

A Personal Productivity Digital Twin: A Framework for Modeling, Predicting, and Optimizing Individual Work Behavior

¹Samiksha Patil, ²Shrutika Chaulkar, ³Suraj Patil, ⁴Suyash Madhvi

¹²³⁴Department of Information Technology, Pillai College of Arts, Commerce and Science, Navi Mumbai, India

Abstract-Digital twin (DT) technology has evolved from its origins in industrial product-lifecycle management toward representations of human beings, giving rise to human, personal, and cognitive digital twins. In parallel, ubiquitous-computing and organizational-psychology research has quantified how task switching, meeting load, and fragmented attention erode individual productivity, while smartphone- and wearable-sensing studies have shown that passively collected behavioral signals can predict productivity- and wellbeing-related outcomes. These two lines of research remain largely disconnected: human digital twin work has concentrated on healthcare and industrial contexts, while personal productivity tools typically issue isolated, session-based recommendations rather than maintaining a persistent, continuously updated model of the individual. This paper proposes a Personal Productivity Digital Twin (PPDT), a conceptual and methodological framework in which an individual's tasks, schedule, workload, and attention patterns are represented as a dynamically updated digital state that is used to predict near-term productivity outcomes and to generate constrained, optimization-driven scheduling recommendations refined through a human-in-the-loop feedback cycle. We review relevant literature on digital twins, human and cognitive digital twins, and productivity sensing; define a layered PPDT architecture; formalize a productivity state vector together with prediction and optimization objectives; and present an operating-cycle algorithm for the twin. Because no single publicly available dataset captures the full multidimensional PPDT state, the framework is presented and evaluated at a conceptual level; we explicitly distinguish literature-reported findings from proposed contributions and outline the experimental, privacy, and ethical work required before empirical deployment.

Keywords - digital twin; human digital twin; personal productivity; behavioral modeling; task scheduling optimization; predictive analytics; privacy-preserving computing

I. INTRODUCTION

Knowledge work today is increasingly mediated by digital tools: calendars, task managers, messaging platforms, and document-editing applications that together generate continuous, fine-grained traces of how individuals spend their time and attention. At the same time, decades of research in human-computer interaction and organizational psychology have documented that this digitally mediated work is highly fragmented. Information workers switch between activities frequently, and interruptions carry a measurable cost in resumption time and error rate [10], [11]. These findings suggest that individual productivity is not solely a function of effort or skill, but also of how work is scheduled, sequenced, and interrupted over time.

Separately, digital twin (DT) technology has matured from a product-lifecycle-management concept for manufactured artifacts [1] into a general paradigm for maintaining a continuously updated virtual representation of a physical counterpart that is used for monitoring, prediction, and decision support [2]. Over the past decade this paradigm has been extended from machines and processes to human beings, giving rise to the notions of the human digital twin (HDT) [3], [4], the personal digital twin (PDT) [5], and the cognitive digital twin [6], which represent an individual's physiological, behavioral, or cognitive state rather than a manufactured object. Most reported applications of human-centered digital twins, however, remain concentrated in healthcare, rehabilitation, and industrial human-in-the-loop settings [3]–[6].

Comparatively little published work applies the digital-twin paradigm specifically to everyday personal productivity, that is, the management of tasks, schedules, focus, and workload in professional and academic life.

A parallel and largely independent body of work has used smartphone and wearable sensing to monitor and predict productivity- and wellbeing-related outcomes. Passive smartphone sensing has been used to correlate behavioral signals with academic performance and wellbeing over an academic term [9], and dedicated productivity-monitoring studies have shown that features derived from smartphone usage and ambient sensing can be used to classify changes in a user's productivity state with conventional machine learning classifiers [8]. These systems, however, typically operate as session-based monitors or classifiers: they estimate or report a productivity-related outcome but do not maintain a persistent, multidimensional digital representation of the individual that is continuously updated, used for forward prediction, and coupled to an explicit optimization or recommendation process.

This gap motivates the central research problem addressed in this paper: how can a digital twin representing an individual's tasks, schedule, workload, and attention patterns be constructed, kept synchronized with the person's ongoing behavior, and used to analyze, predict, and optimize personal productivity in a way that is scientifically grounded, technically well-specified, and attentive to privacy and autonomy? We refer to such a system as a Personal Productivity Digital Twin (PPDT).

The contribution of this paper is threefold. First, we bring together literature from three otherwise separate areas, namely digital twin architectures, personal and cognitive digital twins, and productivity/attention sensing, to motivate and scope the PPDT concept and identify a concrete research gap. Second, we propose a layered PPDT architecture, a mathematical formulation of the productivity state, and a prediction, optimization, and feedback algorithm that operationalizes the architecture. Third, we critically examine the availability of data for empirically validating such a system, distinguish literature-reported findings from proposed and conceptual contributions throughout, and discuss the privacy, ethical, and methodological limitations that must be addressed before a PPDT could be responsibly deployed.

The remainder of this paper is organized as follows. Section II reviews the literature on digital twins, human and cognitive digital twins, productivity and attention research, predictive modeling, and personalization/optimization techniques, and synthesizes this literature into a research gap. Section III presents the proposed PPDT methodology, including its architecture, data considerations, feature engineering, state representation, prediction approach, optimization formulation, and feedback mechanism. Section IV discusses the framework at a conceptual and literature-grounded level, since no experiments have been conducted, and presents evaluation metrics and a comparative analysis against existing approaches. Section V concludes the paper and outlines directions for future, empirically grounded work

II. LITERATURE REVIEW

A. Origins and Evolution of Digital Twins

The digital twin concept originated in industrial product-lifecycle management. Grieves first sketched the idea in 2002 as a "Mirrored Spaces Model," consisting of a physical space, a virtual space, and a data link between them, later formalizing it as a set of virtual information constructs that fully describes a physical manufactured product across its lifecycle [1]. The term "digital twin" itself was subsequently adopted by NASA in the context of vehicle-health management and simulation. Because early definitions were product-centric, more recent efforts have sought a discipline-neutral definition; the U.S. National Academies of Sciences, Engineering, and Medicine define a digital twin as a set of virtual information constructs that mimics the structure, context, and behavior of a natural, engineered, or social system, is dynamically updated with data from its physical twin, has a predictive capability, and informs decisions that realize value, emphasizing bidirectional interaction between the physical and virtual instances as the defining property of a digital twin [2]. This broader, system-agnostic definition is what makes it meaningful to speak of a digital twin of a person rather than only of a machine.



