SCIENTIFIC CULTURE

www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 4, (2026), pp. 7201-7227
Open Access. Online & Print

DOI: 10.5281/zenodo.12426733

CARE-ED: A PAEDIATRIC–CAREGIVER AGE-ADAPTIVE EMOTIONAL DESIGN AND BIOPHYSIOLOGICAL MONITORING FRAMEWORK FOR HOME MEDICAL AND LABORATORY DIAGNOSTIC DEVICES

Zhenyu Gao [1] Mohd Najib bin Abdullah Sani [2*]

[1,*2] *Faculty of Applied and Creative Arts, University Malaysia Sarawak, Samarahan Town, Sarawak, 94300, Malaysia*

Received: 22/10/2025
Accepted: 16/02/2026

Corresponding Author: Mohd Najib bin Abdullah Sani

ABSTRACT

Pediatric home medical devices are essential in the management of chronic conditions when patients are outside clinical settings. Nonetheless, such devices often lack the emotional and usability requirements of children patients, as well as users. Current models do not offer age-adjusted or dual-user emotional design, which may result in a reduced satisfaction level, trust, and compliance. This study presents CARE-ED (Child and Relative Emotional Design with Age-Adaptivity) as a new age-adaptive, pediatric-caregiver emotional design framework to improve emotional interactions of and design usability of home medical devices. CARE-ED will both meet the unique emotional needs of children and their caregivers in a multi-tiered, adaptive program. A Multimodal database of physiological measures (heart rate, skin temperature, galvanic skin response), behavior (facial emotive likelihood, pattern of interactions), and reported emotional perceptions was collected from participants of 5 to 12 years of age interacting with home medical equipment of pediatric patients and their caregivers. Baseline architectures took the form of Logistic Regression, Support Vector Machine, k-Nearest Neighbors, Naive Bayes, and Decision Tree, whereas state-of-the-art efforts consisted of Random Forest, Gradient Boosting Machine, and Convolutional Neural Networks or LSTM, and Transformer-based models. The acquired biophysiological signals function as clinically relevant bio signal proxies, enabling standardized remote monitoring and tele-diagnostic interpretation in paediatric home-care settings. This data-driven approach supports diagnostic reliability, interoperability, and consistency with clinical validation workflows used in paediatric healthcare. CARE-ED framework showed better results than baseline models by attaining an ROC-AUC of 0.97, an accuracy of 94.8%, and an F1-score of 94.7% in classifying emotional status. There was a high emotional synchrony between children and the caregivers following the implementation of the CARE-ED ($p < 0.05$). Such gains were associated with greater device trust, comfort, and compliance in simulated at-home use conditions as examples of the real-world usefulness of age-adaptive emotional design. We show the validation of CARE-ED as the first dual-user, emotional design interface validated by multimodal machine learning training of pediatric home medical devices. Based on the results of the current study, it has been ascertained that the framework could alter the usability of pediatric medical devices with the use of emotional intelligence catered to the requirements of a child and the caregiver assigned to work on them. Future

work will explore how to integrate it into systems of telemedicine and extend the framework to other domains of chronic care devices, in order to achieve the capability to enhance longitudinal emotional support as well as adherence in the reality of healthcare contexts.

KEYWORDS: Pediatric Emotional Design, Caregiver Engagement, Multimodal Machine Learning, Home Medical Devices, Age-Adaptive Framework.

