modern ML techniques in this context demands further research since, while multi-modal datasets enable information fusion from different temporal and spatial resolutions for decision making, the inter-correlation of different modalities may decrease the accuracy of their predictions [158], [159].

The majority of physiological metrics in Table IV have been also used for assessing stress and workload in aviation and transportation (e.g., fNIRS [160], cardiovascular and hormone measures, EEG, EMG, respiration, and eye movements [161], [162]). However, lack of freedom for operating in poorly optimized conditions, different regulations for simulation training, and unequal opportunities for receiving feedback and supervision from surgical peers and mentors result in important distinctions between these two fields [163], [164].

Finally, findings reported in this systematic review may be affected by a few methodological limitations. Papers that were not indexed in databases or did not include designed search terms were, inevitably, not collected—resulting in possible exclusion of studies that measured stress or workload but did not clarify this in their abstracts. Furthermore, studies such as [165] were excluded due to not being accessible until February 2021 although it had assessed workload in simulated and robot-assisted laparoscopy with NASA-TLX and EEG-based PSD, used a wearable set with step-wise linear discriminant analysis (LDA), and included a resting-state baseline. Furthermore, despite incorporating a rigorous methodology to extract study characteristics, vagueness or incompleteness in experimental procedures and variables in source papers may have resulted in missing detailed features. Finally, study characteristics extracted using SPIDER were in line with defined research questions, and features such as electrodes type and locations, sampling frequency and precision, and duration of recordings were out of scope of this study but will be analyzed in a follow-up meta-analysis.

V. CONCLUSION

By reviewing studies on intraoperative assessment of mental stress published in the past 20 years and analyzing their temporal and categorical developments, we have identified gaps in experimental design and data analysis to suggest potential research directions considering the current trends and upcoming requirements for the emerging field of surgical data science. The current systematic review has spanned several trending applications including wearable systems, body area/sensor networks, and informatics in biological and physiological systems and healthcare technologies. This work can attract a broad audience that includes, but is not limited to, researchers in the fields of biomedical engineering, affective computing, surgical human factors, human-computer interaction, surgical training, and smart ORs. Our work stimulates further meta-analyses that will help to present a higher-level view on stress consequences in intraoperative work.

REFERENCES

- [1] M. Birhanu, B. Gebrekidan, G. Tesefa, and M. Tareke, "Workload determines workplace stress among health professionals working in felege-hiwot referral hospital, bahir dar, northwest Ethiopia," *J. Environ. public health*, vol. 2018, 2018.

- [2] A. Russo, A. Paštar, A. Slišković, I. Urlič, Z. Škare, and R. Mulić, "Stress among anesthesiology nurses/technicians working in the intensive care unit (ICU): Experiences and comparisons," *Psychiatria Danubina*, vol. 32, no. 1, pp. 47–52, 2020.
- [3] S. De Hert, "Burnout in healthcare workers: Prevalence, impact and preventative strategies," *Local Regional Anesth.*, vol. 13, pp. 171–183, 2020.
- [4] R. S. Prichard, C. J. O'Neill, J. J. Oucharek, Y. H. Colinda, L. W. Delbridge, and M. S. Sywak, "A prospective study of heart rate variability in endocrine surgery: Surgical training increases consultant's mental strain," *J. Surg. Educ.*, vol. 69, pp. 453–458, Jul. 2012.
- [5] T. Banfi *et al.*, "Effects of sleep deprivation on surgeons dexterity," *Front. Neurol.*, vol. 10, 2019, Art. no. 595.
- [6] M. Weenk *et al.*, "Stress measurement in surgeons and residents using a smart patch," *Am J. Surg.*, vol. 216, no. 2, pp. 361–368, 2018.
- [7] K. L. Chrouser, J. Xu, S. Hallbeck, M. B. Weinger, and M. R. Partin, "The influence of stress responses on surgical performance and outcomes: Literature review and the development of the surgical stress effects (SSE) framework," *Am J. Surg.*, vol. 216, no. 3, pp. 573–584, 2018.
- [8] H. N. Modi, H. Singh, A. Darzi, and D. R. Leff, "Multitasking and time pressure in the operating room: Impact on surgeons' brain function," *Ann. Surg.*, vol. 272, no. 4, pp. 648–657, 2020.
- [9] T. J. Kim and P. H. Seong, "Influencing factors on situation assessment of human operators in unexpected plant conditions," *Ann. Nucl. Energy*, vol. 132, pp. 526–536, 2019.
- [10] N. E. Anton, L. D. Howley, M. Pimentel, C. K. Davis, C. Brown, and D. Stefanidis, "Effectiveness of a mental skills curriculum to reduce novices' stress," *J. Surg. Res.*, vol. 206, no. 1, pp. 199–205, 2016.
- [11] R. S. Lazarus and S. Folkman, *Stress, Appraisal, and Coping*. New York, NY, USA: Springer, 1984.
- [12] R. S. Lazarus, *Stress and Emotion: A New Synthesis*. New York, NY, USA: Springer, 2006.
- [13] J. M. Koolhaas *et al.*, "Stress revisited: A critical evaluation of the stress concept," *Neurosci. Biobehav. Rev.*, vol. 35, no. 5, pp. 1291–1301, 2011.
- [14] P. M. Goodnite, "Stress: A concept analysis," *Nursing Forum*, Wiley Online Library, vol. 49, no. 1, pp. 71–74, 2014.
- [15] J. L. Plass and S. Kalyuga, "Four ways of considering emotion in cognitive load theory," *Educ. Psychol. Rev.*, vol. 31, no. 2, pp. 339–359, 2019.