Fig. 1. The digital twin concept applied to an individual: a physical person and a virtual, continuously synchronized productivity state.

B. Human, Personal, and Cognitive Digital Twins

As digital twin research matured, several extensions targeted human beings as the physical counterpart. The human digital twin (HDT) is defined in a recent survey as a digital replica of a physical-world human, continuously updated from sensor

and record data and used to provide feedback such as diagnoses, predictions, or other suggestions back to the person; the same survey notes that human attributes relevant to an HDT can be grouped into physical, physiological, cognitive-performance, personality, emotional, ethical, and behavioral categories, reflecting the far greater complexity of modeling a human compared to a machine [3]. A complementary survey emphasizes that HDTs use a bidirectional data thread for ongoing synchronization with a person's state and for providing feedback that directly or indirectly influences the person, while also cataloguing open challenges in definition, design, and trust [4].

Within healthcare, the personal digital twin (PDT) concept has been proposed as an enhanced digital twin with “actionable insight” capability intended to support more accurate clinical decision-making; one influential formulation defines the PDT across mental, physical, social, and biological dimensions of an individual's life rather than a single biological subsystem, and situates it within a reference architecture that combines digital twin, blockchain, and AI technologies for personalized healthcare use cases such as COVID-19 monitoring [5]. Separately, the notion of a cognitive digital twin (CDT) extends a digital twin with reasoning and problem-solving capability that goes beyond passive data mirroring toward handling unforeseen situations; in a production-research context, a cognitive digital twin has been characterized as incorporating expert and problem-solving knowledge, in contrast to a purely data-driven “hybrid twin” [6]. Collectively, this literature establishes that human-centered digital twins are an active and rapidly diversifying research area, but also confirms that the applications reported to date are concentrated in clinical, rehabilitation, and cognitive-modeling domains rather than everyday task and time management.

C. Enabling Architectures: IoT- and AI-Enabled Digital Twins

Because a digital twin must be continuously synchronized with its physical counterpart, its practical realization depends on Internet-of-Things (IoT) sensing, cloud/edge computing, and machine learning. A representative architecture for a Web-of-Things-based digital twin integrates heterogeneous sensor streams, a virtual representation layer, and a model-based reinforcement-learning controller that can be trained largely in simulation before being deployed against the real system, addressing the common IoT constraint that large quantities of real interaction data are costly to obtain [7]. Broader surveys of digital twin architectures for the “Internet of Digital Twins” further describe layered designs spanning data acquisition, synchronization, modeling, and decision/visualization layers, and catalogue the security and privacy threats that arise once digital twins encode sensitive information about their physical counterpart [18]. These architectural patterns, namely a sensing layer, a synchronization/state layer, a model layer, and a decision/interaction layer, are directly relevant to a personal productivity twin, whose “physical counterpart” is a person rather than a machine, and whose sensing sources (calendars, task managers, device-usage logs) are considerably more privacy-sensitive than industrial telemetry.

D. Personal Productivity Measurement and Behavioral Sensing

A separate line of ubiquitous-computing research has investigated whether personal productivity can be inferred from passively collected behavioral data. The StudentLife study instrumented 48 college students with continuous smartphone sensing over a ten-week academic term and found that automatically sensed conversation, physical activity, mobility, and sleep correlated significantly with mental well-being and academic outcomes, with reported stress rising and sleep, activity, and positive affect declining as term-time workload increased; the study also demonstrated that a regularized linear regression model could predict a student's cumulative GPA from these passively sensed features [9]. More directly relevant to workplace productivity, a dedicated smartphone-based productivity-monitoring study collected behavioral features from smartphone sensors and combined them with two benchmark datasets (a purpose-built productivity logging application and a smart-home activity dataset) to train logistic regression, support-vector-machine, and k-nearest-neighbor classifiers that detected changes in a user's productivity state, reporting that all three classifiers achieved good predictive performance on both datasets [8]. Complementary wearable-sensing research on stress has shown that self-supervised pretraining on a user's own baseline physiological dynamics can substantially reduce the amount of labeled data needed to personalize a predictive model, which is directly relevant to the “cold start” problem that any newly instantiated personal twin would face for a user with little historical data [17]. Together, these studies establish that behavioral sensing can support productivity- and wellbeing-related prediction, but each system reported estimates a state or outcome without maintaining a persistent, forward-looking, and optimization-linked representation of the individual over time.

E. Attention Fragmentation and Context Switching

Any model of personal productivity must account for the well-documented cost of interrupted and fragmented work. Field observation of information workers found that working spheres were frequently interrupted and that a majority of

interrupted activities were not immediately resumed. This established that fragmentation, measured as both the length of uninterrupted activity and the frequency of interruption, is a pervasive feature of everyday knowledge work rather than an occasional anomaly [10]. A controlled followup study found that when interruptions occur, people often compensate by working faster to finish the interrupted task in a comparable amount of time, but at the cost of measurably higher subjective stress, frustration, and perceived time pressure, and that individual differences in personality traits such as openness to experience predict how disruptive an interruption is for a given person [11]. These findings jointly imply that a productivity-relevant state representation should include not only task and calendar attributes but also a measure of context-switching frequency and its accumulated disruption cost, and that the effect of context switching is likely to be person-specific rather than uniform across individuals.

F. Predictive Modeling Techniques Applicable to Behavioral Time Series

Several families of supervised and sequence models are candidates for forecasting a personal productivity state from historical behavioral features. Ensemble tree methods such as random forests aggregate many decorrelated decision trees to reduce variance and are well suited to tabular, mixed-type behavioral features [19], while gradient-boosted tree implementations such as XGBoost have become a standard choice for structured, tabular prediction problems because of their regularization and scalability properties [20]. For genuinely sequential dependencies, such as how several consecutive days of workload and sleep jointly affect tomorrow's focus, recurrent architectures such as long short-term memory (LSTM) networks are designed to retain and forget information across long input sequences through gated memory cells [21]. Transformer architectures instead use self-attention to model long-range dependencies without the strictly sequential recurrence of LSTMs, at the cost of typically requiring more training data [22]. No single one of these methods is universally preferable; the appropriate choice depends on the amount of historical data available for a given individual, the granularity of the sensing stream, and whether the modeled dependencies are primarily cross-sectional (tabular ensembles) or temporal (recurrent/attention-based models), a trade-off we return to in Section III-F.