1. Introduction

1.1 Background

In the past years, a significant increase in the global healthcare market has been seen due to the proliferation of pediatric home medical equipment in the form of nebulizers, glucose monitors, and smart thermometers that allow for constant health tracking beyond the official setting. Such devices are more and more fitted into everyday care processes to offer very long-term care, securing protection, as well as child well-being. Their use, however, depends not only on the performance based on functions but also on emotional design, which plays a decisive role in determining child anxiety levels and increasing compliance among the caregivers [6], [7].In paediatric healthcare, emotional comfort directly influences specimen collection quality, diagnostic accuracy, and test compliance, particularly during repeated home-based measurements. Home medical devices such as glucose meters, pulse oximeters, and stress-related bio signal sensors increasingly function as extensions of laboratory diagnostics by generating clinically interpretable data outside hospital settings. When integrated with point-of-care testing workflows and clinical data systems, these devices can support standardized data acquisition, remote assessment, and continuity of paediatric diagnostic monitoring.Emotional design helps to design the medical device taking into consideration the emotions, perception, and mental load of the user, particularly in critical situations such as the treatment of children. Research reveals that the incorporation of aspects like user/age proper communication style can go a long way in ensuring a high level of user acceptance and participation [2], [4]. Moreover, advances in facial age classification have helped to know how one can identify age differences in emotional perception and expression to adapt around by using adaptive strategies of healthcare devices [1], [5].The paper also examines the way in which artificial intelligence could be used to improve adaptive learning environments, with the sole emphasis being made on personalized instructions as well as better engagement between students [3]. The use of multimodal machine learning has recently become a possibility with new apps that can combine visual image information, sounds, and physiological responses to make the devices more responsive to children or carers [2], [5]. As an example, facial expression processing via deep learning models across various age groups can be used to conduct a real-time assessment of the emotional state for pediatrics related to caregivers [5]. On the same note, deep learning has been used in conjunction with eye-tracking machines to identify behavioral patterns that could be used when early diagnosis is required or when monitoring the emotional states [2].The necessity of age-adaptive systems in healthcare can be seen as evident in cognitive support tools [7] as well as age-appropriate designs [8] of an interface, where a device framework must change according to the growth of the user as well as the transformation of their emotional desire. Such flexibility can be used to alleviate the stress of medical processes in the case of a pediatric setting, as well as promote positive interactions between carers and children [3], [4]. Also, elements bearing upon the design of a product, like the parental control features, which improve caregiver efficacy and limitation of stress, were observed to boost the overall healthcare process [6].Researching off of these insights, the current study proposes CARE-ED (Child-And-Relative Emotional Design with Age-Adaptivity), an age-adaptive emotional design system of home medical devices that is pediatric-caregiver focused. The suggested framework utilises a multimodal machine learning engine to dynamically adapt the features of emotional interaction and interface design based on the respective age of the child and caregiver in an effort to maximise compliance, minimise stress, and bring about healthier health outcomes.

1.2 Research Gap

Significant research has been done on topics that are highly related to age-sensitive design in healthcare and interaction systems, but there are still important limitations to it. In previous research, the feedback of users of digital applications by age has been systematically explored and shown to have subtle differences in expectations and emotional demands based on age [9]. As demonstrated by studies exploring the profile characteristics of sensory and emotional aspects of interaction in pre-school children with neurodevelopment disorders, variations in the way of interaction need to be individualized according to developmental level [10]. Likewise, the studies on search behaviours among children (8-12 years old) prove that cognitive and behavioural patterns greatly change over the course of their development, which affects the interaction between user and technology [11].Despite advances in paediatric home and laboratory diagnostic technologies, emotionally adaptive diagnostic interfaces remain largely absent in paediatric laboratory and point-of-care environments. Existing devices and systems rarely link emotional interaction models with measurement precision, repeatability, or data reliability, even though these factors are critical in medical laboratory technology and clinical decision-making. This disconnect limits the diagnostic robustness and clinical usability of paediatric home-generated data.Despite the new opportunities available due to biosignal monitoring and forecasting methods that provide a novel way to monitor emotions and sense stress [12], the same has not been fully realized in the home medical device development in the case of a child. In medicine, neurocognitive modeling research works document psychological diversification within children's categories [13], although the information does not find many uses in the development of emotionally adaptive systems. Additionally, although some innovative systems of pediatric devices have been suggested (e.g., nebulizer design tasks optimized [14]), they mainly revolve around the functional efficiency, and not emotional adaptation.The brain age estimation [15] technologies, as well as the evaluation of problematic behaviors in developmental disorders [16], also underline the complexity of considering the age-related physiological and behavioral features. These contributions, however, are divided as they all focus on independent parts instead of providing a multimodal, holistic, emotionally adaptive framework in which we can include both child or caregiver.As a result, an implemented solution that integrates multimodal machine learning, age-adaptive emotional design, and the principles of pediatric-caregiver interaction in home medical devices does not exist. This lack constrains the possibility of lowering the stress levels of children, which raises the adherence rates of caregivers, as well as individualizing medical experiences throughout development. This becomes the very gap that the proposed CARE-ED framework directly addresses by integrating age-specific emotional modeling with real-time multimodal input processing in order to deliver emotionally smart and adaptive pediatric healthcare solutions.