- [16] A. Gaillard, "Comparing the concepts of mental load and stress," *Ergonomics*, vol. 36, no. 9, pp. 991–1005, 1993.
- [17] R. D. Dias, M. C. Ngo-Howard, M. T. Boskovski, M. A. Zenati, and S. J. Yule, "Systematic review of measurement tools to assess surgeons' intra-operative cognitive workload," *Br J. Surg.*, vol. 105, no. 5, pp. 491–501, 2018.
- [18] G. Matthews, J. Szalma, A. R. Panganiban, C. Neubauer, and J. S. Warm, "Profiling task stress with the dundee stress state questionnaire," *Psychol. stress: New research*, vol. 1, pp. 49–91, 2013.
- [19] S. McQueen, M. H. Mobilio, and C.-A. Moulton, "Fractured in surgery: Understanding stress as a holistic and subjective surgeon experience," *Amer. J. Surg.*, vol. 221, no. 4, pp. 793–798, 2021.
- [20] N. E. Anton, P. N. Montero, L. D. Howley, C. Brown, and D. Stefanidis, "What stress coping strategies are surgeons relying upon during surgery?," *Amer. J. Surg.*, vol. 210, pp. 846–851, Nov. 2015.
- [21] B. T. Crewther *et al.*, "Skill acquisition and stress adaptations following laparoscopic surgery training and detraining in novice surgeons," *Surg. Endosc.*, vol. 30, no. 7, pp. 2961–2968, 2016.
- [22] E. Belykh *et al.*, "Systematic review of factors influencing surgical performance: Practical recommendations for microsurgical procedures in neurosurgery," *World Neurosurgery*, vol. 112, pp. e182–e207, 2018.
- [23] M. Roizenblatt *et al.*, "Association of weight-adjusted caffeine and β -blocker use with ophthalmology fellow performance during simulated vitreoretinal microsurgery," *JAMA Ophthalmol.*, vol. 138, no. 8, pp. 819–825, 2020.
- [24] S. Arora, N. Sevdalis, R. Aggarwal, P. Sirimanna, A. Darzi, and R. Kneebone, "Stress impairs psychomotor performance in novice laparoscopic surgeons," *Surg. Endosc.*, vol. 24, no. 10, pp. 2588–2593, 2010.
- [25] P. L. Hénaux, P. Jannin, and L. Riffaud, "Nontechnical skills in neurosurgery: A systematic review of the literature," *World Neurosurg.*, vol. 130, pp. e726–e736, 2019.
- [26] G. I. Lee and M. R. Lee, "Can a virtual reality surgical simulation training provide a self-driven and mentor-free skills learning? investigation of the practical influence of the performance metrics from the virtual reality robotic surgery simulator on the skill learning and asso," *Surg. Endosc.*, vol. 32, no. 1, pp. 62–72, 2018.
- [27] J. Barré *et al.*, "Virtual reality single-port sleeve gastrectomy training decreases physical and mental workload in novice surgeons: An exploratory study," *Obes. Surg.*, vol. 29, no. 4, pp. 1309–1316, 2019.
- [28] M. C. Stetz, J. Y. Kaloi-Chen, D. D. Turner, S. Bouchard, G. Riva, and B. K. Wiederhold, "The effectiveness of technology-enhanced relaxation techniques for military medical warriors," *Mil. Med.*, vol. 176, no. 9, pp. 1065–1070, 2011.
- [29] D. Stefanidis *et al.*, "Implementation results of a novel comprehensive mental skills curriculum during simulator training," *Amer. J. Surg.*, vol. 213, no. 2, pp. 353–361, 2017.
- [30] S. Arora, N. Sevdalis, D. Nestel, M. Woloshynowych, A. Darzi, and R. Kneebone, "The impact of stress on surgical performance: A systematic review of the literature," *Surgery*, vol. 147, no. 3, pp. 318–330, 2010.
- [31] D. A. Reyes, B. Tang, and A. Cuschieri, "Minimal access surgery (MAS)-related surgeon morbidity syndromes," *Surg. Endosc. Other Interv. Tech.*, vol. 20, no. 1, pp. 1–13, 2006.
- [32] K. Georgiou, A. Larentzakis, and A. G. Papavassiliou, "Surgeons' and surgical trainees' acute stress in real operations or simulation: A systematic review," *Surgeon*, vol. 15, no. 6, pp. 355–365, 2017.
- [33] L. M. Naismith and R. B. Cavalcanti, "Validity of cognitive load measures in simulation-based training: A systematic review," *Acad. Med. Associat. Amer. Med. Colleges Med. Educ. Meeting*, vol. 90, no. 11, pp. S24–S35, 2015.
- [34] J. L. Sewell, L. A. Maggio, O. ten Cate, T. van Gog, J. Q. Young, and P. S. O'Sullivan, "Cognitive load theory for training health professionals in the workplace: A BEME review of studies among diverse professions: BEME guide no. 53," *Med. Teach.*, vol. 41, no. 3, pp. 256–270, 2019.
- [35] S. S. Vedula and G. D. Hager, "Surgical data science: The new knowledge domain," *Innov. Surg. Sci.*, vol. 2, no. 3, pp. 109–121, 2020.
- [36] L. Maier-Hein *et al.*, "Surgical data science—from concepts toward clinical translation," *Med. image Anal.*, vol. 76, p. 102306, 2022.
- [37] A. Sami *et al.*, "Real-time observations of stressful events in the operating room," *Saudi J. Anaesth.*, vol. 6, no. 2, p. 136, 2012.