G. Personalization via Reinforcement Learning and Constrained Optimization

Beyond predicting a productivity state, a PPDT must recommend actions, such as task orderings or schedule changes, that are expected to improve that state. This is fundamentally a sequential decision or constrained optimization problem. Reinforcement learning (RL) has been applied to personalization broadly, including to personalized learning-sequence recommendation, where the recommendation problem is formulated as a Markov decision process in which a policy balances exploiting known-good recommendations against exploring potentially better ones [12]. A systematic literature review of RL for personalization more broadly finds that RL has increasingly been applied to systems that interact with humans, but also that comparative and realistic evaluations remain relatively rare and that most published algorithms are used only once, indicating that the field has not yet converged on standard, reusable evaluation practice [13]. In parallel, classical and metaheuristic optimization methods have been applied directly to personal and calendar-level scheduling: a compromise-programming formulation combined with a genetic algorithm has been used to solve a multi-objective task-assignment problem with real scheduling data [15], and a tiered smart-calendar system combining branch-and-bound, Hungarian-algorithm, and genetic-algorithm solvers has been proposed specifically to optimize the timing of personal calendar events subject to user preferences and constraints [16]. These results indicate that both learning-based (RL) and search-based (genetic/exact) optimization techniques are viable candidates for the recommendation layer of a PPDT, with RL better suited to settings where an interaction policy can be learned online and search-based optimization better suited to settings where an explicit, interpretable multi-objective schedule must be produced from a known task and constraint set.

H. Privacy-Preserving Processing of Personal Data

Because a PPDT would continuously process highly personal data, including calendars, application usage, and potentially physiological signals, its design must address the privacy and security risks that are already documented for digital twins of physical systems, only more acutely. Surveys of digital twin security note that once a digital twin encodes detailed information about its physical counterpart, the twin itself becomes an attractive and sensitive target, and identify federated learning as a promising way to train digital-twin models across distributed devices without centralizing raw data [18]. Concrete federated-learning schemes for digital-twin applications have been proposed that keep raw data on the local device and instead exchange encrypted or aggregated model updates, using techniques such as homomorphic encryption to protect the parameters exchanged between local devices and an aggregating server [14]. These privacy-preserving learning techniques are directly transferable to a PPDT, in which the "local device" would be the individual's own phone or computer rather than an industrial edge node.

I. Synthesis and Research Gap

Table IV, presented after the methodology section for ease of comparison, summarizes how existing lines of work relate to the properties required of a Personal Productivity Digital Twin. Industrial digital twins provide mature architectural patterns for state synchronization and feedback but are not designed to represent individual human behavior [1], [2]. Human, personal, and cognitive digital twin research explicitly targets individuals but has concentrated on healthcare, rehabilitation, and cognitive-modeling applications, generally without a task- and schedule-centric productivity focus [3]–[6]. Smartphone- and wearable-based productivity and wellbeing sensing systems demonstrate that behavioral signals are predictive of productivity-relevant outcomes, but they typically operate as classifiers or monitors rather than as persistent, updateable digital twins, and they do not couple prediction to an explicit optimization or recommendation engine [8], [9], [17]. Reinforcement-learning and search-based personalization/scheduling research supplies candidate optimization engines but is rarely integrated with a continuously maintained individual state representation [12], [13], [15], [16]. Limited research to date has therefore combined (i) a persistent, multidimensional digital representation of an individual's productivity-relevant state, (ii) forward prediction of that state, and (iii) constrained, personalized optimization of schedules or task assignments, within a single closed feedback loop centered on everyday personal productivity rather than healthcare or industrial process control. This is the gap the proposed Personal Productivity Digital Twin framework is intended to address.

III. PROPOSED METHODOLOGY

This section presents the proposed Personal Productivity Digital Twin (PPDT) framework. Unless a specific literature citation is given, the architecture, state representation, and algorithm described below constitute the proposed contribution of this paper rather than an established or previously validated system, and are presented as a conceptual framework requiring empirical validation (Section IV).

A. System Architecture

The proposed PPDT follows a layered architecture adapted from general digital-twin and IoT-enabled digital-twin design patterns [2], [7], [18], specialized for an individual human user rather than a machine. The architecture consists of five layers. The first is a personal data source layer comprising the individual's calendar, task manager, device and application usage logs, and, optionally, wearable or self-report data. The second is a data collection and preprocessing layer that ingests, cleans, and temporally aligns these heterogeneous streams. The third is a digital twin state layer that maintains the continuously updated productivity state vector described in Section III-E. The fourth is a prediction and optimization layer that forecasts near-term productivity and generates constrained scheduling recommendations. The fifth is a human interaction and feedback layer through which recommendations are presented to the user and the user's actions and outcomes are captured and fed back into the twin. Fig. 2 depicts this layered flow.

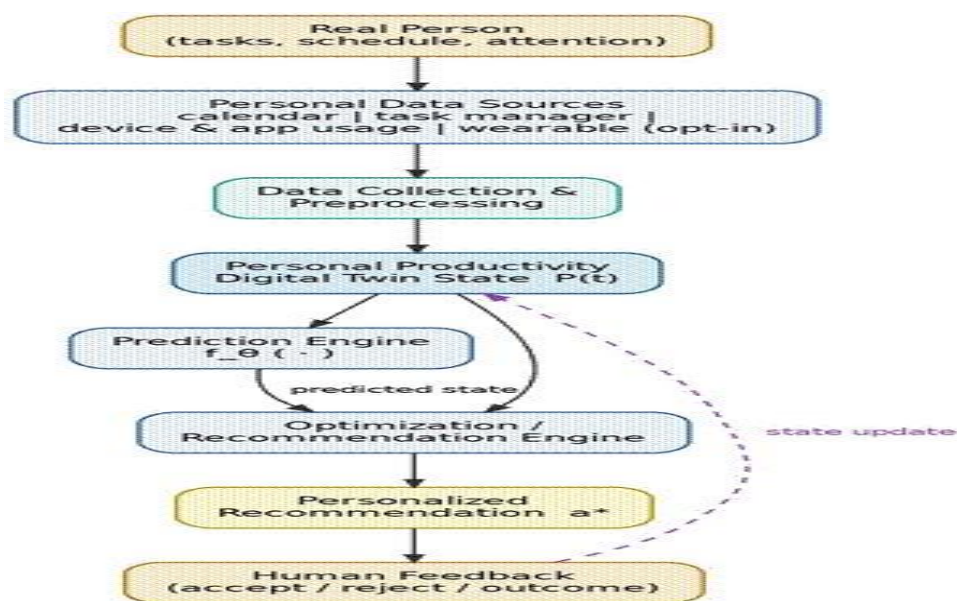


Fig. 2. Layered architecture of the proposed Personal Productivity Digital Twin.

