

Emotion Prediction in Performance-Critical Tasks: A Systematic Review of Physiological Signals and Deep Learning Models

Norhawani Ahmad Teridi¹, Tengku Mohd Tengku Sembok², Muhammad Fairuz Abd Rauf³,
Nurhafizah Moziyana Mohd Yusop⁴, Zuraini Zainol⁵, Shahrulfadly Rustam⁶,
Azlinda Abdul Aziz⁷, Hazri Haidar⁸, Mohd Fahmi Mohamad Amran^{9*}

Faculty of Defence Science and Technology, Department of Computer Science,
National Defence University of Malaysia (NDUM), 57000 Kuala Lumpur, Malaysia^{1, 3, 4, 5, 6, 9}
Kulliyah of Information and Communication Technology, International Islamic University Malaysia (IIUM),
Jalan Gombak 53100 Kuala Lumpur, Malaysia²

Faculty of Computer Science and Mathematical, Universiti Teknologi MARA (UiTM) Melaka, Campus Jasin, Malaysia⁷
Prisma Khas Sdn Bhd, Emerald North Plaza, Damansara Perdana, 47820 Petaling Jaya, Selangor, Malaysia⁸

Abstract—Emotions strongly influence how people think, decide, and perform, making reliable emotion forecasting essential in performance-critical environments. Traditional methods such as facial expressions, speech, and self-reports often lack reliability and continuity. Physiological signals offer a more objective alternative, providing continuous indicators of emotional states, while deep learning models are well-suited to capturing their non-linear temporal characteristics. Unlike prior reviews that primarily focus on general emotion recognition or isolated model performance, this study specifically examines emotion prediction in performance-critical contexts through the combined analysis of physiological signals, deep learning architectures, and task-driven requirements. This systematic review synthesizes recent studies on emotion prediction using physiological data and deep learning models. Following the PRISMA framework, relevant studies published between 2021 and 2025 were identified from the Dimensions AI and Web of Science databases, resulting in 25 eligible articles. The review examines trends in physiological modalities, deep learning architecture, emotion representations, and evaluation practices. Beyond summarizing these trends, the review provides a structured comparative synthesis that organizes existing studies according to physiological signal modality, model architecture, performance-critical task context, emotion representation, and evaluation practices, thereby offering methodological guidance for future emotion prediction system design. Findings show that EEG is the most widely used modality, frequently combined with peripheral signals such as heart rate variability, electrodermal activity and electrocardiography in multimodal systems. Hybrid architectures, particularly CNN–LSTM models, dominate current approaches, although attention-based and lightweight models are gaining traction. Key challenges remain, including inter-subject variability, limited real-world validity, inconsistent emotion modeling and non-standardized evaluation. This review highlights current gaps and offers guidance for developing more robust emotion prediction systems in high-performance contexts.

Keywords—Emotion prediction; physiological signals; deep learning; multimodal fusion; performance-critical tasks

*Corresponding author.

I. INTRODUCTION

The ability to automatically identify a person's emotion is an important development towards building emotionally intelligent systems that are being used increasingly in applications including healthcare, HCI and education [1, 2]. Emotion plays a role in cognition, decision-making and performance on tasks in high-pressure situations such as elite athletics, aviation and military operations [1]. Emotion recognition has traditionally been achieved using subjective measures, including facial expression, voice inflection and textual communication; the subjectivity of these measures can be influenced by the surrounding environment [4], making emotion recognition difficult with respect to privacy (e.g., lack of consent, recordings). For this reason, most current research focuses on collecting and analyzing physiological signals, as they provide reliable and unambiguous representation of emotion based on physiological response.

Physiological signals, such as electroencephalography (EEG), electrodermal activity (EDA), variability in heart rate (HRV) and temperature of the skin (ST), provide excellent opportunities to continuously and understandably monitor emotional reactions in daily life through physiological measures. However, they also create significant challenges for analytical processing due to their high dimensionality and non-linearity. Using deep learning models such as convolutional neural networks (CNNs), long short-term memory (LSTM) networks and hybrid architectures is a new way to robustly extract spatio-temporal features from physiological signals. Compared to traditional machine learning algorithms, deep learning models yield improved results for processing physiological signals; therefore, they can be used more effectively than traditional methods [2, 4]. There are still many constraints on current performance-sensitive task research despite recent improvements. It is often difficult for researchers to repeat results between subjects and what succeeds in a controlled setting might not transfer to a real-world scenario depending on an individual's cognitive load, stress level and other elements that can affect the final results [5, 6]. Even

though multimodal fusion has been suggested and tested through the use of attention methods to increase robustness, this type of research has not made much headway beyond the use of benchmark datasets or task environments [7].

Furthermore, these limitations in study design extend to studies of cognitive load exactly as they relate to use cases in a true performance-critical environment. Despite the models being able to achieve high levels of accuracy in controlled laboratory settings, they tend to fall short when it comes to generalizing their results to individuals in a new context, for example, an environment, under different task conditions, such as cognitive load, stress, time pressure, or when dealing with different types of emotions [5, 6]. Multimodal fusion and attention mechanisms, though useful, have primarily been examined through their use on 'standardized' datasets rather than through their application across real-world contexts in which tasks are performed [7].

In recent years, research in emotion prediction using psychological signals has grown significantly, reflecting increased interest in developing more robust and adaptive systems. However, existing studies often focus on isolated aspects such as model performance or specific signal modalities, without providing a comprehensive synthesis that integrates physiological characteristics, deep learning architectures, and contextual requirements of performance-critical tasks. Furthermore, current review studies tend to address emotion recognition in general settings, with limited emphasis on performance-critical environments where cognitive load, stress, and real-time constraints play a decisive role. This creates a gap in understanding how physiological signals and deep learning models can be effectively aligned with task-specific demands in such high-stakes contexts.

To address this gap, this systematic review critically examines recent literature published between 2021 and 2025 on emotion prediction using physiological signals and deep learning models in performance-critical tasks. In contrast to prior surveys that mainly provide broad overviews of affective computing or general emotion recognition, this review offers a structured analytical contribution through a task-oriented and signal-centric comparative synthesis. Specifically, it organizes the literature across five core dimensions: physiological signal modality, deep learning architecture, multimodal fusion strategy, emotion representation, and evaluation practices in performance-critical settings. This framework enables a clearer comparison of how methodological choices relate to real-world task demands and supports the formulation of methodological guidance for future research. By providing a task-oriented and signal-centric synthesis of existing studies, this review aims to bridge the gap between laboratory-based emotion prediction models and real-world performance-critical applications, while contributing a structured comparative framework for interpreting methodological trends, limitations, and future research directions in emotion prediction systems. Specifically, this review is guided by the following research questions:

RQ1: What types of physiological signals and their representations are used for emotion prediction and what deep learning architectures and multimodal fusion strategies are adopted in performance-critical context tasks?

RQ2: Which deep learning architectures and multimodal fusion strategies are adopted for emotion prediction in performance-critical contexts?

RQ3: How are emotion prediction models evaluated in terms of performance metrics, validation practices and emotion representation schemes and what are the implications for model generalization?

RQ4: What are the key methodological limitations and research gaps hinder the development of robust and deployable emotion prediction systems in real-world performance-critical environments?

By providing a task-oriented and signal-centric synthesis of existing studies, this review aims to bridge the gap between high-performance laboratory models and real-world performance-critical applications, while contributing a structured comparative framework for interpreting methodological trends, limitations, and future design directions in emotion prediction systems.

This study is structured as follows: Section II presents a comprehensive literature review on emotion prediction in performance-critical tasks. Section III describes the systematic review methodology, including the search strategy, selection criteria, and data extraction process. Section IV discusses the findings and synthesis of the reviewed studies, highlighting trends, challenges, and research gaps. Section V presents the cross-domain perspectives. Finally, Section VI concludes the study and outlines directions for future research.

II. LITERATURE REVIEW

A. Emotion in Performance-Critical Tasks

Performance-critical tasks require individuals to operate under high cognitive load, time pressure, or competitive conditions and emotional states play a major role in the execution of these tasks, as they hold a big influence over the individuals' cognitive abilities, such as attention, decision-making, motor control and overall task effectiveness. Emotion is a key factor that affects performance-heavy tasks, which involve a great deal of sustained focus and emotional regulation. Thus, to improve performance monitoring and adapt intervention strategies, it is crucial to first understand how emotional states fluctuate.

Research indicates that identifying emotions quickly and accurately not only aids in improving someone's psychological health but is also necessary for improving performance, promoting safety and reducing human error when completing the task [4].

The modern world of affective computing has shifted from reading emotional states through traditional behavioral cues to physiological biomarkers. This is because the biomarkers offer more objective, continuous and tamper-resistant indicators [2]. In studies like that of the WESAD dataset, utilization shows that wearable IoT-based sensors paved the way to monitor biosignals such as electrodermal activity (EDA), heart rate variability (HRV), electrocardiography (ECG) and skin temperature in real time, even while performing dynamic tasks [8,9]. These signals are valuable, as they reflect autonomic nervous system activity,

which is closely tied to stress, cognitive load and emotional regulation under pressure.

Even so, the work of translating physiological emotion recognition from a wellness-oriented setting to a performance-critical setting is still insufficient. Though the multilevel deep neural network proposed in the abstract and other models as such demonstrate strong classification performance in lab-based benchmarks (achieving up to 87.7% accuracy on WESAD), they cannot guarantee their robustness in valid, task-driven environments [8]. The characteristics of performance-critical tasks introduce variables such as physical movement, environmental noise, interpersonal interaction and variable cognitive demands. These factors can greatly affect the quality of the signal and model generalizability [5].

Moreover, literature that can be found often focuses on the innate emotion categories, like stress and baseline, rather than the finer emotional states, such as frustration, flow and anxiety, which are much more relevant to performance contexts. There is also an undeniable gap in studies that validate emotion recognition models in real-world performance scenarios, where ground truth emotional labels are difficult to procure and the consequences of misclassification are high [6].