- [38] L. J. Moore, M. R. Wilson, J. S. McGrath, E. Waine, R. S. Masters, and S. J. Vine, "Surgeons' display reduced mental effort and workload while performing robotically assisted surgical tasks, when compared to conventional laparoscopy," *Surg. Endosc.*, vol. 29, no. 9, pp. 2553–2560, 2015.
- [39] A. M. Abdelrahman *et al.*, "Impact of single-incision laparoscopic cholecystectomy (SILC) versus conventional laparoscopic cholecystectomy (CLC) procedures on surgeon stress and workload: A randomized controlled trial," *Surg. Endosc.*, vol. 30, no. 3, pp. 1205–1211, 2016.
- [40] H. Singh *et al.*, "Robotic surgery improves technical performance and enhances prefrontal activation during high temporal demand," *Ann. Biomed. Eng.*, vol. 46, no. 10, pp. 1621–1636, 2018.
- [41] M. A. Zenati, K. B. Leissner, S. Zorca, L. Kennedy-Metz, S. J. Yule, and R. D. Dias, "First reported use of team cognitive workload for root cause analysis in cardiac surgery," in *Seminars in Thoracic and Cardiovascular Surgery*, Elsevier, 2019, vol. 31, pp. 394–396.
- [42] L. R. Kennedy-Metz, R. D. Dias, R. Srey, G. C. Rance, C. Furlanello, and M. A. Zenati, "Sensors for continuous monitoring of surgeon's cognitive workload in the cardiac operating room," *Sensors (Switzerland)*, vol. 20, no. 22, pp. 1–11, 2020.
- [43] M. Marrelli, S. Gentile, F. Palmieri, F. Paduano, and M. Tatullo, "Correlation between surgeon's experience, surgery complexity and the alteration of stress related physiological parameters," *PLoS One*, vol. 9, Nov. 2014, Art. no. e112444.
- [44] H. N. Modi *et al.*, "Temporal stress in the operating room: Brain engagement promotes "coping" and disengagement prompts "choking," *Ann. Surg.*, vol. 267, no. 4, pp. 683–691, 2018.
- [45] D. Stefanidis *et al.*, "Effectiveness of a comprehensive mental skills curriculum in enhancing surgical performance: Results of a randomized controlled trial," *Amer. J. Surg.*, vol. 213, no. 2, pp. 318–324, 2017.
- [46] L. X. Chen, T. Fuller, D. D. McIntire, and K. A. Kho, "Introduction to open surgical skills curriculum: Randomized trial of self-paced vs group video tutorial viewing," *J. Surg. Educ.*, vol. 76, no. 2, pp. 453–458, 2019.
- [47] N. E. Anton, E. A. Bean, E. Myers, and D. Stefanidis, "Optimizing learner engagement during mental skills training: A pilot study of small group vs. individualized training," *Amer. J. Surg.*, vol. 219, no. 2, pp. 335–339, 2020.
- [48] L. Kennedy and S. H. Parker, "Timing of coping instruction presentation for real-time acute stress management: Potential implications for improved surgical performance," *J. Healthc. Informat. Res.*, vol. 2, no. 1–2, pp. 111–131, 2018.

- [49] Y. Kassahun *et al.*, "Surgical robotics beyond enhanced dexterity instrumentation: A survey of machine learning techniques and their role in intelligent and autonomous surgical actions," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 4, pp. 553–568, 2016.
- [50] D. R. Seshadri *et al.*, "Wearable sensors for monitoring the internal and external workload of the athlete," *NPJ Digit. Med.*, vol. 2, no. 1, pp. 1–18, 2019.
- [51] F. Al-Shargie, T. B. Tang, and M. Kiguchi, "Stress assessment based on decision fusion of EEG and fNIRS signals," *IEEE Access*, vol. 5, pp. 19889–19896, 2017.
- [52] M. Torkamani-Azar, S. D. Kanik, S. Aydin, and M. Cetin, "Prediction of reaction time and vigilance variability from spatio-spectral features of resting-state EEG in a long sustained attention task," *IEEE J. Biomed. Heal. Informat.*, vol. 24, no. 9, pp. 2550–2558, Sep. 2020.
- [53] S. L. Klosterman, J. R. Estep, J. W. Monnin, and J. C. Christensen, "Day-to-day variability in hybrid, passive brain-computer interfaces: Comparing two studies assessing cognitive workload," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2016, pp. 1584–1590.
- [54] G. Lecuyer, M. Ragot, N. Martin, L. Launay, and P. Jannin, "Assisted phase and step annotation for surgical videos," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 15, no. 4, pp. 673–680, 2020.
- [55] J. V. Cordeiro, "Digital technologies and data science as health enablers: An outline of appealing promises and compelling ethical, legal, and social challenges," *Front. Med.*, vol. 8, 2021, Art. no. 647897.
- [56] L. Maier-Hein *et al.*, "Surgical data science for next-generation interventions," *Nat. Biomed. Eng.*, vol. 1, no. 9, pp. 691–696, 2017.
- [57] K. Tuwairqi, J. H. Selter, and S. Sikder, "Assessment of surgeon fatigue by surgical simulators," *Open Access Surg.*, vol. 8, pp. 43–50, 2015.
- [58] H. M. Mentis, A. Chellali, K. Manser, C. G. Cao, and S. D. Schwaitzberg, "A systematic review of the effect of distraction on surgeon performance: Directions for operating room policy and surgical training," *Surg. Endosc.*, vol. 30, no. 5, pp. 1713–1724, 2016.