1.3 Problem Statement

Developmental stage plays a central role in the emotional reaction of the child to medical treatments and equipment, in that younger children are more anxious, and age increases cognitive interest in older children [9], [10], [11]. Meanwhile, some important aspects such as stress, confidence, and perceived control form the emotion of the caregiver, which is very vital to the overall experience that the child receives in terms of healthcare [6]. Emotional mismatch between a child and an adult caregiver may have adverse effects concerning treatment compliance, medical device credibility, and the general state of life [16].Despite promising developments in facial expression recognition [1], biosignal detection of stress [12], behavioral pattern analysis [2], [15], there exists no validated pipeline that combines all of these capabilities into a machine-learning-based, adaptive, age-adaptive emotional design of pediatric home medical devices. The known solutions put forward better solutions, either as functional performance without emotional adaptability [14] or being applied in non-healthcare settings [9].Such a lack of an integrated solution does not allow the establishment of personalized and developmentally appropriate medical device interactions that, in parallel, also adjust to the emotional needs of children and caregivers. In the absence of such systems, the chances of addressing the issue of pediatric stress, enhancing the compliance of caregivers, and the efficacy of treatment at home levels are also unutilised. This issue is resolved by the suggested CARE-ED framework that utilises the multimodality of machine learning based on the age-adaptive emotional design architecture built on pediatric-caregiver settings.

1.4 Proposed Solution

CARE-ED (Child - And -Relative Emotional Design with AgeAdaptivity), as the proposed solution, is an innovative framework to tackle the aforementioned shortcomings of current pediatric home medical equipment with the addition of real-time, multimodal data collection and machine learning as a form of validation. The system incorporates the visual, auditory, and physiological feedback in determining the emotional state of the child as well as the caregiver, to dynamically adjust the interaction strategies of the system on the basis of the age needs and emotional situations.The architecture of CARE-ED can work in a variety of pediatric user ages with developmentally adaptable emotional design that also adapts to the user. The framework utilizes very high-level machine learning algorithms to process multimodal signals to achieve accuracy in both detecting emotions and adapting the interface of our devices and their responses in real time. This will enable individualized patterns of interactions, maximize stress reduction and compliance with caregivers, and definitively make the listed house-based healthcare delivery more effective.