Emotional dynamics play a big role in affecting the performance of individuals as they execute performance-critical tasks, but it also conquers the realm by shaping adaptive behaviors, an individual's learning efficiency and the development of long-term skills. Emotions are constantly changing regardless of the activity, whether elite sports, military drills, or complicated human-machine interactions, because they evolve with changing task difficulty, type of feedback and environmental stressors. All this happens because temporal emotional transitions are closely linked to the consistency of performance and resilience under pressure rather than static emotional labels. Therefore, it would be more effective to predict emotions for such tasks using models that truly have the capabilities to provide the signals of rapid emotional fluctuations and contextual dependencies that happen during the execution of the task, instead of relying solely on snapshots for emotion assessments.

B. Physiological Signals for Emotion Prediction

Electrophysiological monitoring, specifically electroencephalography (EEG), is mainly used to examine the relationship between emotional states and performance outcomes in performance-critical tasks, as it provides a clear and direct insight into cortical activity. As studied in prior research, competitive environments can significantly influence cortical efficiency, which includes attention, emotional regulation and cognitive workload. It was shown that expert performers exhibit much more efficient neural processing compared to novices [10]. However, electrophysiological monitoring remains difficult due to its susceptibility to many factors, like environmental noise, motion artifacts and inter-subject variability. Over the last few years, there have been promises of improving the accuracy of reading emotions using an electroencephalogram (EEG), particularly for precision sports such as shooting. These new technologies allow us to encode and represent signals more accurately, particularly through sparse coding and time-frequency transformations [11].

Besides cortical emissions from the brain, other psychophysiological indicators (e.g., heart rate variability, or HRV) can also offer insight into emotional regulation, stress resilience and the autonomic balance of the body. Many studies have shown strong correlations between HRV profiles and emotional regulation strategies. Some of these studies have found that greater levels of HRV are generally associated with increased ability to maintain sustained focus when under duress [12]. Therefore, HRV acts as a prominent peripheral indicator in a high-performance context by measuring autonomic nervous system activity associated with an emotion (stress or cognitive load). Also, HRV measurements can easily complement EEG assessments of an individual's emotional state.

Continuous and non-invasive methods of measuring physiological signals, including heart rate variability (HRV), skin temperature and electrodermal activity, have been made possible by the rapidly growing field of wearable technology. These physiological parameters provide valuable insights into emotional arousal, especially in dynamic or stressful situations [9]. The integration of these signals is critical for environments that are high-stakes, as emotional reactions change quickly based on other factors such as stress, movement and changing environmental conditions.

Individuals can develop interoceptive awareness by recognizing their internal body signals, such as breathing and heart rate. Recently, studies have explored interoceptive awareness in high-performance athletes. Research indicates that higher interoceptive awareness leads to improved emotional regulation and consistent performance in precision sports. Therefore, interoceptive awareness allows athletes to maintain physiological equilibrium when under pressure [13]. Environmental or situational disruptions can easily interfere with interoceptive awareness, which will affect emotional stability and reduce performance in the given task [14].

Overall, studies have shown a spotlight on the value of combining central and peripheral physiological signals because it shows a more comprehensive assessment of emotional states during performance-critical tasks. Nevertheless, the major challenges remain, which are variability in signal quality, acquisition conditions and contextual influences. Thus, these limitations prove the need for robust modelling approaches that can handle multimodal physiological data under dynamic and ecologically valid conditions. Therefore, the following section will discuss the initiative of adopting advanced deep learning techniques.

C. Deep Learning Approach for Emotion Prediction

Deep learning has greatly changed emotion recognition, as it enables the automatic extraction of high-level domain-invariant feature representations directly from physiological data, such as electroencephalogram (EEG) signals [15]. Its reliance on "hand-crafted" features, such as time, frequency and non-linear analysis, makes traditional machine learning (ML) deemed inefficient because it is low-level, redundant and requires extensive professional expertise to design. The DL model, however, learns modality-specific representations through deep layers, so it effectively sustains performance even when faced with complex and non-stationary input data.

Due to its capability, recent studies have shown that it has increasingly adopted deep learning architectures tailored to the temporal and non-stationary nature of physiological signals. Recurrent neural networks (RNNs), including Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, are widely used due to their effectiveness in capturing sequential dependencies and temporal dynamics in emotion-related physiological data. These are particularly suitable for performance-critical tasks as emotional states evolve continuously in response to task difficulty, cognitive load and environmental stressors.

The results from [16] showed that GRU-based RNN models can adequately classify EEG signals acquired while the subject completed an interactive, serious-game activity across various levels of mental stress. The results support the theory that modelling temporal structures in EEG data (as opposed to relying exclusively on static feature representations) is critical when classifying the state of mind within interactive and task-focused environments. This matches closely to how performance-critical situations occur due to dynamic rather than passive induction of emotion.

An alternative example provided by [17] used hybrid deep learning models such as long short-term memory (LSTM), gated recurrent unit (GRU) and one-dimensional convolutional neural network (1D-CNN) to process remote photoplethysmography (rPPG) data to detect stress with an overall accuracy of 95.83%. Additionally, this study conducted a comprehensive evaluation of the effectiveness of hybrid models, specifically designed to utilize temporal modelling and feature extraction, demonstrating that signal transformation to the frequency domain yields superior results. As demonstrated in both studies, there exists a degree of physiological variability among subjects. Therefore, both studies encountered difficulties in generalizing their findings across different subjects. The resulting recommendations to utilize subject-specific or adaptive models are aimed at producing stable and robust models for use in real-world applications. These studies highlight the use of deep learning to leverage both temporal dynamics and spectral information to predict emotions accurately.

Additionally, [18] presented an optimized one-dimensional convolutional neural network (1D-CNN) framework for emotion detection based on electroencephalography (EEG). This includes channel and feature selection strategies along with graph-based empirical mode decomposition (EMD) data augmentation. The study achieved mean accuracy rates of 97.6%, 95.3% and 89.0% for the MAHNOB-HCI, SEED and DEAP datasets, respectively. These results confirm that lightweight 1D-CNN architectures can produce excellent results with a reduced computational cost by selectively employing channels and refining features. In [19], the authors took a step beyond unimodal physiological data by developing a deep learning multimodal framework that fused text, audio and video modalities to predict psychological states. The authors' hybrids of BERT and CNN and BERT and LSTM (Long Short-Term Memory) produced accuracies of up to 89.3%. This further illustrates the benefit of fusing multiple modalities with transfer learning to provide a comprehensive representation of emotional states.

Recognizing that physiological and behavioral responses vary from person to person makes it important to keep in mind that there's difficulty in creating generalizations via models. Thus, advocating for subject-specific or adaptive rather than generalized modelling strategies provides potential with respect to improving stability in real-world applications. Taken together, these strategies demonstrate the ability of deep learning to utilize temporal dynamics, spectral features, optimally designed architecture and multimodal integration for accurate contextually aware emotion prediction. Future research should be directed toward improving the interpretability of the model, addressing data limitations through realistic augmentation and ensuring ethical delivery in sensitive mental health settings.

D. Emotion Prediction in Performance-Critical Context

Emotion prediction will have completely different outcomes when predicting an emotion in a performance-critical versus a standard affective computing environment. This is because emotional responses are tightly knitted with task demands, cognitive workload and time pressure. In such environments, physiological changes may reflect any other processes or factors that overlapped. For instance, emotional arousal, mental stress, fatigue and attentional shifts. The addition of such factors creates a challenge to isolate emotion-specific patterns without task-aware modeling. There have been EEG-based studies that show deep learning can learn discriminative representations when given complex physiological inputs. However, the interpretation and stability of those representations can vary across tasks and individuals, as they are all different [20].

A key characteristic of performance-critical contexts is that emotional states evolve dynamically during ongoing task execution rather than being elicited through static stimuli. Thus, task-driven protocols were used to model the dynamics using deep learning to induce graded stress and performance pressure. For example, the recent study in the IEEE Journal of Biomedical and Health Informatics employed EEG data collected within a serious game and used a GRU-based recurrent architecture to estimate the multilevel stress. This supports the view that temporal modeling is important as emotional states fluctuate as task engagement happens [16]. However, ecological validity remains a persistent challenge. This is because many models report strong performance under controlled acquisition conditions, yet performance-critical settings introduce movement artifacts, environmental noise, interpersonal interaction and unpredictable task progression that can degrade signal quality and reduce generalizability of the signals. Thus, it encourages the use of multimodal physiological sensing, in which combining central and peripheral signals can provide complementary evidence to improve its stability or robustness. Recent work has proposed fusion-based multimodal emotion recognition approaches, including EEG and ECG integration, to polish emotion classification performance and capture multidimensional affective states in a more comprehensive manner [3].

In addition to multi-mode representations, there is a rising interest in more sophisticated multi-modal fusion techniques for representing relationships among the various modalities. For instance, [21] recently proposed a method for recognizing

emotion in time series of multimodal data via online multimodal hypergraph learning. They mirror the increasing focus on systems that accommodate complex cross-modal dependence as opposed to simply appending data together or utilizing late fusion alone. An additional issue is that emotions are presented and evaluated in different ways. Hence, performing cross-study comparisons becomes very complicated. Examples of differences included discrete to dimensional emotion categories and differentiating between stress from baseline classifications. This limitation is closely linked to heterogeneous metric reporting and validation design in which accuracy-only reporting may obscure class imbalance and overstate model readiness for real-world use. Recent studies of multimodal EEG-based emotion detection have also noted difficulties in the design of fusing signals from different modalities, defining how sensor modalities will represent emotions and creating ways of classifying signals into specific emotion categories. In addition to these issues, they have further reinforced the need for more standardized evaluation metrics and reporting procedures in this area [22].