- [59] P. Oomens, V. X. Fu, G. J. Kleinrensink, and J. Jeekel, "The effect of music on simulated surgical performance: A systematic review," *Surg. Endosc.*, vol. 33, no. 9, pp. 2774–2784, 2019.
- [60] L. Hull, S. Arora, R. Aggarwal, A. Darzi, C. Vincent, and N. Sevdalis, "The impact of nontechnical skills on technical performance in surgery: A systematic review," *J. Amer. Coll. Surg.*, vol. 214, no. 2, pp. 214–230, 2012.
- [61] V. X. Fu *et al.*, "The effect of preferred music on mental workload and laparoscopic surgical performance in a simulated setting (OPTIMISE): A randomized controlled crossover study," *Surg. Endosc.*, 2020.
- [62] J. Koskinen, R. Bednarik, H. Vrzakova, and A. P. Elomaa, "Combined gaze metrics as stress-sensitive indicators of microsurgical proficiency," *Surg. Innov.*, vol. 27, no. 6, pp. 614–622, 2020.
- [63] J. Gao *et al.*, "Quantitative evaluations of the effects of noise on mental workloads based on pupil dilation during laparoscopic surgery," *Amer. Surg.*, vol. 84, pp. 1951–1956, Dec. 2018.
- [64] H. N. Modi *et al.*, "Association of residents' neural signatures with stress resilience during surgery," *JAMA Surg.*, vol. 154, no. 10, 2019.
- [65] I. Pavlidis, D. Zavlin, A. R. Khatri, A. Wesley, G. Panagopoulos, and A. Echo, "Absence of stressful conditions accelerates dexterous skill acquisition in surgery," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, 2019.
- [66] R. Phitayakorn, R. D. Minehart, M. W. Hemingway, M. C. Pian-Smith, and E. Petrusa, "Relationship between physiologic and psychological measures of autonomic activation in operating room teams during a simulated airway emergency," *Amer. J. Surg.*, vol. 209, no. 1, pp. 86–92, 2015.
- [67] G. Pimentel, S. Rodrigues, P. A. Silva, A. Vilarinho, R. Vaz, and J. P. Silva Cunha, "A wearable approach for intraoperative physiological stress monitoring of multiple cooperative surgeons," *Int. J. Med. Inform.*, vol. 129, pp. 60–68, Sep. 2019.
- [68] H. Snyder, "Literature review as a research methodology: An overview and guidelines," *J. Bus. Res.*, vol. 104, pp. 333–339, 2019.
- [69] A. Liberati *et al.*, "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration," *PLoS Med.*, vol. 6, no. 7, pp. e1–e34, 2009.
- [70] M. L. Rethlefsen *et al.*, "PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews," *Syst. Rev.*, vol. 10, Jan. 2021, Art. no. 39.
- [71] M. H. Zubair, L. R. Hussain, K. N. Williams, and K. J. Grannan, "Work-related quality of life of US general surgery residents: Is it really so bad?," *J. Surg. Educ.*, vol. 74, no. 6, pp. e138–e146, 2017.
- [72] R. S. Chung and N. Ahmed, "How surgical residents spend their training time: The effect of a goal-oriented work style on efficiency and work satisfaction," *Arch. Surg.*, vol. 142, no. 3, pp. 249–252, 2007.
- [73] F. Çalişkan Tür, I. Toker, C. T. Şaşmaz, S. Hacı, and B. Türe, "Occupational stress experienced by residents and faculty physicians on night shifts," *Scand. J. Trauma. Resusc. Emerg. Med.*, vol. 24, no. 1, pp. 1–8, 2016.
- [74] C. C. Lebares, E. V. Guvva, N. L. Ascher, P. S. O'Sullivan, H. W. Harris, and E. S. Epel, "Burnout and stress among US surgery residents: Psychological distress and resilience," *J. Am. Coll. Surg.*, vol. 226, no. 1, pp. 80–90, 2018.
- [75] D. Harris, M. Wilson, and S. Vine, "Development and validation of a simulation workload measure: The simulation task load index (SIM-TLX)," *Virtual Real.*, vol. 24, no. 4, pp. 557–566, 2020.
- [76] G. Jones, M. Hocine, J. Salomon, W. Dab, and L. Temime, "Demographic and occupational predictors of stress and fatigue in french intensive-care registered nurses and nurses' aides: A cross-sectional study," *Int. J. Nurs. Stud.*, vol. 52, no. 1, pp. 250–259, 2015.
- [77] Y. Demirtas *et al.*, "Plastic surgeon's life: Marvelous for mind, exhausting for body," *Plast. Reconstr. Surg.*, vol. 114, no. 4, pp. 923–931, 2004.
- [78] A. P. Tendulkar, G. P. Victorino, T. J. Chong, M. K. Bullard, T. H. Liu, and A. H. Harken, "Quantification of surgical resident stress "on call," *J. Amer. Coll. Surg.*, vol. 201, no. 4, pp. 560–564, 2005.
- [79] I. Hassan *et al.*, "Negative stress-coping strategies among novices in surgery correlate with poor virtual laparoscopic performance," *J. Brit. Surg.*, vol. 93, no. 12, pp. 1554–1559, 2006.
- [80] C. M. Wetzel *et al.*, "The effects of stress and coping on surgical performance during simulations," *Ann. Surg.*, vol. 251, no. 1, pp. 171–176, 2010.
- [81] A. Cooke, D. Smith, and A. Booth, "Beyond PICO: The SPIDER tool for qualitative evidence synthesis," *Qual. Health Res.*, vol. 22, no. 10, pp. 1435–1443, 2012.