Fig. 3 focuses on the data-flow perspective. It shows that raw personal data does not need to leave the device in identifiable form: on-device feature extraction and, where models are shared across users, federated model updates can keep the underlying calendar text, application names, or location traces local, consistent with the digital-twin privacy guidance discussed for sensitive physical counterparts in [14] and [18].

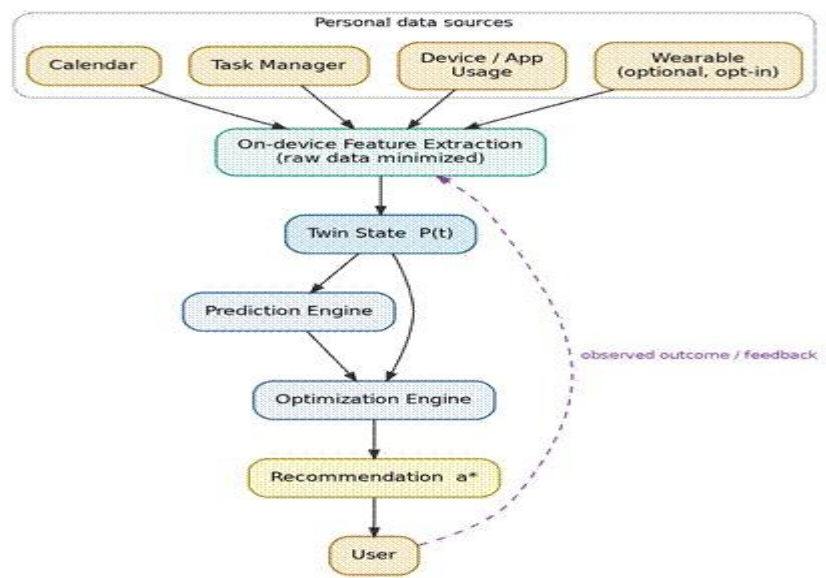


Fig. 3. Data flow between personal data sources, the digital twin, and the prediction and optimization engines.

B. Dataset and Data Collection

A central methodological constraint of this work is that no single publicly available dataset captures the full multidimensional state of a Personal Productivity Digital Twin as defined in Section III-E. The closest partial precedents are (i) the StudentLife dataset, which provides ten weeks of continuous smartphone sensing (activity, conversation, sleep, mobility) linked to academic and wellbeing outcomes for 48 students, but which was not designed around task-manager or calendar data and targets an academic rather than general knowledge-work population [9]; and (ii) the productivity-monitoring datasets used by Khan et al., namely a purpose-built productivity-logging application (“Prodmap”) and the CASAS smart-home activity dataset, which provide behavioral features linked to self-reported productivity change but were collected for a binary productivity-classification task rather than continuous multidimensional state estimation [8]. Neither dataset includes the full set of variables specified in Table I (e.g., explicit task priority, deadline pressure, and meeting load), and we are not aware of a public dataset that jointly captures calendar structure, task-manager metadata, device-usage-derived attention metrics, and self-reported productivity outcomes for the same individuals over an extended period.

Given this limitation, we do not claim to have trained or validated the proposed models on real user data. Two paths are available for future empirical work, and both are labeled explicitly rather than presented as already executed: (a) a controlled, IRB-approved data-collection study that instruments consenting participants’ calendars, task managers, and (optionally) device-usage logs over a multi-week period, following the ethical and consent practices established in prior smartphone-sensing studies [8], [9]; or (b) a synthetic/simulated dataset, generated from parameterized statistical models of task arrival, duration, deadline, and interruption processes calibrated against the aggregate statistics reported in the interruption science literature [10], [11], used to test the internal consistency, computational feasibility, and optimization behavior of the framework before any human-subject deployment. Any results obtained from such a synthetic dataset must be reported as simulated/synthetic results and not represented as real-world productivity outcomes.

C. Data Preprocessing

Because the PPDT integrates heterogeneous, asynchronously sampled sources, preprocessing must address four issues consistent with general digital-twin data-synchronization practice [2], [7]. Missing-value handling is required because device-usage or wearable streams can be interrupted by connectivity loss or user opt-out. Conservative imputation, such

as last-observation-carried-forward for slowly varying variables or explicit missingness indicators for sensing gaps, is preferred over aggressive interpolation that could fabricate apparent behavioral patterns. Normalization is required because features are measured on incompatible scales, for instance minutes of focused work versus a 1–5 self-reported energy rating. Person-specific normalization, that is, z-scoring against an individual's own historical distribution, is preferable to population normalization, since productivity-relevant baselines are known to vary substantially between individuals [17]. Temporal alignment and aggregation resample all sources onto a common time grid, such as hourly or daily bins, appropriate to the prediction horizon. Finally, privacy-preserving preprocessing performs feature extraction on-device where possible, so that only aggregated or encrypted features, rather than raw calendar text, application names, or location traces, are transmitted to any shared model. This is in line with the federated and encrypted update schemes proposed for privacy-sensitive digital twins [14].

D. Feature Engineering

Table I lists the variables used to construct the productivity state vector. Variables are selected only where a defensible source and, where available, supporting literature exist; several composite features (e.g., workload score, deadline pressure) are proposed derived quantities rather than directly observed signals, and are marked accordingly.