In summary, physiological responses are greatly dependent on the individual and context-dependent as well. The new work being done on using neurophysiological workload recognition is particularly relevant to this type of work in emotion prediction in constantly changing task environments. It reflects a direction that is highly relevant to emotion prediction in dynamic operational settings where task conditions change over time [23].

For the effective prediction of emotion in performance-critical contexts, we must develop standardized benchmarks, subject-independent validation and interpretable model designs with the potential to be deployed in high-stakes settings. The synthesis and gaps identified in this review correspond directly to those three factors.

III. METHODOLOGY

A. Methodology Used

This research adhered to the PRISMA guidelines during the research process to uphold transparency within the research while also enabling methodological rigor. While prior surveys in affective computing and emotion recognition have broadly examined emotion detection techniques and model performance, they often emphasize general-purpose settings without explicitly considering performance-critical environments. In contrast, this review adopts a task-oriented and context-driven approach by focusing on studies that integrate physiological signals, deep learning architectures, and real-world performance demands. This distinction enables the present review to advance beyond existing surveys by aligning methodological evaluation with performance-critical task requirements. The literature search was carried out on two large-scale databases: Dimensions AI and WoS. These databases were selected because they provide extensive coverage of peer-reviewed academic journals from a variety of fields, including AI, affective computing, neuroscience and performance analytics.

Fig. 1 illustrates the citation trend of articles from 2016 to 2025, showing a consistent increase in both total citations and

average citation counts. This trend reflects growing research interest and reinforces the need for a comprehensive systematic review.

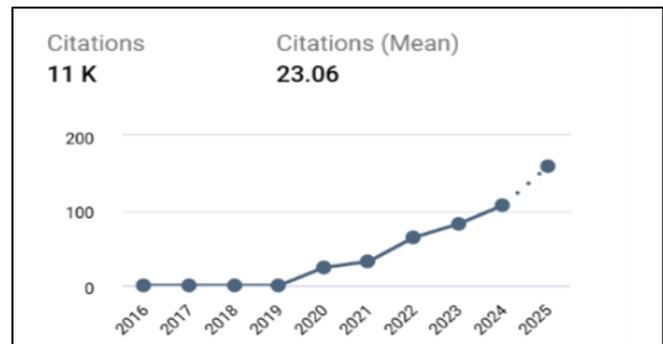


Fig. 1. The citation rate of articles.

The original research journal articles were searched for and published between Jan 2021 and Dec 2025 in order to reflect up-to-date research on physiological signal processing techniques using deep learning to predict emotion, but also to ensure the research is still in line with current methodology. The overall database search and screening process is summarized using a PRISMA flow diagram, as presented in Fig. 2.

B. Study Selection

The study selection process followed the PRISMA-recommended multiple-stage screening process. All obtained records from Dimensions AI and Web of Science were searched and ranked before screening through to establish their relevance to the review objectives. The first step in this process was screening titles, during which obviously irrelevant studies such as those unrelated to emotion prediction, physiological signals, or performance-related tasks were excluded. The second stage consisted of screening abstracts in order to make stated or implied conclusions about each article's use of physiological data, machine learning and deep learning for emotion prediction. If an article did not meet these criteria, it was excluded as well.

Once the group of articles from stage 2 was established, the last step involved conducting full-text reviews of the remaining articles in order to determine whether they contained the required information:

1) *Physiological signals*: The study must involve physiological data such as EEG, HRV, ECG, EDA, or GSR. Studies relying solely on facial expressions, speech, or self-report data were excluded.

2) *Deep learning models*: The study must apply deep learning techniques, including LSTM, CNN, GRU, hybrid architectures, or transformer-based models. Studies using only traditional machine learning models (e.g., SVM, k-NN) without deep learning were excluded.

3) *Performance or task context*: Research must take place in a context directly related to performance or tasks. This includes, but is not limited to, performance in a sport, evaluating workload, completing tasks that result in stress, completing tasks where, for example, participants are under time pressure, as well as competing against other individuals or

groups to accomplish actual tasks that are routinely completed in a work environment.

Studies were excluded if physiological signals were not used for prediction, if emotion analysis was not a primary objective, or if the task context was not performance-related.

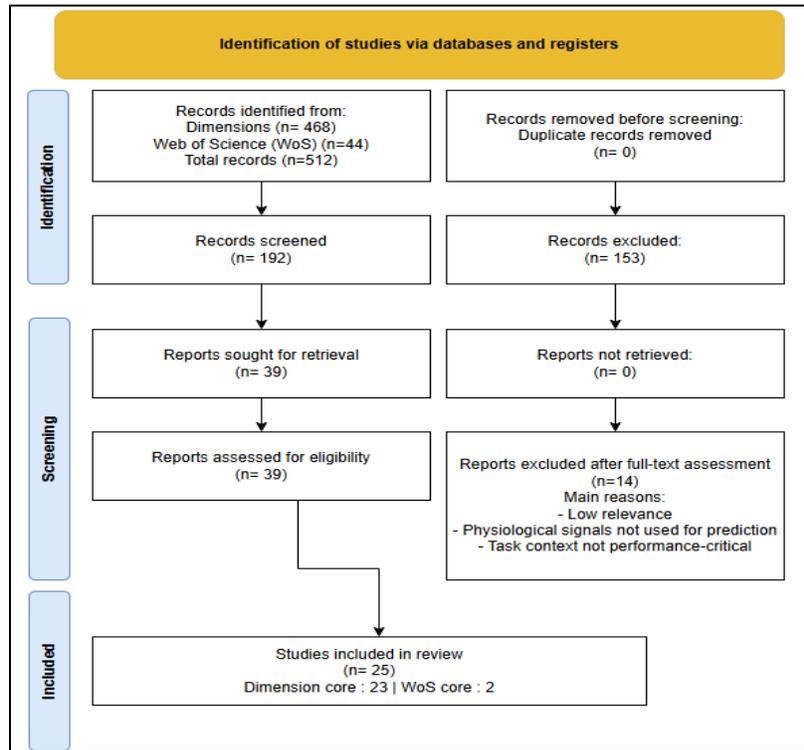


Fig. 2. Literature search strategy process step-by-step using the PRISMA flowchart.

C. Quality Assessment

The quality assessment was performed on only full-text studies and used to exclude unsuitable studies from the systematic literature review by ensuring that they met the definition of studies that used physiological signals and deep learning (DL) models to predict emotions in performance-oriented tasks.

The quality assessment included reviewing and validating the actual methodology used in each study (specifically, the methodological use of physiological signals, DL models and task context) as was described in each study's full-text article. The quality assessment framework was drafted to assist with final eligibility validation by applying the framework to all candidates' studies that were retained from initial screening to ensure that the same quality assessment criteria were used across all candidate studies. Each candidate study was assessed against three core criteria (E1-E3) to determine eligibility, "yes" [Y] to all three criteria for inclusion in the final synthesis of results.

1) *E1: Use of physiological signals as model inputs:* Physiological signals (EEG, ECG, HRV, EDA/GSR) must be used as direct inputs for the modeling, in which they are represented in any valid format (as raw signal, extracted features, or time-frequency). Any signal that is used for descriptive or post-hoc analysis would not count.

2) *Deep learning models used for emotion-related prediction:* A model based on deep learning must be employed for prediction, such as a CNN, LSTM, GRU, Transformer, or

hybrid model (i.e., models on which deep learning is based), and the predictions must relate to emotions. The predicted outcomes must be related to an emotion class (as an example, to an arousal/valence score or to a level of stress) when emotion is treated as the proxy for prediction.

3) *E3: Presence of a genuine performance-critical task context:* The prediction of emotions must occur within an actual performance-critical task, for example, competitive sports, workload-intensive activities, stressful or time-pressured tasks, or actual real-world/competitive environments. This task must involve meaningful performance requirements; generally, there would be some measurable performance associated with a task or other operational relevance that could support those performance metrics.

This task-specific inclusion criterion further differentiates the present review from conventional surveys, which typically do not restrict their analysis to performance-critical contexts.

The fifth step of the quality assessment was to perform a structured data extraction and coding process after conducting the final eligibility validation. This data extraction process was undertaken with the goal of achieving methodological consistency, comparability across studies and transparency in the synthesis of the evidence. Only papers that met all eligibility criteria (E1-E3) were included in this stage of the assessment.

The investigator created a Data Extraction Table (DET) a priori to systematically extract essential methodological and

outcome-based attributes from each of the included studies. The data extraction focused on descriptive characteristics and quality indicators that represent the robustness and applicability of the proposed methods.

The DET comprised five main categories, as described below:

- **General Study Information:** The papers included in the study were varied in terms of genre and study period and provided more than enough detail for coding purposes, but this was done only once and may have changed subsequently. In addition, the source of data used in each study was extracted, including whether the study relied on publicly available benchmark datasets or self-collected physiological data obtained from controlled or real-world experimental settings.
- **Physiological Signal Characteristics:** Various types of physiological signals were used for analysis, including EEG, ECG, HRV, EDA and GSR; thus, physiological signals were extracted from the database. Additionally, single modality vs. multi-modality, the appropriate pre-processing steps used and operational integrity were also extracted from the database for further analysis and transparency in methodology.
- **Deep Learning Model Configuration:** The differences in deep learning architectures and the variety of input types (raw signals, features from the physiological signal, or spectrograms) were recorded and coded for future comparative modelling.
- **Emotion Representation and Performance Measures:** Different representations of emotion (categorical versus arousal and valence) and performance measures about Task-related performance (i.e., response to accuracy, response time, scoring and errors) were included to confirm that performance was primarily evaluated per outcome.
- **Results and Limitations:** In addition to reporting the maximum recorded success for each reported study (best practices), information about the number of datasets and acknowledged limitations was coded to permit cross-study comparisons and the generation of ideas and constraints around generality across studies.