- [82] K. Ukegijini, T. Kastiunig, B. Widmann, R. Warschkow, and T. Steffen, "Impact of intraoperative noise measurement on the surgeon stress and patient outcomes. a prospective, controlled, single-center clinical trial with 664 patients," *Surg.*, vol. 167, no. 5, pp. 843–851, 2020.
- [83] J. B. Lemaire, J. E. Wallace, A. M. Lewin, J. de Grood, and J. P. Schaefer, "The effect of a biofeedback-based stress management tool on physician stress: A randomized controlled clinical trial," *Open Med.*, vol. 5, no. 4, pp. 154–165, 2011.
- [84] I. Amirian, L. T. Andersen, J. Rosenberg, and I. Gögenur, "Decreased heart rate variability in surgeons during night shifts," *Can. J. Surg.*, vol. 57, no. 5, pp. 300–304, 2014.
- [85] U. Yildiz Findik, A. Ozbas, I. Cavdar, S. Yildizeli Topcu, and E. Onler, "Assessment of nursing students' stress levels and coping strategies in operating room practice," *Nurse Educ. Pract.*, vol. 15, no. 3, pp. 192–195, 2015.
- [86] B. Böhm, N. Rötting, W. Schwenk, S. Grebe, and U. Mansmann, "A prospective randomized trial on heart rate variability of the surgical team during laparoscopic and conventional sigmoid resection," *Arch. Surg.*, vol. 136, no. 3, pp. 305–310, 2001.
- [87] M. I. Klein *et al.*, "Performance, stress, workload, and coping profiles in 1st year medical students' interaction with the endoscopic/laparoscopic and robot-assisted surgical techniques," *Proc. Hum. Factors Ergon. Soc.*, vol. 2, pp. 885–889, 2008.
- [88] M. I. Klein *et al.*, "Mental workload and stress perceived by novice operators in the laparoscopic and robotic minimally invasive surgical interfaces," *J. Endourol.*, vol. 26, no. 8, pp. 1089–1094, 2012.
- [89] N. Hubert, M. Gilles, K. Desbrosses, J. Meyer, J. Felblinger, and J. Hubert, "Ergonomic assessment of the surgeon's physical workload during standard and robotic assisted laparoscopic procedures," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 9, no. 2, pp. 142–147, 2013.
- [90] M. I. Klein *et al.*, "Mental stress experienced by first-year residents and expert surgeons with robotic and laparoscopic surgery interfaces," *J. Robot. Surg.*, vol. 8, no. 2, pp. 149–155, 2014.
- [91] J. Heemskerk *et al.*, "Relax, it's just laparoscopy! a prospective randomized trial on heart rate variability of the surgeon in robot-assisted versus conventional laparoscopic cholecystectomy," *Dig. Surg.*, vol. 31, no. 3, pp. 225–232, 2014.
- [92] B. R. Lowndes *et al.*, "Surgical team workload comparison for 4-port and single-port laparoscopic cholecystectomy procedures," *Appl. Ergon.*, vol. 78, pp. 277–285, 2019.
- [93] M. I. Klein, P. R. DeLucia, and R. Olmstead, "The impact of visual scanning in the laparoscopic environment after engaging in strain coping," *Hum. Factors*, vol. 55, no. 3, pp. 509–519, 2013.
- [94] E. C. Lee, A. Rafiq, R. Merrell, R. Ackerman, and J. T. Dennerlein, "Ergonomics and human factors in endoscopic surgery: A comparison of manual vs telerobotic simulation systems," *Surg. Endosc. Other Interv. Tech.*, vol. 19, no. 8, pp. 1064–1070, 2005.

- [95] Z. Wang, I. Reed, and A. M. Fey, "Toward intuitive teleoperation in surgery: Human-centric evaluation of teleoperation algorithms for robotic needle steering," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2018, pp. 5799–5806.
- [96] M. N. Theodoraki *et al.*, "Mental distress and effort to engage an image-guided navigation system in the surgical training of endoscopic sinus surgery: A prospective, randomised clinical trial," *Eur. Arch. Oto-Rhino-Laryngol.*, vol. 272, pp. 905–913, Jul. 2015.
- [97] K. Stelter, M. N. Theodoraki, S. Becker, V. Tsekmistrenko, B. Olzowy, and G. Ledderose, "Specific stressors in endonasal skull base surgery with and without navigation," *Eur. Arch. Oto-Rhino-Laryngol.*, vol. 272, no. 3, pp. 631–638, 2015.
- [98] R. Berguer, W. D. Smith, and Y. H. Chung, "Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery," *Surg. Endosc.*, vol. 15, no. 10, pp. 1204–1207, 2001.
- [99] M. Schuetz *et al.*, "Three different types of surgeon-specific stress reactions identified by laparoscopic simulation in a virtual scenario," *Surg. Endosc. Other Interv. Tech.*, vol. 22, no. 5, pp. 1263–1267, 2008.
- [100] M. Klein, L. P. H. Andersen, M. Alamili, I. Gögenur, and J. Rosenberg, "Psychological and physical stress in surgeons operating in a standard or modern operating room," *Surg. Laparosc. Endosc. Percutaneous Tech.*, vol. 20, no. 4, pp. 237–242, 2010.
- [101] P. B. Andreatta, M. Hillard, and L. P. Krain, "The impact of stress factors in simulation-based laparoscopic training," *Surgery*, vol. 147, no. 5, pp. 631–639, 2010.