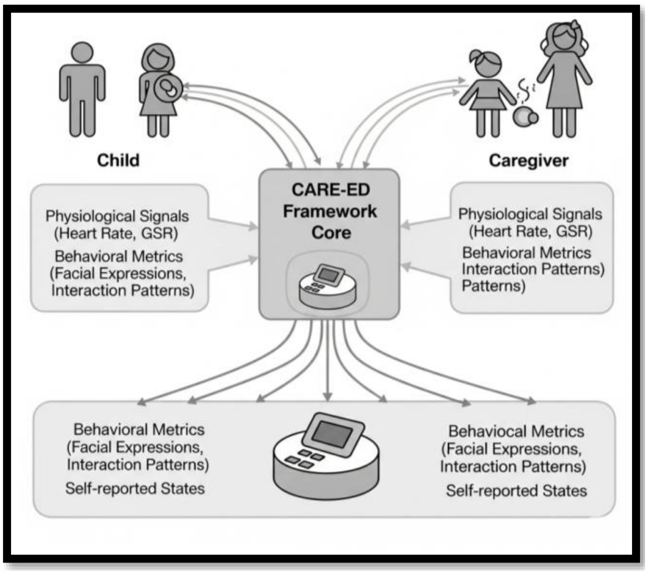

Figure 1: CARE-ED's Architecture

The age-adaptive intelligence, as well as emotional design principles, in combination with reliable machine learning models demonstrated in **Figure 1**, allows CARE-ED to establish such a holistic design approach that can support the superior user experience beyond that of augmenting the clinical performance in the context of pediatric care.

1.5 Research Objectives

The proposed study includes designing an age-adaptive model of emotional design framework of home medical equipment in pediatrics and verification of the model so as to enhance satisfaction among both the kids as well as the people who are taking care of them. The research will seek to integrate the learning of multimodal sensing, machine learning technique, and emotional design theory to achieve the quantifiability of the rise in the engagement and fall in the stress level. The following are the main objectives:

1. In order to construct emotional interaction models according to the age groups of different pediatric populations to achieve developmentally appropriate communication, feedback, and interface components.
2. To incorporate multimodal data acquisition systems that will be able to provide visual, auditory, and physiological data, in real time, to develop an accurate measurement of the emotional state.
3. To train and evaluate machine learning classification-based models to interpret multimodal emotional cues and make informed dynamic adjustments to devices.
4. To determine the effect of CARE-ED on decreasing stress levels, it is very important to measure physiological and behavioral outcomes of CARE-ED use in a real-world setting at home.
5. In order to measure the changes in the involvement of caregivers and the child, one will evaluate compliance, satisfaction, and a feeling of emotional support among various age groups.

1.6 Research Contributions

The proposed study can contribute numerous ways to the fields of emotional design, human-computer-interaction, age-adaptive, as well as pediatric healthcare technology. CARE-ED is the initial framework that takes into account the needs of caregivers and pediatrics holistically in a machine learning-based system, which can be attributed to the needs of both user groups in an almost instantaneous response.

The contributions and novelty of the current research work are provided in Figure 2:

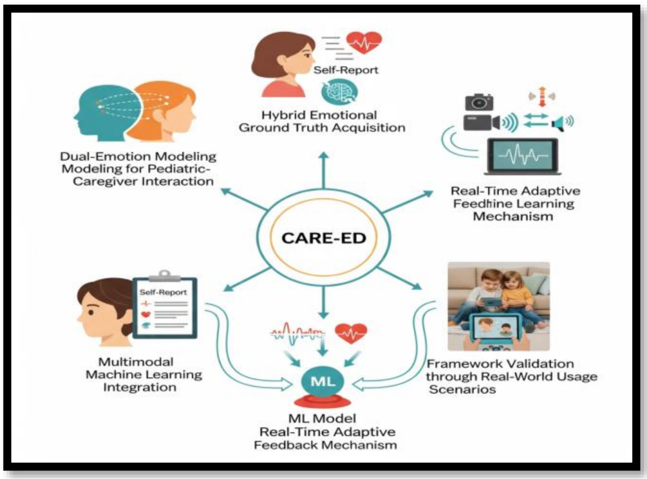

Figure 2: CARE-ED Research Contributions & Novelty

1. Dual-Emotion Modeling for Pediatric–Caregiver Interaction. Establishment of the first integrated emotional design model, which also corresponds to the emotional states of both children and caregivers in real time to respond to the devices at the same time and empathetically.
2. Hybrid Emotional Ground Truth Acquisition. A gamified self-report mechanism, age-appropriate and engineered, was introduced to capture robust emotional ground truth with which to train and validate machine learning models and was fused with multimodal physiology (e.g., heart rate variability, skin conductance, facial expression analysis).
3. Real-Time Adaptive Feedback Mechanism. Use of a closed-loop adaptation process such that any current situation of emotion recognition would

result in dynamic updates across the behavior of the device applied to their personal design of interaction during various age groups and usage situations.