After the conclusion of the Data Extraction Table (DET), a Relevance-based screening was performed to determine the ultimate set of studies included in this synthesis. Each study was assessed systematically and categorized according to its level of fit for the objectives of this review.

The assessment of relevance was based on the extent to which each study met core eligibility criteria (E1–E3) with an extra emphasis on: 1) the application of deep learning models for predicting emotions directly and 2) using emotion predictions in a real-world performance-critical task. The relevance level of the studies was categorized into three levels. High relevance referred to studies that met all three eligibility criteria, applied a deep-learning model to explicitly predict emotions related to an event, and addressed tasks of substantial real-world significance, thereby forming the core evidence for

the systematic review. Medium relevance referred to studies that satisfied the primary eligibility criteria but exhibited one or more limitations in task realism, the level of detail describing the performance context, or the emphasis placed on emotion-related outcomes, while still contributing valuable information to the review. Low relevance referred to studies that lacked sufficient detail to meet the eligibility criteria or failed to satisfy all three criteria, such as treating emotion prediction as a secondary or peripheral outcome, lacking a clearly defined performance context, or providing inadequate methodological detail, and therefore were either excluded from the systematic review evidence or retained with a clearly documented rationale for their inclusion or exclusion.

Only studies with High and Medium relevance were used as part of the final analysis and synthesis. This method of filtering studies with respect to relevance served to keep the final collection of evidence in the systematic review methodologically sound, collectively coherent, and directly related to the objectives of the systematic review.

D. Search Result

The original database search produced a total of 512 records, with 468 coming from Dimensions AI and 44 from Web of Science. After title and abstract screening, 192 records were eligible for further consideration. After reviewing, 153 records were excluded because they did not meet the criteria for inclusion. Next, 39 complete articles were obtained and reviewed for eligibility. Of those articles, 14 studies were excluded after the full-text review based on their lack of relevance, absence of physiological signals to predict emotion, or because they did not support a task that required performance. A total of 25 studies met all criteria for inclusion and were ultimately included in the systematic review. Of those studies, 23 were retrieved from Dimensions AI, and 2 were retrieved from Web of Science.

E. Risk of Bias Assessment

To assess the risks of bias that exist within these studies and that may hinder the accuracy of the findings from these studies, a risk-of-bias assessment was performed. The majority of identified risks constituted aspects of the studies that are likely to influence the strength of the findings; for example, small sample sizes, lack of control conditions, unbalanced dataset types and NOT reporting training and validation procedures.

Many of the studies have been conducted in controlled laboratory environments, which limits our ability to generalize results from laboratory settings to performance-critical real-life settings. In addition, the variations among the devices and techniques used to obtain physiological signals from participants across studies can also create measurement bias across studies. Despite these limitations, the use of multiple datasets, numerous task domains and comparative model performance evaluations among studies to determine model accuracy helped minimize the influence of each individual study's bias on our interpretation of results and our identification of research gaps in discussion.

IV. RESULTS AND DISCUSSION

The findings discussed here derive from a systematic review of 25 studies that investigate the use of physiological signals

with deep learning models for emotion prediction in performance-critical tasks. The synthesized results identify prevailing trends with respect to signal modalities, learning architectures, fusion strategies and application domains and discuss how different physiological signals contribute to emotion recognition performance as well as the ways in which deep learning approaches have been utilized to model the affective states of individuals in high-demand and task-oriented environments. The discussion also highlights methodological patterns, research gaps and performance evaluation practices that may be deduced from the literature reviewed, all relevant for guiding future research on intelligent emotion-aware systems.

A. Overview of Selected Studies

A collection of 25 primary research articles covering emotion forecasting at work-critical job levels using physiological information with deep-learning methods was included in the final selection (2021-2025) for this review. The included studies were from many different focus areas, such as elite sports, cognitive workload measurements, psychological stress evaluations, human-computer interactions, and task performance assessments.

The majority of the studies used controlled laboratory experimental environments, although participant numbers and experimental designs varied across the 25 studies. A few of the studies collected small laboratory samples for physiological data because they did not have access to larger sample sizes; however, many of the studies used publicly accessible datasets for creating and validating their emotion prediction models. Accordingly, the reviewed studies used either publicly available benchmark datasets or self-collected physiological datasets, mostly from controlled laboratory settings.

All of the 25 studies in the final selection of studies used physiological data to produce predictions for human emotion recognition, either as standalone signal sources or in a multi-

signal (i.e., combined from 2 or more sources) manner. To identify complex temporal and nonlinear patterns in their data, many of the studies developed and utilized deep-learning models. There were many similarities between the studies in relation to emotion prediction, but there were also a lot of differences in the selection of signals, development of model architectures, and adoption of evaluation methods, limiting researchers from directly comparing across studies. The next section explores these issues in detail. For a summary of the studies in the final selection (concerning the task completion context, used physiological signals, used signal fusion methods, used model architectures, and representations of emotion), see Table I.

B. Physiological Signals Used

The studies that made up the reviewed literature used different types of physiological signals as input to models of emotional states in highly demanding tasks. Different types of physiological signals were used to reflect the many ways different physiological signals are used to model these emotions and their associated stress levels. The table below lists and describes all of the physiological signals employed in the studies reviewed.

As indicated by Table II, the most used signal modality across the published studies was electroencephalography (EEG), used in 22 of the studies reviewed. EEG as a method of input was used separately and/or as supplemental input to other forms of input within the multimodal classification systems examined. EEG input types have included raw EEG signals, features based upon frequency bands, and time- and frequency-transformed EEG signal data, enabling the deep learning-based methods to learn patterns of neural activity that correspond to discrete emotions or arousal levels. In addition, most of the studies that used EEG as a measure used it primarily within the areas of affective computing (e.g., emotion detection), work-related performance evaluations, and sport performance evaluations.

TABLE I. SUMMARY OF SELECTED STUDIES ON DEEP LEARNING-BASED EMOTION PREDICTION USING PHYSIOLOGICAL SIGNALS

Paper ID	Author (Year)	Domain / Task Context	Physiological Signal(s)	Model type	Input format	Fusion method	Emotion representation	Best Result Reported
P01	[24]	Human-Computer Interaction (HCI) / Stimulus elicitation (watching videos)	EEG, GSR, ST, HR	CNN + LSTM (Hybrid)	Raw signals (processed)	Hybrid (Early Fusion)	Discrete (7 categories: calm, happy, disgust, surprise, anger, sad, fear)	96.21% Accuracy
P02	[25]	Affective Computing / Human-Computer Interaction (HCI)	EEG, GSR, ECG	2D-CNN, 1D-CNN and LSTM	Raw signals (structured format)	Late Fusion	Arousal-Valence classes: HVHA, HVLA, LVHA, LVLA	99.0% Accuracy (AMIGOS), 90.8% Accuracy (DREAMER)
P05	[26]	Medical & Mental Health / Brain-Computer Interface (BCI)	EEG	Bi-Directional Long Short-Term Memory (Bi-LSTM)	Selected features (statistical, wavelet, Hurst exponent)	None (Single-modal)	Arousal-Valence + Liking	Valence: 99.45%, Arousal: 96.87%, Liking: 99.68% (Accuracy)
P06	[1]	Affective Computing / Naturalistic	HR, Temperature, EDA	LSTM	Raw signals with emotional cues	Early Fusion	Arousal-Valence	96% Accuracy (independent)

		Conversations (Debates)						93% (combined)
P07	[9]	Tourism & Consumer Applications / Real-World Recommender System	HR	Hybrid CNN + LSTM	Raw signals or basic HR features	Hybrid	Discrete (self-labelled categories)	Not in source (reported as "promising")
P08	[27]	Embedded Systems & Edge Computing / Wearable Applications	ECG, EDA	1D-CNN (lightweight)	Hand-crafted features + Deep Learning integration	Hybrid (Model integration)	Discrete (Fear detection / Binary)	F1-score: 80%, Accuracy: 74%
P11	[28]	Neuroscience & Embedded Systems / FPGA implementation	EEG	CNN-LSTM + ResNet-152	Features (extracted via ICA)	Late Fusion	Not in source	Improved speed/area (numerical accuracy not in source)
P13	[7]	Smart Healthcare / Mental Health Monitoring	EEG and ECG	CNN-LSTM (Optimized with Bayesian)	Raw signals or basic features	Early Fusion	Discrete (3 categories: neutral, tense, concerned)	Accuracy 92.1%, AUC-ROC 0.96
P14	[6]	Human-Computer Interaction (HCI) / Wearable & Real-time Systems	EEG (Selected channels)	1DCNN-BiLSTM (Hybrid)	Raw signals or basic features	Hybrid (Model design)	Not in source (Multiclass)	85.16% Accuracy
P15	[5]	EEG-based Emotion Recognition / Generalizable AI	EEG (Multichannel)	ATDD-LSTM (Attention and Domain Discriminator)	Raw signals or basic features	Not applicable (Attention mechanism)	Not in source	Superior performance (numerical values not in source)
P16	[29]	Online Learning & Education	EEG (5 channels: FP1, AF3, F7, T7, FP2)	1D-CNN	Raw signals	None (Single-modal)	Discrete (7 categories: boredom, confusion, frustration, curiosity, excitement, concentration, anxiety)	Accuracy >99%
P17	[30]	Mental Health & Clinical Applications / Real-Time Emotion Recognition	EEG	Machine Learning (XGBoost, LightGBM, KNN)	Statistical and mathematical features	None (Comparative study)	Not in source	96.77% Accuracy (with KNN)
P18	[31]	Mental Health & Stress Management / Diagnostic Aid	EEG	CNN-LSTM (StressNet)	Images (Azimuthal projection-based images)	Hybrid (Sequential)	Discrete (Binary: Stress vs Normal)	97.8% Accuracy
P19	[32]	Wearable Technology / Affective Computing	EEG and BVP	ConvNet-LSTM	Raw signals or basic features	Early and Late Fusion Comparison	Dimensional (Arousal-Valence)	71.61% Accuracy (Early Fusion)
P20	[33]	Medical Rehabilitation / Assistive Technology (Stroke Patients)	EEG	LSTM	Spectrogram/Scalogram (2D Images)	None (Single-modal)	Arousal-Valence (3, 5, or 9 categories)	Valence: 88%, Arousal: 90% Accuracy (9-class)
P21	[34]	Industrial Safety / Occupational Health (Electric Power Workers)	ECG (Heart rate signal)	1D-CNN	Processed raw signals (3D stacking operation)	None (Single-modal)	Discrete (4 types)	98.57% Accuracy, 98.73% F1-score
P22	[35]	Biometrics & Gait Analysis / Virtual Reality (VR)	EEG and Inertial signals	Attention-Based CNN Fusion Model	Images (Wavelet-based "virtual emotion images")	Hybrid (Decision Fusion)	Not in source	98.73% Accuracy
P23	[36]	Mental Health & Psychology / VR Therapy	EEG and HR	CNN-LSTM (Hybrid)	Raw signals or features	Hybrid	Psychological Stress (Arousal-Valence)	Superior performance (numerical values not in source)
P25	[37]	Human-Computer Interaction (HCI) / Social Robotics	EEG and Speech	TMNet (Transformer-fused CNN-BiLSTM-BiGRU)	Features or processed representations	Late Fusion (via Transformer)	Not in source	98.89% Accuracy