TABLE I. SUMMARY OF VARIABLES USED IN THE PROPOSED PERSONAL PRODUCTIVITY DIGITAL TWIN

Variable	Description	Data Type	Source / Basis
Task priority	Ordinal importance assigned to a task	Ordinal	Task manager
Task duration (est./actual)	Estimated vs. actual time spent on a task	Continuous (min)	Task manager / calendar
Deadline pressure (proposed)	Derived urgency = $f(\text{time remaining, task duration})$	Continuous [0,1]	Derived from calendar + task manager
Focus ratio (F)	Proportion of active time in uninterrupted work sessions	Continuous [0,1]	Device/application usage logs [8]
Context-switch frequency (C)	Count of application/task switches per hour	Count/hour	Device-usage logs; grounded in [10], [11]
Meeting load (M)	Hours scheduled in meetings per day	Continuous (hrs)	Calendar
Historical completion rate	Proportion of past tasks completed by deadline	Continuous [0,1]	Task manager history
Time-of-day factor (T)	Cyclical encoding of hour-of-day productivity tendency	Continuous cyclical /	System clock + historical logs
Workload score (W, proposed)	Composite of open tasks, meeting load, and deadline pressure	Continuous	Derived (proposed composite)
Energy/fatigue proxy (E)	Self-report or wearable-derived fatigue indicator	Continuous [0,1]	Self-report; optional wearable [17]
Goal progress (G)	Proportion of a tracked goal completed to date	Continuous [0,1]	Goal-tracking application (if used)
Distraction frequency	Count of non-work application openings per hour	Count/hour	Device-usage logs
Break frequency	Count of self-initiated breaks per day	Count/day	Device-usage logs / self-report

E. Digital Twin State Representation

We define the productivity-relevant state of the individual at time t as the vector

$$P(t) = [F(t), W(t), C(t), E(t), T(t), D(t), M(t), G(t)] \quad (1)$$

where $F(t)$ is the focus ratio, $W(t)$ is the composite workload score, $C(t)$ is the context-switch frequency, $E(t)$ is the energy/fatigue proxy, $T(t)$ is the time-of-day productivity factor, $D(t)$ is deadline pressure, $M(t)$ is meeting load, and $G(t)$ is goal progress, each as defined in Table I and normalized to a common scale prior to use. This eight-dimensional vector is proposed as a minimal but literature-grounded representation: F , C , and M are directly motivated by the attention-fragmentation literature [10], [11]; D and the components of W follow directly from task-manager and calendar metadata; and E and G allow the state to incorporate optional wearable/self-report and goal-tracking signals without making them mandatory, addressing the sensor-availability limitation discussed in Section III-B. The vector is intentionally extensible: additional dimensions (e.g., sleep duration, social-interaction level) may be appended where a user consents to the corresponding data source, following the same person-specific normalization described in Section III-C.

F. Productivity Prediction

The prediction task is to estimate the near-term future state, or a scalar productivity outcome derived from it, from the current and historical twin state and contextual information:

$$\hat{P}(t+) = f_{\theta}(P(t), H(t-k:t), Ctx(t)) \quad (2)$$

where $H(t-k:t)$ denotes the k most recent historical state observations, $Ctx(t)$ denotes contextual features such as day-of-week or upcoming deadline density, and f_{θ} is a model with parameters θ learned per individual (or initialized from a population-level model and then personalized). Model selection should follow the data-availability and dependency-structure considerations discussed in Section II-F rather than a fixed default choice. For a newly onboarded user with limited historical data, a regularized linear or treeensemble model (e.g., random forest [19] or gradient-boosted trees such as XGBoost [20]) is proposed as a more data-efficient and interpretable starting point, consistent with the linear model used for GPA prediction from sensing data in [9] and the classifier-based approach used for productivity-state change detection in [8]. As sufficient longitudinal history accumulates for a given user, a sequence model such as an LSTM [21], or, where very large amounts of fine-grained data are available, a transformer-based model [22], may better capture temporal dependencies such as multi-day fatigue accumulation. We emphasize that this is a proposed model-selection strategy, not a result of a conducted model comparison; Section IV presents this comparison only at the conceptual/literature-based level.

G. Productivity Optimization

Given a predicted state and a candidate set of task orderings or schedule modifications $A = \{a_1, \dots, a_n\}$, the π PPDT selects a recommendation by maximizing a multi-objective utility function subject to explicit constraints:

$$U(a) = \alpha \cdot \text{Prod}(a) - \beta \cdot \text{Switch}(a) - \gamma \cdot \text{Overload}(a) - \delta \cdot \text{Risk}(a) \quad (3)$$

where $\text{Prod}(a)$ is the predicted productivity outcome under action a ; $\text{Switch}(a)$ is a penalty proportional to the number and estimated cost of context switches induced by a , grounded in the disruption costs reported in [10], [11]; $\text{Overload}(a)$ penalizes scheduling that exceeds a personalized workload ceiling; $\text{Risk}(a)$ penalizes an increased probability of deadline violation; and $\alpha, \beta, \gamma, \delta \geq 0$ are personalized weights reflecting an individual's preferences, to be calibrated rather than fixed a priori. The optimization is subject to constraints of the form $\sum_i \text{duration}(\text{task}_i, a) \leq \text{AvailableTime}(t)$, $\text{end}(\text{task}_i, a) \leq \text{deadline}(\text{task}_i)$, $W(a) \leq W_{\max}$, $a_i \in \text{Prefs}$ (4) where $\text{AvailableTime}(t)$ is the unscheduled time in the individual's calendar, $W_{\max}$ is a personalized maximum-workload threshold, and Prefs encodes hard user preferences (e.g., no meetings before a stated hour). This formulation is deliberately compatible with two complementary solution strategies identified in the literature: for small, well-specified action sets it can be solved with exact or near-exact combinatorial methods such as branch-and-bound or genetic algorithms, following the multi-objective task-assignment and smart-calendar formulations of [15] and [16]; for settings where the twin must learn an interaction policy online from repeated recommend-observe cycles, it can instead be framed as a Markov decision process and solved with reinforcement learning, following the personalization approaches surveyed in [12] and [13]. We do not claim experimental superiority of either strategy; the choice is proposed as contingent on whether the action space is enumerable at each decision point (favoring search-based optimization) or must be learned from interaction history (favoring RL).

H. Feedback and Digital Twin Updating

The final architectural component closes the loop between recommendation and lived behavior. After a recommendation a^* is presented, the system observes the human's actual action a_{obs} , which may differ from a^* if the recommendation is modified or ignored, along with the resulting outcome $O(t+\Delta)$: whether the task was completed, rescheduled, or

abandoned, and the individual's self-reported or sensed productivity state afterward. The discrepancy between the predicted and observed state is used to update both the prediction model parameters θ and the personalized optimization weights (α , β , γ , δ). This follows the general digital-twin principle that the virtual representation must be kept synchronized with its physical counterpart through a persistent bidirectional data link [1], [2]. Fig. 4 depicts this cycle, and Algorithm 1 specifies it procedurally.