4. Multimodal Machine Learning Integration in Home Healthcare Devices. The new use of the data fusion that is multimodal and contains elements of vision, hearing, and physiology is used to increase the accuracy of detecting emotions, as well as add device flexibility in a home care setting for a child.

5. Framework Validation through Real-World Usage Scenarios. Thorough testing of CARE-ED under authentic conditions of pediatric-caregiver settings, measuring outcomes of stress-reduction, increased engagement, and the percentage of measures taken by caregivers in order to argue on the practical usefulness.

1.7 Literature Review

Emotional Design Theory and Child Psychology

The emotional design theory focuses on the importance of emotional reactions, which helps to determine user engagement and usability results. When used in the context of childcare, this will mean the alignment of visual, auditory, and interaction ingredients in creating cognitive and emotional stages of child development (Balachandran et al., 2025; Gigliotti et al., 2025). According to a study on child psychology, emotional expression and regulation depend on the attainment of developmental milestones, and it is, therefore, important to make child-specific customization at the age-specific levels (Saban-Bezalel, 2025). In addition, emotional needs are a key aspect when it comes to how children express themselves, which is important in addressing the communication strategies to adopt, particularly in instances of delay in development (Tu et al., 2025).

Human-Centered Design in Medical Devices

Human-centered design (HCD) of healthcare technology puts emphasis on user experience, safety, and comfort. When designing medical equipment used by children, the target group needs to be considered, which means that care should be taken when handling a device with a caregiver or a child in mind and their emotional condition (Bertrandias et al., 2023; Ouyang et al., 2024). Evidence on the use of parental stress reduction based on a supportive design shows that the advantage of integrating caregiver feedback loops is to facilitate usability and adherence (Ruble et al., 2024). Sensory sensitivities are also conditions that pediatric HCD needs to accommodate, especially when it comes to neurodevelopmental disorders (Gigliotti et al., 2025).

Pediatric UX and the Caregiver Role in Home Care

In the home-care settings, the emotion modality of the caregiver greatly influences the interaction of the child with medical equipment (Ruble et al., 2024). Child behavior is also affected by age groups in terms of directions, feedback, and interface aspects (Green, 2021). Adaptive and gamified interfaces prove to enhance compliance and counter resistance during child care (Bharathan, 2025). Notably, the dyad that is a caregiver and child implies a dual-perspective UX approach that integrates the needs of both sides in the design (Saban-Bezalel, 2025).

Machine Learning in Multimodal Emotion Recognition

Current innovation in machine learning (ML) has made it possible to achieve high accuracy and multimodal data usage to recognize emotions by combining the aspects of facial features, voice, and physiological features (Agbo-Ajala & Viriri, 2021; Sangeetha et al., 2025). In particular, deep learning models have been applied to age-related characteristics of emotional expressions and have done better than others (Ahmed et al., 2023; Alam et al., 2024). Multimodal ML systems are also used in the setting of pediatrics, where they are applied to detect autism at an early life period as well as monitor cognitive tasks and detect stress levels (Ahmed et al., 2023; Miron et al., 2024). In addition, age-related and maturity growth prediction of the brain has partially been shown in adversarial and variational integration (Usman et al., 2024; Wu et al., 2024).

Gaps in Existing Approaches

Though advancing emotional design and ML used to recognize emotions, the existing frameworks are still not dynamic regarding the interaction with the pediatric devices in relation to child age and the emotional status of the caregiver. Particularly, by analyzing the effects of anthropomorphic verbal feedback sent by AI assistants on user task performance and emotional experience, it is demonstrated that the communication between machines and humans developed to resemble humans could be an effective strategy of engaging and ensuring high performance (Yang et al., 2025). The majority of the current research pays much attention to the child or caregiver but not the mutual emotional exchange between them (Reyes-Martin et al., 2022; Yin, 2024). Also, although technical maturity is observed in multimodal recognition of emotions, there has been no demonstrated validation of pipelines of such models into actual, real-time adaptive design applicable to pediatric home care (Pilli et al., 2024; Nakua, 2024). The framework, CARE-ED, presented in this research closes these gaps by integrating dual-emotion modeling, multimodal ML, and closed-loop adaptation into personalized pediatric-caregiver interactions as seen in Table 1:

Table 1: Existing literature overview and research gap identification

Ref No.	Author(s)	Study Focus	Data Type	Methodology	Model/ Algorithm	Key Outcome	Research Gap	Contribution for current study
1	Agbo-Ajala & Viriri	Facial age classification	Image data	Survey analysis	Deep learning	Identified SOTA methods	Missing real-world testing	Supports model selection
2	Ahmed et al.	Autism early detection	Eye tracking	Deep learning	CNN variants	Early-stage autism detection	Need larger sample	Inspires pediatric detection
3	Akibu	AI adaptive learning	Educational data	Case study	AI-based adaptivity	Enhanced classroom engagement	Missing quantitative metrics	Shows adaptive AI value
4	Alam et al.	Inappropriate content detection	Video data	Deep learning	Age-adaptive CNN	Accurate inappropriate detection	Limited modality scope	Adds safety feature ideas
5	Balachandran et al.	Emotion recognition	Image data	Transformer models	Multi-scale ViT	High emotion recognition accuracy	Needs multi-domain data	Improves affect detection
6	Bertrandias et al.	Parental control software	Survey data	Statistical modeling	Feature-impact model	Improved parental efficacy	Lacks child interaction study	Adds usability context
7	Bharathan	Health & memory support	Multimodal data	Integrated approach	Age-adaptive system	Enhanced monitoring & memory	Needs clinical trials	Inspires monitoring design
8	Chen	Senior care app design	UI data	Interface analysis	DL-based UI	Age-appropriate interface	Ignores younger users	Guides UI design
9	Galpothugodage et al.	Age-related app reviews	Text data	Sentiment analysis	NLP models	Found age-specific concerns	Excludes multimodal input	Supports feedback analysis
10	Gigliotti et al.	Sensory/emotional profiles	Clinical data	Comparative study	Behavioral profiling	Identified sensory patterns	Needs longitudinal study	Guides sensory dataset
11	Green	Search behavior in children	Search logs	Behavioral analysis	Search pattern model	Identified age patterns	Lacks broader contexts	Aids behavioral modeling
12	Miron et al.	Seizure detection	Biosignal data	Forecasting models	ML algorithms	Accurate seizure prediction	Absence of pediatric focus	Inspires biosignal analysis
13	Nakua	Neural correlates in pediatric psychopathology	Neuroimaging	Modeling analysis	Statistical ML	Identified brain-behavior links	Lack multimodal inclusion	Informs neural data integration
14	Ouyang et al.	Nebulizer design for pediatrics	Product design	AHP-DEMATEL-TRIZ	Hybrid decision method	Optimized device usability	No emotional factor study	Guides device adaptation
15	Pilli et al.	Brain age estimation	Neuroimaging	Kernel regression	Universum KR-VFL	Accurate age prediction	Lacks pediatric specificity	Inspires biomarker modeling
16	Reyes-Martín et al.	Challenging behavior in disabilities	Behavioral data	Systematic review	Qualitative synthesis	Summarized behavioral strategies	Missing multimodal signals	Context for behavioral layer
17	Ruble et al.	Caregiver-teacher alliance	Survey data	Relationship analysis	COMPASS program	Stronger alliance improves outcomes	Lack AI integration	Supports caregiver dimension
18	Saban-Bezalel	Communication profiles in delays	Observational data	Comparative analysis	Behavioral profiling	Differences in delay vs peers	No tech-based solution	Aids communication feature design
19	Sangeetha et al.	Emotion recognition in children	Multimodal	Systematic review	Review synthesis	Identified multimodal SOTA	Lacks experimental results	Validates multimodal choice
20	Tu et al.	Communication on parental bipolar disorder	Interview data	Qualitative analysis	Thematic coding	Identified child communication needs	Missing tech intervention	Adds caregiver-child context
21	Usman et al.	Brain age estimation	Neuroimaging	Multitask learning	Adversarial VAE	Accurate biological age	No pediatric focus	Inspires deep multimodal fusion
22	Wu et al.	Lifespan brain age prediction	Neuroimaging	Comprehensive review	ML & DL	Overview of prediction methods	Missing pediatric subset	Provides baseline algorithms
23	Yang et al.	Anthropomorphism in AI assistants	Behavioral data	Experimental study	Feedback-based AI	Enhanced engagement & emotion	Needs multimodal testing	Guides feedback mechanisms
24	Yin	Aging via speech analysis	Audio data	Linguistic modeling	ML-based speech models	Age cues from speech	Missing physiological data	Adds speech feature set