P26	[38]	Smart Driving / Vehicle Safety	HR, Breathing Frequency, EDA, EMG	CNN + LSTM (Hybrid)	Raw signals or basic features	Hybrid (Model and System level)	Not in source	Significant enhancement (numerical values not in source)
P27	[39]	Sports & Exercise Quality / Physical Therapy	EEG, ECG, GSR, HR and Facial expressions	Residual Network with Attention Modules	Features	Hybrid (Early Fusion & Multi-Head Mutual Attention)	Discrete	88.8% Accuracy
P28	[40]	Digital Health / Daily Emotion Monitoring and Forecasting	HR and Step Count	Cluster-guided attention model	Raw signals or basic time-series features	Hybrid	Discrete Binary (Positive vs Negative)	74.4% Accuracy (Current), 72.9% (1-hour forecast)
P29	[24]	Behavioral Analysis / Gaming (Gameplay)	EEG	ResNet18 (CNN) + ELM-W-AE	Images (CWT Scalograms)	Not in source	Discrete (4 categories)	99.6% Accuracy
W1	[41]	Brain-Computer Music Interface (BCMI) / Music Technology	EEG, HRV, EDA, ECG	Not in source	Features (EEG and physiological features)	Hybrid	Arousal-Valence	Accuracy improvement of 10.2% (Valence) and 9.3% (Arousal)
W10	[8]	Mental Healthcare / IoT-based Wearable Sensing	ECG, EDA/GSR, ACC, TEMP, RESP, BVP	Hierarchical Deep Neural Network (CNN-based)	Multivariate time-series data	Model-level / Late Fusion	Discrete (3 categories: baseline, stress, amusement)	87.7% Accuracy (WESAD dataset)

Note: Paper IDs P01–P29 correspond to studies retrieved from the Dimensions AI database, while W1–W10 correspond to studies retrieved from the Web of Science database.

TABLE II. PHYSIOLOGICAL SIGNALS USED IN DEEP LEARNING-BASED EMOTION PREDICTION STUDIES

Physiological Signal	No. of Studies	Usage Mode	Input Representation	Emotion / State Predicted	Typical Task / Application Context
EEG	22	Single, Multimodal	Raw signal, frequency bands, time–frequency, features	Discrete emotion, arousal–valence, stress	Affective computing, workload tasks, sport-related tasks
HR / HRV	11	Multimodal	Time-domain features, frequency-domain features	Stress, arousal	Workload tasks, stress induction, wearable monitoring
EDA / GSR	7	Multimodal	Raw signal, statistical features	Stress, arousal	Controlled lab tasks, wearable-based studies
ECG	7	Multimodal	RR intervals, waveform features	Stress, emotional state	Healthcare-related emotion monitoring
Skin Temperature	3	Multimodal	Raw signal, mean variation	Stress-related emotional states	Wearable sensing, laboratory tasks
Respiration / BVP	4	Multimodal	Time-series features	Stress-related states	IoT-based wearable datasets
Accelerometer (motion)	4	Multimodal (auxiliary)	Activity features	Context-aware stress/emotion	Wearable and ambulatory settings

Most of the studies investigated peripheral physiological signals in multimodal configurations. Heart rate (HR) and heart rate variability (HRV) were used in a total of eleven studies and were typically represented by time- and/or frequency-domain features. These physiological signals were closely linked to detecting stress and arousal; in particular, they are well suited for workload-intensive and wearable monitoring applications. Electrodermal activity (EDA) and electrocardiography (ECG) were used in a total of seven studies and were principally used to capture the sympathetic nervous system responses and to detect the stress and emotional response from participants using EDA and ECG in controlled laboratory and healthcare-related settings.

Other physiological signal modalities such as skin temperature, respiration, or blood volume pulse (BVP) and motion derived from accelerometer-based sensors are used less frequently than the above modalities and serve as complementary inputs within multi-modal systems.

Contributing contextual or auxiliary information, these signals can assist with the interpretation of emotional states when used in wearable sensing and ambulatory settings, due to the influence of movement and physiological variability.

As shown in Table II, there is a significant reliance on EEG as the primary modality used with deep learning-based emotion prediction systems, while peripheral physiological signals are used in conjunction with the EEG signal to increase the robustness and contextual awareness for emotion prediction. The use of multiple modalities, including integrating central and autonomic responses to create emotional states when modeling performance-critical task emotion states, is emphasized through this multimodal trend.

C. Deep Learning Models Applied

Various deep learning frameworks have been applied to analyze and investigate physiological traits within a performance-relevant and dynamic timeframe for emotional

assessment task completion through accomplished physiological signal processing. The more popular models included Convolutional Networks (CNN), Long Short-Term Memory Networks (LSTM), Gated Recurrent Units (GRU), and Hybrid Models.

Convolutional Neural Network-based methods were typically used for feature extraction (e.g., EEG, Spatial/Time Frequency), while Recurrent Models (LSTM & GRU) were used to time correlate physiological signals over the course of uninterrupted task execution. LSTM models were more prevalent due to their effectiveness in modelling long-term temporal patterns, while GRU models were occasionally preferred for their computational efficiency.

An increasing number of studies adopted hybrid architectures, such as CNN combined with LSTM or GRU layers, to jointly learn spatial and temporal features. More recent work also explored attention mechanisms and lightweight architectures to improve interpretability and support real-time deployment. A summary of the deep learning models and architectural trends across the reviewed studies is presented in Table III.

TABLE III. SUMMARY OF DEEP LEARNING MODELS USED FOR EMOTION PREDICTION.

Model Type	Signal Modality	Task Context	Key Strength
CNN	EEG	Cognitive tasks, stress	Spatial feature learning
LSTM	EEG, HRV	Continuous tasks	Temporal dependency modelling
GRU	HRV, ECG	Mental workload	Computational efficiency
CNN-LSTM	Multimodal	Performance-critical tasks	Spatiotemporal learning
Attention based	EEG, multimodal	Real-time prediction	Interpretability

D. Input Representation and Fusion Strategies

Multimodal fusion strategies were widely adopted in the reviewed studies to enhance emotion prediction performance by integrating complementary physiological signals. These strategies aimed to combine information from central signals, such as EEG, with peripheral signals, including HRV, ECG and EDA, to obtain a more comprehensive representation of emotional states in performance critical tasks.

The most reported approach was featuring level fusion, where features extracted from multiple physiological modalities were concatenated or jointly learned before classification. The use of this strategy enabled early-stage comprehension of the interplay between different types of data by way of deep learning models typically resulting in greater predictive accuracy for many outcome measures. However, the typical requirements associated with performing feature-level fusion included significantly higher computational complexity and careful consideration of the alignment of features.

Another approach to performing fusion across the modalities was to use decision-level fusion, where each modality was used to train an independent model and then the output from each model across the various modalities was fused either through

voting schemes based on the majority of votes received per modality, through weighted averaging using weights defined by the respective mean-square error associated with the trained model for each modality, or through the use of ensemble techniques (i.e., stacking of models), thereby providing more flexibility and robustness than feature-level fusion, as well as resulting in less cross-modality interaction in some instances.

The various representations of physiological data in conjunction with the manner in which the studies previously discussed have performed fusing of modalities for emotion and stress prediction over time can be clearly identified in Table IV. Compared to prior methodologies, a shift towards more structured and model-aware strategies with regard to the performance of data fusion has occurred over time.

The input representations can be broadly grouped into three categories:

- Raw or minimally processed signals: There are studies (e.g., P02, P19, W10) that utilize either raw physiological sequences or multivariate time series. This provides both temporal dynamics and no hand-crafted “bias”, thus relying much more heavily on the model capacity than would be expected from the ability to learn the discriminatory nature of the data.
- Time-frequency and image-based transformations: Several studies (e.g., P13, P18, and P22) have produced some form of transformation of the physiological signals to either a spectrogram, azimuthal projection, or image of a virtual emotional effect (e.g., VR). Similar or greater capabilities using CNN architectures also allow for improved spatial feature learning in EEG signals.
- Hybrid and multi-domain features: P27 has combined both hand-crafted multimodal features and attention mechanisms. In P25, both speech MFCCs and EEG sequences are integrated. This reflects an attempt to achieve the best of both worlds by using domain knowledge while utilizing data-driven learning.

TABLE IV. COMPARISON OF INPUT REPRESENTATION AND FUSION STRATEGY.