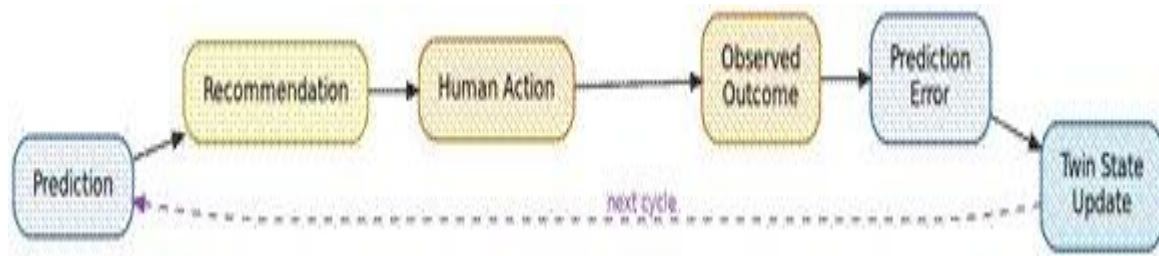


Fig. 4. Digital twin feedback loop linking prediction, recommendation, human action, and twin update.

Algorithm 1: Personal Productivity Digital Twin Operating Cycle Input: raw personal data stream $D(t)$ from consented sources; historical twin-state archive H ; user preference set P_{refs} ; initial model parameters θ (population-initialized) Output: continuously updated twin state $P(t)$ and a personalized recommendation stream

- 1: Initialize $P(0)$ from a short baseline calibration period
- 2: for each decision epoch $t = 1, 2, 3, \dots$ do
- 3: Collect raw data $D(t)$ from calendar, task manager, device-usage, and (if opted in) wearable sources
- 4: Preprocess $D(t)$: align timestamps, impute missing values, normalize against the user's own historical baseline
- 5: Extract features and update state vector
- 6: $P(t) \leftarrow [F(t), W(t), C(t), E(t), T(t), D(t), M(t), G(t)]$ Predict near-term state:
- 7: $P_{\text{hat}}(t+\Delta) \leftarrow f_{\theta}(P(t), H(t-k:t), C_{\text{tx}}(t))$
- 8: Generate candidate action set $A = \{a_1, \dots, a_n\}$ (feasible schedules / task orderings)
- 9: for each a in A : compute $U(a)$ subject to constraints (Eq. 4)
- 10: Select recommendation $a^* \leftarrow \text{argmax}_{\{a \text{ in } A\}} U(a)$
- 11: Present a^* to the user via the interaction layer
- 12: Observe human action a_{obs} and outcome $O(t+\Delta)$
- 13: Compute prediction error $e \leftarrow \text{Actual}(t+\Delta) - P_{\text{hat}}(t+\Delta)$
- 14: Update θ and personalized weights ($\alpha, \beta, \gamma, \delta$) using e (incremental / online update)
- 15: Append $P(t+\Delta)$ to twin-state archive H
- 16: end for

IV. RESULTS AND DISCUSSION

As established in Section III-B, no dataset currently available to the authors jointly captures the full PPDT state across a general knowledge-work population, and no experiments were conducted for this paper. This section therefore presents a conceptual and literature-grounded discussion of expected system behavior; evaluation methodology, and how the proposed framework compares with existing approaches. Wherever a quantitative figure is mentioned, it is explicitly attributed to the literature that reported it; no accuracy, error, or improvement figures are original to this paper.

A. Experimental Setup (Proposed, Not Yet Conducted)

A rigorous empirical evaluation of the PPDT would require, at minimum: (i) a multi-week, consented data collection period per participant, long enough to establish an individual baseline and observe workload variation, following the term-length precedent set by the ten-week StudentLife protocol [9]; (ii) a held-out temporal split (training on earlier weeks, evaluating on later weeks) rather than a random split, since productivity data are autocorrelated in time; (iii) a cold-start evaluation condition simulating a newly onboarded user with only baseline-period data, given the personalization challenges documented for wearable-based personal sensing models [17]; and (iv) an interleaved online evaluation of the recommendation engine, in which recommendation acceptance and downstream outcomes are logged prospectively rather than only retrospectively predicted. None of these steps have been executed for the present paper; they are specified here

as the minimum protocol required before the framework's predictive or optimization claims could be empirically supported.

B. Evaluation Metrics

Table III lists metrics proposed for evaluating a future PPDT implementation, distinguishing prediction metrics from outcome- and human-factors-level metrics.

TABLE III. PROPOSED EVALUATION METRICS FOR A PERSONAL PRODUCTIVITY DIGITAL TWIN

Metric	Applies To	Description	Justification
MAE / RMSE	Prediction	Magnitude of error between predicted and observed productivity score	Standard regression accuracy metrics
R^2	Prediction	Proportion of variance in the outcome explained by the model	Indicates overall model fit, cf. GPA-prediction use in [9]
Task completion rate	Optimization outcome	Proportion of scheduled tasks completed on time	Direct, interpretable productivity indicator
Deadline violation rate	Optimization outcome	Proportion of tasks completed after their deadline	Captures scheduling risk, relates to Risk(a) in Eq. 3
Context-switch frequency (post-recommendation)	Behavioral	Change in switching frequency after adopting a recommendation	Grounded in fragmentation-cost literature [10], [11]
Recommendation acceptance rate	Human factors	Proportion of recommendations the user follows without modification	Indicates perceived usefulness/trust
Personalization gap	Prediction	Difference between personalized and population-average model error	Assesses whether personalization adds value over a generic model
Prediction / recommendation latency	System	Time to compute a recommendation once new data arrives	Assesses real-time feasibility of the pipeline

C. Performance Analysis (Literature-Reported, Not Our Experimental Results)

Although the present framework has not itself been evaluated, related literature provides indicative, though not directly transferable, evidence that behaviorally derived features carry predictive signal for productivity-related outcomes. Khan et al. reported that logistic regression, support-vector-machine, and nearest-neighbor classifiers each achieved good predictive performance when detecting productivity-state change from smartphone- and activity-derived features across two independent benchmark datasets [8]; the StudentLife study similarly reported significant correlations between passively sensed behavioral features and academic performance, and demonstrated that a simple regularized linear model could predict cumulative GPA from such features [9]. These are literature-reported results from prior systems evaluated on their own datasets and populations; they are cited here only as evidence that the general class of behavioral-sensing-to-productivity-outcome prediction is feasible in principle, not as a performance claim for the PPDT proposed in this paper, whose accuracy on any specific dataset remains an open empirical question.