- [102] S. Arora *et al.*, "The imperial stress assessment tool (ISAT): A feasible, reliable and valid approach to measuring stress in the operating room," *World J. Surg.*, vol. 34, no. 8, pp. 1756–1763, 2010.
- [103] C. Engelmann *et al.*, "Effects of intraoperative breaks on mental and somatic operator fatigue: A randomized clinical trial," *Surg. Endosc.*, vol. 25, no. 4, pp. 1245–1250, 2011.
- [104] C. Conrad *et al.*, "A quality improvement study on avoidable stressors and countermeasures affecting surgical motor performance and learning," *Ann. Surg.*, vol. 255, no. 6, pp. 1190–1194, 2012.
- [105] L. P. H. Andersen, M. Klein, I. Gögenur, and J. Rosenberg, "Psychological and physical stress among experienced and inexperienced surgeons during laparoscopic cholecystectomy," *Surg. Laparosc. Endosc. Percutaneous Tech.*, vol. 22, no. 1, pp. 73–78, 2012.
- [106] B. Zheng, X. Jiang, G. Tien, A. Meneghetti, O. N. M. Panton, and M. S. Atkins, "Workload assessment of surgeons: Correlation between NASA TLX and blinks," *Surg. Endosc.*, vol. 26, no. 10, pp. 2746–2750, 2012.
- [107] S. J. Vine, P. Freeman, L. J. Moore, R. Chandra-Ramanan, and M. R. Wilson, "Evaluating stress as a challenge is associated with superior attentional control and motor skill performance: Testing the predictions of the biopsychosocial model of challenge and threat," *J. Exp. Psychol. Appl.*, vol. 19, no. 3, pp. 185–194, 2013.
- [108] J. R. Plyuter, A. F. Rutkowski, and J. J. Jakimowicz, "Immersive training: Breaking the bubble and measuring the heat," *Surg. Endosc.*, vol. 28, no. 5, pp. 1545–1554, 2014.
- [109] R. J. Berg *et al.*, "The impact of heat stress on operative performance and cognitive function during simulated laparoscopic operative tasks," *Surg.*, vol. 157, pp. 87–95, Jan. 2015.
- [110] P. Waterland, F. S. Khan, E. Ismaili, and C. Cheruvu, "Environmental noise as an operative stressor during simulated laparoscopic surgery," *Surg. Laparosc. Endosc. Percutaneous Tech.*, vol. 26, no. 2, pp. 133–136, 2016.
- [111] J. T. Flinn, A. Miller, N. Pyatka, J. Brewer, T. Schneider, and C. G. Cao, "The effect of stress on learning in surgical skill acquisition," *Med. Teach.*, vol. 38, no. 9, pp. 897–903, 2016.
- [112] Y. Sonoda, D. Onozuka, and A. Hagihara, "Factors related to teamwork performance and stress of operating room nurses," *J. Nurs. Manag.*, vol. 26, no. 1, pp. 66–73, 2018.
- [113] P. Oomens, V. X. Fu, V. E. Kleinrensink, G. J. Kleinrensink, and J. Jeekel, "The effects of preferred music on laparoscopic surgical performance: A randomized crossover study," *World J. Surg.*, vol. 44, no. 8, pp. 2614–2619, 2020.
- [114] A. Harvey, G. Bandiera, A. B. Nathens, and V. R. LeBlanc, "Impact of stress on resident performance in simulated trauma scenarios," *J. Trauma Acute Care Surg.*, vol. 72, no. 2, pp. 497–503, 2012.
- [115] J. Causer, J. N. Vickers, R. Snelgrove, G. Arsenaault, and A. Harvey, "Performing under pressure: Quiet eye training improves surgical knotting performance," *Surg.*, vol. 156, no. 5, pp. 1089–1096, 2014.
- [116] A. Gouin *et al.*, "Evolution of stress in anaesthesia registrars with repeated simulated courses: An observational study," *Anaesth. Crit. Care Pain Med.*, vol. 36, no. 1, pp. 21–26, 2017.
- [117] M. B. Goldberg *et al.*, "Optimizing performance through stress training - an educational strategy for surgical residents," *Amer. J. Surg.*, vol. 216, no. 3, pp. 618–623, 2018.
- [118] A. Bakhsh, G. F. Martin, C. D. Bicknell, C. Pettengell, and C. Riga, "An evaluation of the impact of high-fidelity endovascular simulation on surgeon stress and technical performance," *J. Surg. Educ.*, vol. 76, no. 3, pp. 864–871, 2019.
- [119] T. Abe *et al.*, "The effect of repeated full immersion simulation training in ureterorenoscopy on mental workload of novice operators," *BMC Med. Educ.*, vol. 19, no. 1, pp. 1–9, 2019.
- [120] A. Kurmann, M. Peter, F. Tschan, K. Mühlemann, D. Candinas, and G. Beldi, "Adverse effect of noise in the operating theatre on surgical-site infection," *Brit. J. Surg.*, vol. 98, no. 7, pp. 1021–1025, 2011.
- [121] T. Tien, P. H. Pucher, M. H. Sodergren, K. Sriskandarajah, G. Z. Yang, and A. Darzi, "Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair," *Surg. Endosc.*, vol. 29, no. 2, pp. 405–413, 2015.
- [122] A. Rieger, S. Fenger, S. Neubert, M. Weippert, S. Kreuzfeld, and R. Stoll, "Psychophysical workload in the operating room: Primary surgeon versus assistant," *Surg. Endosc.*, vol. 29, no. 7, pp. 1990–1998, 2015.