Study	Input Representation	Primary Fusion Strategy
P02	2D Spatial Images (EEG) + 1D Raw (ECG/GSR)	Late Fusion (Majority Voting)
P13	Time-Frequency Spectrograms / Raw Signals	Early Fusion (CNN-LSTM)
P18	2D Azimuthal Projections	Hybrid (Integrated spatial-temporal)
P19	Raw Physiological Sequences	Early & Late Comparison
P22	DWT Virtual Emotion Images	Late Fusion (Critic/Weighted Voting)
P25	MFCCs (Speech) + 1D Sequences (EEG)	Late Fusion (Transformer-based)
P27	Hand-crafted Multi-modal Features	Hybrid (Residual Attention)
W10	Multivariate Time-Series	Model-Level Fusion (Unified feature rep)

Fusion strategies show a clearer trend than representations, which are grouped into four categories:

- There has been a large use of late fusion strategies found in the literature (e.g., P02, P22, P25), with a lot of late fusion using either voting or weighted decision techniques. The trend of late fusion in this literature suggests that the preferred choice of a modular design is that each of the different modalities is modelled independently, making them better able to withstand or accommodate loss of data/how reliable the data will be before it is fused together.
- Early fusion strategies are less common and primarily used where the different modalities being fused are closely aligned temporally and/or by structure (e.g., P13). The risk associated with using these early fusion strategies is that one modality may dominate the resulting fused representation, making it more difficult to interpret.
- Hybrid fusion strategies (e.g., P18, P27) are new directions of research. The fusion techniques used in hybrid fusion strategies often use attention or residual learning to combine both space and time features within

the model, showing that these hybrid models produce an increased synergistic capacity for representing the modalities that were fused.

- Model-level fusion (e.g., W10) is an emerging trend in the development of unified or endpoint feature representations. Since their functional purpose is the same, upon model inception, their goal is to address the limits of either early or late fusion.

In general terms, the table indicates a trend away from decision-level fusion towards integrated architectural designs that incorporate both data and environment. Increased use of hybrid and model-level fusion indicates a realization that emotions and stress are complex, dimensional entities that cannot be accurately represented through simple concatenation or voting alone. However, since there is no consistent representation or fusion strategy for the various methods used to represent and fuse data, it appears that the field has not yet reached a unified methodology that would apply to all implementations and would allow for the comparison of results across studies.

TABLE V. FUSION STRATEGIES USED IN MULTIMODAL EMOTION PREDICTION STUDIES

Fusion Strategy	Description	Common Signal Combinations	Key Advantages	Reported Limitations
Feature level fusion	Features from multiple modalities are combined before model training	EEG + HRV, EEG + ECG, EEG + EDA	Captures cross-modal relationships early; often improves accuracy	High dimensionality; requires signal synchronization
Decision level fusion	Independent models are trained per modality and outputs are combined	EEG, HRV, ECG (separate models)	Robust to noisy or missing signals; flexible design	Limited interaction between modalities
Hybrid fusion	Combination of feature and decision level fusion	EEG + peripheral signals	Balances early integration and robustness	Increased model complexity
No fusion (single modality)	Emotion prediction using one physiological signal	EEG only, HRV only	Simpler architecture; lower computational cost	Lower robustness and prediction performance

Table V presents a comprehensive overview of the different methods that can be used for multimodal emotion detection research, including their design principles, combinations of signals, advantages and disadvantages. The most frequently used fusion method is feature-level fusion, as it allows for the early integration of multiple physiological feature types and usually provides improved prediction accuracy. Nevertheless, feature-level fusion has the disadvantage of using high-dimensional data requiring precise signal synchronization.

Decision-level fusion is based on combining multiple independent models, each trained on a single modality of data. This method is useful when one or more modalities may have corrupted or missing signal information, as decision-level fusion is more robust to these types of problems. In addition, this method of fusion allows for flexible model development but restricts the possible interaction between modalities. Two methods of fusion are hybrid fusion strategies, which combine features from both the feature-level and decision-level methods and the use of a single modality without any fusion methods. The use of hybrid strategies was explored to achieve a balance between the advantages of early data integration and robustness, especially for complex performance-critical tasks. Lastly, the use of single modality approaches has also been performed due to the lower computational costs and ease of development, even though these methods generally exhibit lower reliability and lower prediction accuracy than the fusion methods.

E. Performance Metrics and Evaluation Practices

The papers created for this review used a lot of different performance measures to evaluate the performance of the Emotion Detection Model (EDM). The performance measures have many different types of objectives and/or experimental design, emotion representation and process differences. Many studies reported their accuracy as the most frequently used performance measure for EDMs because many of the studies report accuracy larger than 95% and the studies using both discrete emotion categories and/or either stress detection are classified as using binary or classifying stress detection, respectively. These results indicate that the EDM was successful in being able to learn from physiological signals using the deep learning algorithms.

The classification accuracy measure is only one of the performance measures used for measuring EDM, but several papers did report measuring EDM using other performance measures such as: precision, recall, F1 score and confusion matrices. These paving processes will allow for more accurate measurement of the performance of the EDM, as they will allow for measuring the trade-off between false positives and false negatives. The importance of measuring accuracy beyond classifying the incorrect class would be more of an impact within performance-critical applications, as an example of classification would have consequences.

The studies included in Table VI show that there is a tremendous difference in how emotions are captured and evaluated. In most of the studies, classification accuracy was the primary outcome measured by the various studies; usually, reports included the use of other complementary metrics such as F1 score, precision, etc., although these metrics were used more frequently on multi-target emotion prediction tasks. However, several studies utilized dimensional models of emotion and therefore relied on regression-type evaluation measures (mean absolute error) for measuring prediction quality.

TABLE VI. PERFORMANCE METRICS ACROSS KEY STUDIES

Paper ID	Primary Metrics Used	Best Reported Results	Emotion Representation
P01	Accuracy	96.21%	7 Discrete Emotions
P02	Accuracy, F1-Score	99.0% (AMIGOS), 90.8% (DREAMER)	4-Quadrant (HVHA, HVLA, etc.)
P05	Accuracy, Precision, Recall, F1	99.45% (Valence), 99.68% (Liking)	Valence, Arousal, Liking
P13	Accuracy, AUC-ROC, Recall, F1	92.1% Accuracy, 0.96 AUC-ROC	Neutral, Tense, Concerned
P14	Accuracy, Specificity, F1	85.16% (using 8 channels)	4 Discrete Emotions
P17	Acc, Precision, Recall, F1, MCC, Kappa, AUC	96.77% Accuracy (KNN model)	Valence, Arousal
P19	Accuracy, Loss	71.61% (Early Fusion)	4-Quadrant Dimensional
P21	Accuracy, F1-Value	98.57% Accuracy, 98.73% F1	4 Emotional Categories
P25	Accuracy, Precision, Recall, F1, MAE	98.89% Accuracy, 0.05 MAE	5 Primary Emotions
P28	Accuracy, Macro-F1, Forecast Horizon	74.4% (Current), 65.5% (2-hr Forecast)	Positive vs. Negative
P29	Accuracy, Sensitivity, Specificity, Kappa	99.6% Accuracy (Augmented)	4 Games (Boring, Calm, etc.)
W10	Accuracy	87.70%	Baseline, Stress, Amusement

Overall, these studies have reported highly successful performance, but the large number of different types of metrics, validation types/protocols and types of emotion capture used limits direct performance comparison across studies, accentuating the need for standardization of emotion prediction evaluation mechanisms/devices. Consequently, the studies in the table indicate there are major issues related to the standardization of emotion prediction evaluations. Therefore, it is vital that any conclusions drawn based on existing research are done so with caution considering the discrepancies found among studies' metrics, validation methods and datasets. Caution should be taken in drawing any inferences from reported accuracy values with respect to the applicability of a studied emotion prediction model to real-world application where performance is critically important.

The overall evidence shows that although emotion prediction systems based on deep learning techniques have shown some success, there needs to be much more focus on using standard evaluation metrics, independent verification of

subject matter, and transparent reporting to be able to reproduce the results obtained and make them practically usable.

F. Identified Research Gap and Challenge

As shown throughout the course of this review, there are many significant shortcomings in both the research surrounding deep learning-based emotion prediction with physiological measures and also in the methodology used to conduct these studies, especially when predicting emotions in high-performance task environments. There are many reports of successful results across multiple studies; however, due to limitations in the methods of fusion, evaluation, and overall experimental design, it is difficult to generalize or practically deploy the current approaches being used to predict emotions.

One major gap in the research is the lack of standardized multimodal fusion frameworks. The majority of the studies reviewed use feature-level fusion (as highlighted in Table V), which offers great advantages in terms of capturing cross-modal relationships and increasing predictive accuracy; however, feature-level fusion also produces high-dimensional feature sets and requires strict synchronization requirements, making it difficult to scale and use in real-time. Decision-level and hybrid methods of fusion, while providing better robust solutions, are infrequently studied; therefore, further research is required to examine adaptive and dynamic fusion methods for balancing accuracy, robustness and computational efficiency.

A major obstacle has been the evaluation of performance measurements. As shown in Table VI, there are numerous differences in how various studies performed evaluations of emotional expression and models used to validate the evaluations, including variations in evaluating performance metrics, emotion representation schemes and validation protocols for each study. Even though most researchers report the accuracy of classification models as their performance metric, using accuracy as the sole metric to compare performances can inappropriately mask problems with the level of class (dis)similarity (imbalance) and generalizability of the model across sample populations. Additionally, the inconsistency of how commonly used complementary performance metrics such as F1 Score, AUC ROC and regression-based errors are reported among various research studies further compounds the challenge of establishing valid cross-study comparisons or identifying models that will truly generalize.