D. Productivity Optimization Analysis (Conceptual Walkthrough)

To illustrate, not to validate, the intended behavior of the optimization layer, consider a hypothetical, illustrative user state in which the twin estimates elevated deadline pressure $D(t)$, elevated meeting load $M(t)$, and a below-baseline focus ratio $F(t)$. Under the objective in Eq. 3, an increase in $D(t)$ and $M(t)$ would be expected to raise the Risk(a) and Overload(a) penalty terms for schedules that pack additional deep-work tasks into an already dense day, favoring recommendations

that either defer lower-priority tasks, consolidate meetings to reduce context switching (lowering Switch(a)), or explicitly protect a contiguous focus block before a looming deadline. This example is presented solely to clarify how the proposed objective function is intended to trade off competing considerations; it uses illustrative, non-empirical values and must not be read as a demonstrated or measured productivity improvement.

E. Comparative Analysis

Table IV situates the proposed PPDT relative to the main categories of prior work reviewed in Section II, along five qualitative dimensions: whether the approach maintains a persistent, individual-level state model; whether it has forward predictive capability; whether it produces optimization-driven recommendations (as opposed to passive monitoring); whether it closes an explicit feedback loop back into the model; and its primary application domain. The comparison is conceptual and feature-based rather than a benchmark of measured performance, since the underlying systems were evaluated on different tasks, populations, and datasets and are not directly comparable numerically.

TABLE IV. CONCEPTUAL COMPARATIVE ANALYSIS OF EXISTING APPROACHES AND THE PROPOSED PPDT

Approach	Persistent Individual State	Predictive Capability	Optimization Recommendation /	Feedback Loop	Primary Domain
Industrial digital twins [1], [2]	Yes (assetlevel)	Model-based simulation	Maintenance / process optimization	Yes	Manufacturing / engineering
Human / cognitive digital twin surveys [3], [4], [6]	Conceptual; mostly healthcareoriented	Partial, domainspecific	Limited	Partial	Healthcare, rehabilitation, cognition
Personal digital twin (healthcare) [5]	Yes, health-specific	Yes (clinical)	Clinical decision support	Yes	Personalized healthcare
Smartphone productivity/wellbeing sensing [8], [9]	No persistent twin; session/period based	Yes (classification/regression)	None (monitoring only)	No	Productivity & wellbeing sensing
RL / search-based scheduling & personalization [12], [13], [15], [16]	No persistent multidimensional twin	Partial (policylevel)	Yes (task/schedule optimization)	Yes (policy updates)	Scheduling / recommendation
Proposed PPDT	Yes; continuously updated multidimensional state	Yes; learned, personalized	Yes; constrained multi-objective	Yes; explicit human-in-the-loop	Personal productivity

The comparison indicates that the proposed framework's contribution lies less in any single novel component. Persistent state modeling, predictive modeling, and optimization-based scheduling each have precedent individually. Its contribution lies instead in combining all three within one continuously updated, feedback-driven loop specifically scoped to everyday personal productivity, a combination that Sections II and IV suggest has received limited attention to date.

V. PRIVACY, SECURITY, AND ETHICAL CONSIDERATIONS

A system that continuously models an individual's tasks, schedule, and attention raises privacy and autonomy concerns beyond those typical of industrial digital twins, and these concerns must shape the design rather than being addressed only after deployment. Data minimization and on-device processing, as described in Section III-C, should be the default rather than the exception, transmitting only the minimum aggregated features required for prediction rather than raw calendar or application content. Where model training must be shared across users to bootstrap a new user's model, privacy-preserving learning techniques such as federated learning and, where feasible, differential privacy or homomorphic encryption of model updates should be used, consistent with proposed digital-twin privacy architectures [14], [18]. Consent and data ownership must be explicit and revocable: a user should be able to see which data sources feed the twin, disable any source, and delete accumulated history, and the twin should degrade gracefully (falling back to simpler, less personalized recommendations) rather than failing when data sources are withdrawn.

Beyond data protection, a PPDT raises distinct autonomy and psychological risks. A system that continuously scores and schedules a person's work can shade into behavioral surveillance if deployed or monitored by an employer rather than controlled by the individual, and the interruption-cost literature itself warns that individual differences in how people respond to monitoring and interruption are substantial [11], suggesting that a one-size-fits-all optimization objective could be experienced as intrusive or stress-inducing by some users even if it is statistically productivity-improving on average. Explainability of recommendations, the ability to override or ignore the twin without penalty, and safeguards against "overoptimization" (for example, hard limits preventing the objective in Eq. 3 from scheduling away all discretionary or rest time in pursuit of a higher productivity score) are proposed as necessary design constraints rather than optional features. Algorithmic bias is also a concern: a model trained predominantly on one population's work patterns (e.g., a specific student cohort, as in [9]) may generalize poorly to individuals with different work structures, schedules, or cultural norms around time use, reinforcing the need for the personalization and cold-start handling discussed in Sections III-C and III-F rather than reliance on a single population-level model.

VI. LIMITATIONS

This work has several limitations that should be considered when interpreting its contribution. First, and most importantly, no experiments were conducted. The framework, state formulation, and algorithm are proposed and conceptually motivated by literature rather than empirically validated, and their real-world predictive accuracy and optimization benefit remain unknown. Second, as discussed in Section III-B, no existing public dataset captures the full proposed state vector, so even a retrospective evaluation would first require a dedicated data-collection effort. Third, productivity is a multidimensional and partly subjective construct. The composite features proposed in Table I, such as workload score and deadline pressure, are reasonable derived quantities but are not themselves independently validated psychometric instruments, and cognitive states such as focus or fatigue are only indirectly observable through behavioral proxies. Fourth, individual variability and behavioral drift mean that a model calibrated during one period of a person's life, such as a stable work routine, may degrade as circumstances change, for example following a new role or life event, requiring ongoing recalibration rather than a one-time training pass. Fifth, the cold-start problem, that is, insufficient historical data for a newly onboarded user, remains a genuine open challenge, only partially addressed by the self-supervised personalization strategies noted in [17]. Sixth, generalization across individuals and cultural contexts is not established. The sensing literature reviewed in Section II draws primarily on specific populations, such as university students in [9], whose patterns may not transfer to other work contexts. Finally, computational and user-acceptance constraints, including the latency required for real-time recommendations and whether users would trust and adopt an automated scheduling agent at all, are practical deployment questions this paper does not resolve.