- [123] A. Bunevicius, D. Bilskiene, A. Macas, and A. Tamasauskas, "Hemodynamic response during aneurysm clipping surgery among experienced neurosurgeons," *Acta Neurochir. (Wien)*, vol. 158, no. 2, pp. 221–227, 2016.
- [124] B. Joseph *et al.*, "Stress among surgical attending physicians and trainees: A quantitative assessment during trauma activation and emergency surgeries," *J. Trauma Acute Care Surg.*, vol. 81, no. 4, pp. 723–728, 2016.
- [125] M. M. Dedmon *et al.*, "Measuring mental stress during otologic surgery using heart rate variability analysis," *Otol. Neurotol.*, vol. 40, no. 4, pp. 529–534, 2019.
- [126] B. Kayani, S. Konan, S. S. Huq, J. Tahmassebi, and F. S. Haddad, "Robotic-arm assisted total knee arthroplasty has a learning curve of seven cases for integration into the surgical workflow but no learning curve effect for accuracy of implant positioning," *Knee Surgery, Sport. Traumatol. Arthrosc.*, vol. 27, no. 4, pp. 1132–1141, 2019.
- [127] K. Bajunaid *et al.*, "Impact of acute stress on psychomotor bimanual performance during a simulated tumor resection task," *J. Neurosurg.*, vol. 126, no. 1, pp. 71–80, 2017.
- [128] Á. del Blanco, J. Torrente, B. Fernández-Manjón, P. Ruiz, and M. Giner, "Using a videogame to facilitate nursing and medical students' first visit to the operating theatre. a randomized controlled trial," *Nurse Educ. Today*, vol. 55, pp. 45–53, 2017.
- [129] S. G. Hart, "NASA-task load index (NASA-TLX): 20 years later," in *Proc. Hum. Factors Ergon. Soc.*, 2006, pp. 904–908.
- [130] M. R. Wilson, J. M. Poolton, N. Malhotra, K. Ngo, E. Bright, and R. S. Masters, "Development and validation of a surgical workload measure: The surgery task load index (SURG-TLX)," *World J. Surg.*, vol. 35, no. 9, pp. 1961–1969, 2011.
- [131] A. Tluczek, J. B. Henriques, and R. L. Brown, "Support for the reliability and validity of a six-item state anxiety scale derived from the state-trait anxiety inventory," *J. Nurs. Meas.*, vol. 17, no. 1, pp. 19–28, 2009.
- [132] M. B. Weinger and N. T. Smith, "Vigilance, alarms, and integrated monitoring systems," *Anesth. Equip. Princ. Appl. Malvern, PA Mosby Year Book*, no. 125, pp. 350–384, 1993.
- [133] A. F. The, I. Reijmerink, M. van der Laan, and F. Cnossen, "Heart rate variability as a measure of mental stress in surgery: A systematic review," *Int. Arch. Occup. Environ. Health*, vol. 93, no. 7, pp. 805–821, 2020.
- [134] W. C. Dobbs *et al.*, "The accuracy of acquiring heart rate variability from portable devices: A systematic review and meta-analysis," *Sport. Med.*, vol. 49, no. 3, pp. 417–435, 2019.
- [135] A. Kominos, "Challenges in multimodal notifications for monitoring cyberphysical systems," in *Proc. 1st Int. Conf. Soc. Autom.*, 2019, pp. 1–7.
- [136] S. Wang *et al.*, "Advances in data preprocessing for bio-medical data fusion: An overview of the methods, challenges, and prospects," *Inf. Fusion*, vol. 76, pp. 376–421, 2021.
- [137] I. Domingues, H. Muller, A. Ortiz, B. V. Dasarathy, P. H. Abreu, and V. D. Calhoun, "Guest editorial: Information fusion for medical data: Early, late, and deep fusion methods for multimodal data," *IEEE J. Biomed. Heal. Informat.*, vol. 24, no. 1, pp. 14–16, Jan. 2020.
- [138] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges, and prospects," in *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, Sep. 2015.
- [139] S. H. Fairclough and L. Venables, "Prediction of subjective states from psychophysiology: A multivariate approach," *Biol. Psychol.*, vol. 71, no. 1, pp. 100–110, 2006.

- [140] N. Gold, M. G. Frasca, C. L. Herry, B. S. Richardson, and X. Wang, "A doubly stochastic change point detection algorithm for noisy biological signals," *Front. Physiol.*, vol. 8, no. JAN, pp. 1–16, 2018.
- [141] A. Jafarifarmand and M. A. Badamchizadeh, "EEG artifacts handling in a real practical brain-computer interface controlled vehicle," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1200–1208, Jun. 2019.
- [142] T. Zhang et al., "Physiological measurements of situation awareness: A systematic review," *Hum. Factors*, no. 3, pp. 1–22, Nov. 2020.
- [143] R. Bednarik et al., "Pupil size as an indicator of visual-motor workload and expertise in microsurgical training tasks," in *Proc. ACM Symp. Eye Track. Res. Appl. Symp.*, 2018, pp. 1–5.
- [144] R. D. Dias, S. J. Yule, and M. A. Zenati, "Augmented cognition in the operating room," in *Digital Surgery*, Cham, Switzerland: Springer, 2020, pp. 261–268.
- [145] J. W. Kwon, S. Sung, S. B. Lee, H. M. Lee, S. H. Moon, and B. H. Lee, "Intraoperative real-time stress in degenerative lumbar spine surgery: Simultaneous analysis of electroencephalography signals and heart rate variability: A pilot study," *Spine J.*, vol. 20, no. 8, pp. 1203–1210, 2020.