Existing research in the area of emotion recognition has also been constrained by limitations related to datasets and experimental designs. Many researchers use controlled laboratory conditions and subject dependent validation schemes to validate their findings. Therefore, those controlled laboratory conditions or subject-based validation of emotion recognition models may dramatically inflate the performance measures that have been reported and reduce the level of ecological validity associated with those findings. Although there have been attempts to use multi-modal sensing modalities and wearable sensing, there continues to be a significant shortage of real-world performance scenarios used to evaluate the accuracy of emotion detection in the real world. These circumstances demonstrate an urgent need for larger, more heterogeneous datasets with subject-independent validation protocols, because

without these, researchers will be unable to develop models of emotion recognition that could generalize across individuals and situations.

Hybrid and attention-based architectures are getting more popular; however, they have yet to show widespread usage. The focus of future research should be on lightweight, interpretable deep learning that can work in a resource-constrained environment while providing illuminating signals about relevance and how decisions are made. These advances will be of vital importance where the safety of the system is.

Future studies must develop standards for how we evaluate emotions, develop systems that can integrate input from multiple sources, and develop experiments that reflect real-world conditions. Addressing these issues will be necessary to enable emotion prediction systems to be used reliably outside of a laboratory environment where performance is critical.

V. CROSS-DOMAIN PERSPECTIVES

The use of physiological signals and deep learning to predict emotions is physically derived from an interdisciplinary base of knowledge, including affective computing, biomedical signal processing, numerical modeling and advanced data analysis methods. The previous research into using the EEG for classification of emotions was evidence of what can be possible (i.e., extracting patterns in emotion from the signals of the brain) and support for the use of EEG as one of the primary modalities for detecting affective states [42]. These early findings will continue to support the application of deep learning to understand emotion-aware systems by validating the physiological underpinnings of those systems.

Research into neuromarketing and affective analytics further illustrates how physiology-based inferences about emotion are limited and that they are modified by the context in which they occurred. This supports the need for studies that utilize formal multimodal, context-aware and transparent approaches, particularly in performance-critical environments in which emotional states are closely linked to cognition and decision-making.

Methodological advances in numerical modeling, such as modified Euler and harmonic-polygon-based schemes, emphasize accuracy, stability and robustness in solving non-linear systems [43, 44]. Although originating in computational mathematics, these principles parallel current challenges in deep learning-based physiological signal modeling, where noise, non-stationarity and inter-subject variability can degrade prediction reliability.

Related work in computer-aided medical analysis and image-based segmentation demonstrates the importance of structured feature representation and explainable model design for high-stakes applications [45, 46]. Likewise, studies on feature selection in sentiment and health-related analytics highlight trade-offs between model complexity, interpretability and generalizability [47]. Collectively, these cross-domain insights reinforce the importance of developing lightweight, interpretable and robust deep learning frameworks for emotion prediction that are suitable for deployment in real-world performance-critical tasks.

VI. CONCLUSION

This systematic review aimed to examine how physiological signals and deep learning models have been applied to emotion prediction in performance-critical tasks, with particular emphasis on physiological signal modalities, deep learning architectures, multimodal fusion strategies, evaluation practices, and methodological limitations. Guided by the research questions, this review synthesized recent studies published between 2021 and 2025 to identify the dominant methodological trends, key research gaps, and implications for building robust emotion prediction systems in real-world performance-critical environments.

The review indicates that EEG remains the dominant physiological modality for emotion prediction, while peripheral signals such as HRV, ECG, and EDA are frequently incorporated to strengthen multimodal modeling. Across the reviewed studies, deep learning approaches, particularly CNN, LSTM, and hybrid architectures, demonstrated strong capability in modelling complex temporal and cross-signal patterns. However, the broader synthesis suggests that predictive effectiveness is not determined by model architecture alone; it also depends on the quality of signal representation, the realism of the task context, and the consistency of evaluation practices.

The review further shows that multimodal emotion prediction generally offers greater potential than single-modality approaches, especially for capturing both central and autonomic responses in demanding environments. Nevertheless, the field remains constrained by several recurring limitations, including inconsistent fusion strategies, limited ecological validity, lack of subject-independent validation, and non-standardized performance evaluation methods. These issues make cross-study comparison difficult and continue to hinder the translation of laboratory findings into reliable real-world emotion prediction systems.

To advance this field, future research should prioritize standardized evaluation frameworks, robust and adaptive fusion strategies, and ecologically valid experimental designs that better reflect real-world performance-critical conditions. Greater emphasis should also be placed on subject-independent validation, interpretable lightweight deep learning models, and deployment-oriented testing in realistic operational settings. These directions are essential for improving the reliability, generalizability, and practical usability of emotion-aware systems.

In conclusion, this review provides a structured synthesis of recent literature on emotion prediction in performance-critical tasks and highlights the need to align physiological sensing, deep learning design, and evaluation methodology more closely with real-world operational demands. By identifying dominant trends, methodological weaknesses, and future research priorities, this study provides a clearer foundation for the development of more reliable and deployable emotion prediction systems.

ACKNOWLEDGMENT

The authors gratefully acknowledge the National Defence University of Malaysia (NDUM) for providing the facilities and

academic support necessary for this study. We also wish to express our sincere appreciation to Prisma Khas Sdn Bhd for their valuable support and collaboration throughout the project.