VII. FUTURE WORK

Several directions follow directly from the limitations above. The most immediate is a controlled, consented data-collection study explicitly designed around the PPDT state vector in Table I, combining calendar and task-manager metadata with device-usage-derived attention features, to enable the first genuine empirical evaluation of the framework. Methodologically, future work should compare the tabular and sequential prediction models discussed in Section III-F directly on such data, and should empirically compare the search-based and reinforcement-learning optimization strategies discussed in Section III-G rather than treating the choice as purely conceptual. On the modeling side, promising extensions include multimodal twins that incorporate wearable physiological signals where users opt in [17]; large-language-model-based or agentic components that can explain recommendations in natural language or negotiate schedule tradeoffs conversationally; federated and privacy-preserving training across many users' devices to bootstrap cold-start models without centralizing raw data [14]; and human-in-the-loop reinforcement learning that explicitly models a user's tolerance for automated schedule changes rather than assuming full compliance with recommendations. Finally, longitudinal studies spanning months rather than weeks would be needed to assess whether a PPDT's benefits (if any) persist as behavioral drift and changing life circumstances are encountered, and dedicated user studies are needed to assess the psychological and autonomy effects discussed in Section V before any workplace-scale deployment is considered.

VIII. CONCLUSION

This paper proposed the Personal Productivity Digital Twin (PPDT), a conceptual framework that extends the digital-twin paradigm, previously applied mainly to industrial systems and, more recently, to healthcare-oriented human digital twins [1]–[6], to the domain of everyday personal productivity. Motivated by well-established evidence that task fragmentation

and interruption carry measurable productivity and stress costs [10], [11], and by evidence that passively sensed behavioral data can predict productivity-relevant outcomes [8], [9], we argued that existing productivity tools generally lack a persistent, continuously updated, and optimization-linked representation of the individual, and that existing human/personal digital twin research has not concentrated on this specific problem. We proposed a five-layer PPDT architecture, an eight-dimensional productivity state vector, prediction and multi-objective optimization formulations grounded in established modeling and scheduling techniques [12], [13], [15], [16], [19]–[22], and an operating-cycle algorithm that closes the loop between prediction, recommendation, human action, and twin update. Because no dataset currently spans the full proposed state and no experiments were conducted, this contribution should be read as a scoped, literature-grounded framework and research agenda rather than a validated system. We have accordingly distinguished throughout between established findings, our proposed methodology, and the conceptual, non-experimental discussion in Section IV, and have outlined the privacy, ethical, and empirical work, summarized in Sections V through VII, required before such a system could be responsibly built and evaluated.

IX. DISCLOSURE AND CONFLICT OF INTEREST

The authors declare no known conflict of interest. No external funding is reported for this work. [Funding source, ethics approval, and data availability statements to be completed by the authors/institution prior to submission.]

REFERENCES

- 1) M. Grieves and J. Vickers, "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," in *Transdisciplinary Perspectives on Complex Systems*, F.-J. Kahlen, S. Flumerfelt, and A. Alves, Eds. Cham, Switzerland: Springer, 2017, pp. 85–113.
- 2) National Academies of Sciences, Engineering, and Medicine, *Foundational Research Gaps and Future Directions for Digital Twins*. Washington, DC, USA: The National Academies Press, 2024.
- 3) Y. Lin et al., "Human digital twin: a survey," *Journal of Cloud Computing*, vol. 13, art. 143, 2024.
- 4) M. W. Lauer-Schmaltz, P. Cash, J. P. Hansen, and A. Maier, "Towards the human digital twin: Definition and design, a survey," Technical University of Denmark, preprint, 2024.
- 5) R. Sahal, S. H. Alsamhi, and K. N. Brown, "Personal digital twin: A close look into the present and a step towards the future of personalised healthcare industry," *Sensors*, vol. 22, no. 15, art. 5918, 2022.
- 6) X. Zheng, J. Lu, and D. Kiritsis, "The emergence of cognitive digital twin: vision, challenges and opportunities," *Int. J. Prod. Res.*, vol. 60, no. 24, pp. 7610–7632, 2022.
- 7) L. Bedogni and F. Chiariotti, "A web of things architecture for digital twin creation and model-based reinforcement control," *arXiv preprint arXiv:2301.12761*, 2023.
- 8) S. A. Khan, A. A. Farhan, L. G. Fahad, and S. F. Tahir, "Personal productivity monitoring through smartphones," *J. Ambient Intell. Smart Environ.*, vol. 12, no. 6, pp. 491–507, 2020.
- 9) R. Wang et al., "StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. (UbiComp)*, 2014, pp. 3–14.
- 10) G. Mark, V. M. Gonzalez, and J. Harris, "No task left behind? Examining the nature of fragmented work," in *Proc. SIGCHI Conf. Human Factors Comput. Syst. (CHI)*, 2005, pp. 321–330.
- 11) G. Mark, D. Gudith, and U. Klocke, "The cost of interrupted work: More speed and stress," in *Proc. SIGCHI Conf. Human Factors Comput. Syst. (CHI)*, 2008, pp. 107–110.
- 12) X. Tang, Y. Chen, X. Li, J. Liu, and Z. Ying, "A reinforcement learning approach to personalized learning recommendation systems," *Br. J. Math. Stat. Psychol.*, vol. 72, no. 1, pp. 108–135, 2019.

- 13) F. den Hengst, E. M. Grua, A. el Hassouni, and M. Hoogendoorn, "Reinforcement learning for personalization: A systematic literature review," *Data Science*, vol. 3, no. 2, pp. 107–147, 2020.
- 14) J. Li and D. Wang, "Federated learning for digital twin applications: a privacy-preserving and low-latency approach," *PeerJ Computer Science*, vol. 11, art. e2877, 2025.
- 15) S. T. Ngo, J. Jaafar, I. A. Aziz, and B. N. Anh, "A compromise programming for multi-objective task assignment problem," *Computers*, vol. 10, no. 2, art. 15, 2021.
- 16) J. Cowton and L. Yang, "A smart calendar system using multiple search techniques," in *Proc. 15th UK Workshop on Computational Intelligence (UKCI)*, Exeter, U.K., 2015.
- 17) T. Islam and P. Washington, "Individualized stress mobile sensing using self-supervised pre-training," *Applied Sciences*, vol. 13, no. 21, art. 12035, 2023.
- 18) Y. Wang, Z. Su, S. Guo, M. Dai, T. H. Luan, and Y. Liu, "A survey on digital twins: Architecture, enabling technologies, security and privacy, and future prospects," *IEEE Internet of Things Journal*, 2023 (arXiv preprint arXiv:2301.13350).
- 19) L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- 20) T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, 2016, pp. 785–794.
- 21) S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- 22) A. Vaswani et al., "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.