- [146] L. J. Weckesser, F. Dietz, K. Schmidt, J. Grass, C. Kirschbaum, and R. Miller, "The psychometric properties and temporal dynamics of subjective stress, retrospectively assessed by different informants and questionnaires, and hair cortisol concentrations," *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, 2019.
- [147] W. Schlotz, R. Kumsta, I. Layes, S. Entringer, A. Jones, and S. Wüst, "Covariance between psychological and endocrine responses to pharmacological challenge and psychosocial stress: A question of timing," *Psychosom. Med.*, vol. 70, no. 7, pp. 787–796, 2008.
- [148] W. R. Lovallo, N. H. Farag, A. S. Vincent, T. L. Thomas, and M. F. Wilson, "Cortisol responses to mental stress, exercise, and meals following caffeine intake in men and women," *Pharmacol Biochem Behav.*, vol. 83, no. 3, pp. 441–447, 2006.
- [149] A. I. Almojali, S. A. Almalki, A. S. Alothman, E. M. Masuadi, and M. K. Alaqeel, "The prevalence and association of stress with sleep quality among medical students," *J. Epidemiol. Glob. Health.*, vol. 7, no. 3, pp. 169–174, 2017.
- [150] W. L. Lim et al., "EEG-based mental workload and stress monitoring of crew members in maritime virtual simulator," in *Trans. Comput. Sci. XXXII*, Berlin, Germany: Springer, 2018, pp. 15–28.
- [151] H. Jebelli, M. Mahdi Khalili, and S. Lee, "A continuously updated, computationally efficient stress recognition framework using electroencephalogram (EEG) by applying online multitask learning algorithms (OMTL)," *IEEE J. Biomed. Heal. Informat.*, vol. 23, no. 5, pp. 1928–1939, Sep. 2019.
- [152] A. Asif, M. Majid, and S. M. Anwar, "Human stress classification using EEG signals in response to music tracks," *Comput. Biol. Med.*, vol. 107, pp. 182–196, 2019.
- [153] R. Katmah, F. Al-Shargie, U. Tariq, F. Babiloni, F. Al-Mughairbi, and H. Al-Nashash, "A review on mental stress assessment methods using EEG signals," *Sensors*, vol. 21, no. 15, 2021, Art. no. 5043.
- [154] C. Cannard, T. Brandmeyer, H. Wahbeh, and A. Delorme, "Self-Health Monitoring and Wearable Neurotechnologies," *Handb. Clin. Neurol.*, vol. 168, pp. 207–232, 2020.
- [155] H. N. Modi, H. Singh, G. Z. Yang, A. Darzi, and D. R. Leff, "A Decade of Imaging Surgeons' Brain Function (Part II): A Systematic Review of Applications for Technical and Nontechnical Skills Assessment," *Surgery*, vol. 162, no. 5, pp. 1130–1139, 2017.
- [156] B. Brandon et al., "Multimodal human and environmental sensing for longitudinal behavioral studies in naturalistic settings: Framework for sensor selection, deployment, and management," *J. Med. Internet Res.*, vol. 21, no. 8, 2019, Art. no. e12832.
- [157] A. Sano, P. Johns, and M. Czerwinski, "Designing opportune stress intervention delivery timing using multi-modal data," in *Proc. 7th Int. Conf. Affect. Comput. Intell. Interact.*, 2017, pp. 346–353.
- [158] H. Vrzakova, A. Begel, L. Mehtätalo, and R. Bednarik, "Affect recognition in code review: An in-situ biometric study of reviewer's affect," *J. Syst. Softw.*, vol. 159, 2020, Art. no. 110434.
- [159] T. Zhou, J. S. Cha, G. Gonzalez, J. P. Wachs, C. P. Sundaram, and D. Yu, "Multimodal physiological signals for workload prediction in robot-assisted surgery," *ACM Trans. Hum.-Robot Interact.*, vol. 9, no. 2, pp. 1–26, 2020.
- [160] Y. Zhu, C. Rodriguez-Paras, J. Rhee, and R. K. Mehta, "Methodological approaches and recommendations for functional near-infrared spectroscopy applications in HF/E research," *Hum. Factors*, vol. 62, no. 4, pp. 613–642, 2020.
- [161] D. Tao, H. Tan, H. Wang, X. Zhang, X. Qu, and T. Zhang, "A systematic review of physiological measures of mental workload," *Int. J. Environ. Res. Public Health*, vol. 16, no. 15, pp. 1–23, 2019.
- [162] O. Tolvanen, A.-P. Elomaa, M. Itkonen, H. Vrzakova, R. Bednarik, and A. Huotari, "Eye-tracking indicators of workload in surgery: A systematic review," *J. Investig. Surg.*, vol. 35, no. 6, pp. 1340–1349, 2022.
- [163] D. Shouhed, B. Gewertz, D. Wiegmann, and K. Catchpole, "Integrating human factors research and surgery: A review," *Arch. Surg.*, vol. 147, no. 12, pp. 1141–1146, 2012.
- [164] P. Gogalniceanu, F. Calder, C. Callaghan, N. Sevdalis, and N. Mamode, "Surgeons are not pilots: Is the aviation safety paradigm relevant to modern surgical practice?," *J. Surg. Educ.*, vol. 78, no. 5, pp. 1393–1399, 2021.
- [165] G. D. i Flumeri et al., "EEG-based workload index as a taxonomic tool to evaluate the similarity of different robot-assisted surgery systems," in *Proc. Int. Symp. Hum. Ment. Workload: Models Appl.*, Springer, Cham, vol. 1107, 2019, pp. 105–117.