REFERENCES

- [1] Zitouni, M. S., Park, C. Y., Lee, U., Hadjileontiadis, L. J., & Khandoker, A. (2023). LSTM-Modeling of Emotion Recognition Using Peripheral Physiological Signals in Naturalistic Conversations. *IEEE Journal of Biomedical and Health Informatics*, 27(2), 912–923. <https://doi.org/10.1109/JBHI.2022.3225330>
- [2] Liao, Y., Gao, Y., Wang, F., Zhang, L., Xu, Z., & Wu, Y. (2025). Emotion recognition with multiple physiological parameters based on ensemble learning. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-96616-0>
- [3] Wang, Z., & Wang, Y. (2025). Emotion recognition based on multimodal physiological electrical signals. *Frontiers in Neuroscience*, 19. <https://doi.org/10.3389/fnins.2025.1512799>
- [4] Dar, M. N., Akram, M. U., Subhani, A. R., Khawaja, S. G., Reyes-Aldasoro, C. C., & Gul, S. (2024). Insights from EEG analysis of evoked memory recalls using deep learning for emotion charting. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-61832-7>
- [5] Du, X., Ma, C., Zhang, G., Li, J., Lai, Y. K., Zhao, G., Deng, X., Liu, Y. J., & Wang, H. (2022). An Efficient LSTM Network for Emotion Recognition From Multichannel EEG Signals. *IEEE Transactions on Affective Computing*, 13(3), 1528–1540. <https://doi.org/10.1109/TAFFC.2020.3013711>
- [6] Singh, K., Ahirwal, M. K., & Pandey, M. (2025). Selected channel based multiclass emotion classification from wearable human brain EEG signal. *Measurement: Sensors*, 39. <https://doi.org/10.1016/j.measen.2025.101874>
- [7] Saini, D. K. J. B., Shieh, C. S., Sankpal, L. J., Mehrotra, M., Bhosale, K. S., & Raut, Y. (2025). Deep Learning-based Optimized Model for Emotional Psychological Disorder Activities Identification in Smart Healthcare System. *Journal of Research, Innovation and Technologies*, 4(2), 143–157. [https://doi.org/10.57017/jorit.v4.2\(8\).02](https://doi.org/10.57017/jorit.v4.2(8).02)
- [8] Kumar, A., Sharma, K., & Sharma, A. (2021). Hierarchical deep neural network for mental stress state detection using IoT based biomarkers. *Pattern Recognition Letters*, 145, 81–87. <https://doi.org/10.1016/j.patrec.2021.01.030>
- [9] Santamaria-Granados, L., Mendoza-Moreno, J. F., Chantre-Astaiza, A., Munoz-Organero, M., & Ramirez-Gonzalez, G. (2021). Tourist experiences recommender system based on emotion recognition with wearable data. *Sensors*, 21(23). <https://doi.org/10.3390/s21237854>
- [10] Zhang, J., Shi, Y., Wang, C., Cao, C., Zhang, C., Ji, L., Cheng, J., & Wu, F. (2021). Preshooting Electroencephalographic Activity of Professional Shooters in a Competitive State. *Computational Intelligence and Neuroscience*, 2021. <https://doi.org/10.1155/2021/6639865>
- [11] Wang, Y., Sun, Y., Fang, L., & Zhang, C. (2022). LEVERAGING SPARSE CODING FOR EEG BASED EMOTION RECOGNITION IN SHOOTING. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2022-May*, 1421–1425. <https://doi.org/10.1109/ICASSP43922.2022.9747506>
- [12] Horvath, E., Kovacs, M. T., Toth, D., & Toth, L. (2022). A study of the relationship between anxiety, cognitive emotion regulation and heart rate variability in athletes. *Journal of Physical Education and Sport*, 22(2), 528–534. <https://doi.org/10.7752/jpes.2022.02066>
- [13] Li, P., Lu, Q., Wu, Q., Liu, X., & Wu, Y. (2021). What Makes an Elite Shooter and Archer? The Critical Role of Interoceptive Attention. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.666568>
- [14] Ahmad, W. N. W., Adib, M. A. H. B. M., Nashir, I. M., Txi, M. R. S., Salleh, F. N. M., & Tang, J. R. (2023). Anxiety status of junior archers in COVID-19 during training isolation period towards the shooting performance. *International Journal of Public Health Science*, 12(1), 361–370. <https://doi.org/10.11591/ijphs.v12i1.21982>
- [15] Xiaohu, W., Ren, Y., Luo, Z., He, W., Hong, J., & Huang, Y. (2023). Deep learning-based EEG emotion recognition: Current trends and future perspectives. In *Frontiers in Psychology* (Vol. 14). Frontiers Media S.A. <https://doi.org/10.3389/fpsyg.2023.1126994>
- [16] Gonzalez-Vazquez, J. J., Bemat, L., Ramon, J. L., Morell, V., & Ubeda, A. (2024). A Deep Learning Approach to Estimate Multi-Level Mental Stress from EEG Using Serious Games. *IEEE Journal of Biomedical and Health Informatics*, 28(7), 3965–3972. <https://doi.org/10.1109/JBHI.2024.3395548>
- [17] Fontes, L., Machado, P., Vinkemeier, D., Yahaya, S., Bird, J. J., & Ihianle, I. K. (2024). Enhancing Stress Detection: A Comprehensive Approach through rPPG Analysis and Deep Learning Techniques. *Sensors*, 24(4). <https://doi.org/10.3390/s24041096>
- [18] Aldawsari, H., Al-Ahmedi, S., & Muhammad, F. (2023). Optimizing 1D-CNN-Based Emotion Recognition Process through Channel and Feature Selection from EEG Signals. *Diagnostics*, 13(16). <https://doi.org/10.3390/diagnostics13162624>
- [19] Singh, P., Patel, S., Mendoza, R., Cruz, I., Martinez, A., & Kim, N. (n.d.). *Improving Psychological State Prediction Using Deep Learning Techniques*. <https://doi.org/10.13140/RG.2.2.21741.70888>
- [20] Aguiñaga, A. R., Delgado, L. M., López-López, V. R., & Téllez, A. C. (2022). EEG-Based Emotion Recognition Using Deep Learning and M3GP. *Applied Sciences (Switzerland)*, 12(5). <https://doi.org/10.3390/app12052527>
- [21] Pan, T., Ye, Y., Cai, H., Huang, S., Yang, Y., & Wang, G. (2023). Multimodal Physiological Signals Fusion for Online Emotion Recognition. *MM 2023 - Proceedings of the 31st ACM International Conference on Multimedia*, 5879–5888. <https://doi.org/10.1145/3581783.3612555>
- [22] Pillalamari, R., & Shanmugam, U. (2025). A review on EEG-based multimodal learning for emotion recognition. *Artificial Intelligence Review*, 58(5). <https://doi.org/10.1007/s10462-025-11126-9>
- [23] Abinaya, G., & Dinakaran, K. (2025). ACXNet hybrid deep learning model for cross task mental workload estimation using EEG neural manifolds. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-19144-x>
- [24] Ari, B., Siddique, K., Alcin, O. F., Aslan, M., Sengur, A., & Mehmood, R. M. (2022). Wavelet ELM-AE Based Data Augmentation and Deep Learning for Efficient Emotion Recognition Using EEG Recordings. *IEEE Access*, 10, 72171–72181. <https://doi.org/10.1109/ACCESS.2022.3181887>
- [25] Dar, M. N., Akram, M. U., Khawaja, S. G., & Pujari, A. N. (2020). Cnn and lstm-based emotion charting using physiological signals. *Sensors (Switzerland)*, 20(16), 1–26. <https://doi.org/10.3390/s20164551>
- [26] Algami, M., Saeed, F., Al-Hadhrani, T., Ghabban, F., & Al-Sarem, M. (2022). Deep Learning-Based Approach for Emotion Recognition Using Electroencephalography (EEG) Signals Using Bi-Directional Long Short-Term Memory (Bi-LSTM). *Sensors*, 22(8). <https://doi.org/10.3390/s22082976>
- [27] Gutiérrez-Martin, L., López-Ongil, C., & Miranda-Calero, J. A. (2026). DeepBind: An End-to-End Fear Detection System Optimized for Extreme-Edge Deployment. *IEEE Journal of Biomedical and Health Informatics*, 30(1), 688–699. <https://doi.org/10.1109/JBHI.2025.3587961>
- [28] Ezilarasan, M. R., & Leung, M. F. (2024). An Efficient EEG Signal Analysis for Emotion Recognition Using FPGA. *Information (Switzerland)*, 15(6). <https://doi.org/10.3390/info15060301>
- [29] Moise, G., Dragomir, E. G., Şchiopu, D., & Iancu, L. A. (2024). Towards Integrating Automatic Emotion Recognition in Education: A Deep Learning Model Based on 5 EEG Channels. *International Journal of Computational Intelligence Systems*, 17(1). <https://doi.org/10.1007/s44196-024-00638-x>
- [30] Olashore, T. D., Boyinbode, O. K., Daramola, O. A., & Kinga, M. T. (2025). Development of Real-Time Electroencephalography-Based Emotion Recognition Models Using Machine Learning Techniques. *Cureus Journal of Computer Science*. <https://doi.org/10.7759/s44389-025-06161-0>
- [31] Mane, S. A. M., & Shinde, A. (2023). StressNet: Hybrid model of LSTM and CNN for stress detection from electroencephalogram signal (EEG). *Results in Control and Optimization*, 11. <https://doi.org/10.1016/j.rico.2023.100231>
- [32] Nakisa, B., Rastgoo, M. N., Rakotonirainy, A., Maire, F., & Chandran, V. (2020). Automatic Emotion Recognition Using Temporal Multimodal

- Deep Learning. *IEEE Access*, 8, 225463–225474. <https://doi.org/10.1109/ACCESS.2020.3027026>
- [33] Wang, C. W., Zhang, Y., Osawa, K., Nakagawa, K., & Tanaka, E. (2025). Development of emotion recognition for rehabilitation feedback system using wavelet transform and LSTM. *ROBOMECH Journal*, 12(1). <https://doi.org/10.1186/s40648-025-00322-5>
- [34] Ying, G., & Hao, C. (2023). A novel one-dimensional convolutional neural network-based method for emotion recognition of electric power industry workers. *Energy Reports*, 9, 763–771. <https://doi.org/10.1016/j.egy.2023.04.297>
- [35] Zhao, Y., Guo, M., Chen, X., Sun, J., & Qiu, J. (2024). Attention-Based CNN Fusion Model for Emotion Recognition during Walking Using Discrete Wavelet Transform on EEG and Inertial Signals. *Big Data Mining and Analytics*, 7(1), 188–204. <https://doi.org/10.26599/BDMA.2023.9020018>
- [36] Li, H., & Sun, H. (2024). Emotional analysis of joint sports quality expansion tasks based on multi-modal feature fusion. *Systems and Soft Computing*, 6. <https://doi.org/10.1016/j.sasc.2024.200092>
- [37] Alam, M. M., Dini, M. A., Kim, D. S., & Jun, T. (2025). TMNet: Transformer-fused multimodal framework for emotion recognition via EEG and speech. *ICT Express*, 11(4), 657–665. <https://doi.org/10.1016/j.icte.2025.04.007>
- [38] Zhou, S., Zhang, N., Duan, Q., Liu, X., Xiao, J., Wang, L., & Yang, J. (2024). Monitoring and Analyzing Driver Physiological States Based on Automotive Electronic Identification and Multimodal Biometric Recognition Methods. *Algorithms*, 17(12). <https://doi.org/10.3390/a17120547>
- [39] Li, Z., Ma, W., Zhang, J., Li, Y., Wang, L., Hao, Y., & Li, X. (2024). DSRP: A Database for Stress Reduction Using Physiological Signals. *IEEE Access*, 12, 135089–135102. <https://doi.org/10.1109/ACCESS.2024.3454090>
- [40] Kim, W., Kutsuzawa, G., & Maruyama, M. (2025). Emotion recognition and forecasting from wearable data via cluster-guided attention with cross-species pretraining. *Intelligent Systems with Applications*, 27. <https://doi.org/10.1016/j.iswa.2025.200560>
- [41] Daly, I., Williams, D., Malik, A., Weaver, J., Kirke, A., Hwang, F., Miranda, E., & Nasuto, S. J. (2020). Personalised, Multi-Modal, Affective State Detection for Hybrid Brain-Computer Music Interfacing. *IEEE Transactions on Affective Computing*, 11(1), 111–124. <https://doi.org/10.1109/TAFFC.2018.2801811>
- [42] V. S. Mohan, M. F. M. Amran, Y. H. Yahaya, N. M. M. Yusop, T. M. T. Sembok M. A. Ahmad, “Emotion classifications in electroencephalogram (EEG) signals,” *International Journal of Recent Technology and Engineering*, vol. 8, no. 3, pp. 2736–2740, doi: 10.35940/ijrte.B2650.098319, 2019.
- [43] S. Nooraida, N. M. M. Yusop, M. S. Anis, A. M. M. Fahmi, A. W. F. Syarul A. M. Iliana, “Cube polygon: A new modified Euler method to improve accuracy of ordinary differential equation (ODE),” *Journal of Physics: Conference Series*, vol. 1532, no. 1, p. 012020, doi: 10.1088/1742-6596/1532/1/012020, 2020.
- [44] N. M. M. Yusop, M. K. Hasan, M. Wook, M. F. Mohamad Amran S. R. Ahmad, “A new Euler scheme based on harmonic-polygon approach for solving first order ordinary differential equation,” *AIP Conference Proceedings*, vol. 1891, no. 1, p. 020102, doi: 10.1063/1.5005435, 2017.
- [45] N. Awang, R. Sulaiman, A. Shapi'i, A. H. A. Rashid, M. F. M. Amran S. Osman, “A comparative study of computer aided system preoperative planning for high tibial osteotomy,” in *Advances in Visual Informatics, LNCS*, vol. 9429. Cham, Switzerland: Springer, pp. 189–198, doi: 10.1007/978-3-319-25939-0_17, 2015.
- [46] R. Sulaiman, D. Marlinawati M. F. Mohamad Amran, “COMSeg technique for MRI knee cartilage segmentation,” *International Review of Applied Sciences and Engineering*, vol. 10, no. 2, pp. 147–155, doi: 10.1556/1848.2019.0018, 2019.
- [47] S. R. Ahmad, N. M. M. Yusop, A. M. Asri M. F. M. Amran, “A review of feature selection algorithms in sentiment analysis for drug reviews,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 12, doi: 10.14569/IJACSA.2021.0121217, 2021.