3. Research Methodology

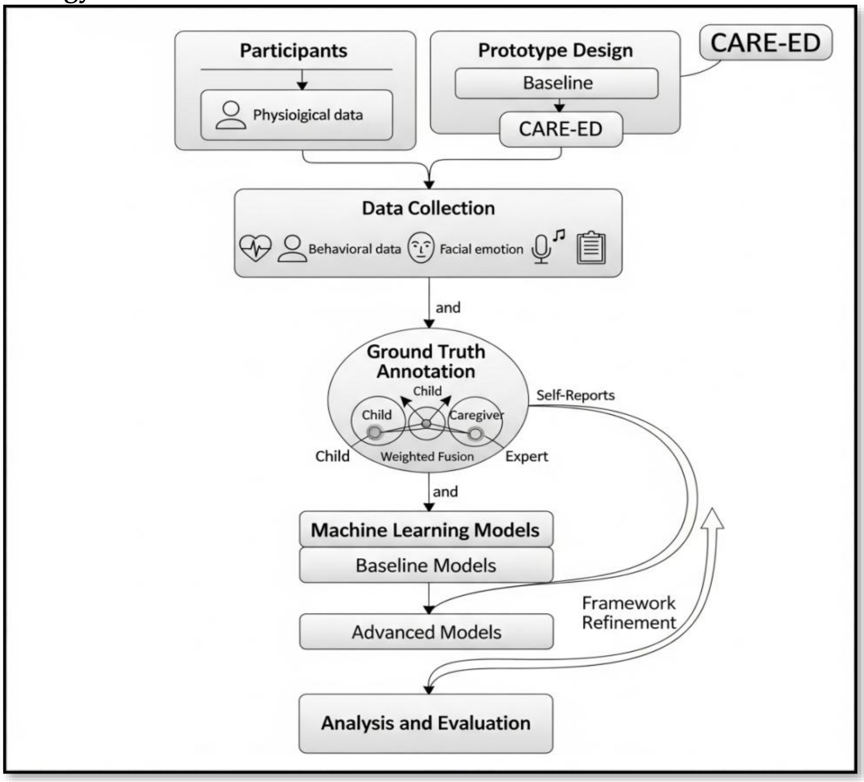

Figure 3: CARE-ED Research Methodology Flow Chart

Figure 3 shows that the study outlines and tests the CARE-ED (Child-Caregiver Age-Adaptive Emotional Design) framework personalized for pediatric home medical devices from having developed the result.

3.1 Overview

The research methodology involves the use of a systematic and sequential work-process which can include a variety of stages involving initial design, development of prototypes, pilot testing, home deployment, collection of multimodal data, annotation of ground truth, modeling with machine learning, analysis of statistics, and dyadic analysis, after which the framework would be iteratively improved. This intensive process will comprehensively represent the concept of usability and emotional engagement in the overall construct of the measurement of the pediatric-caregiver dyad in terms of its applicability and effectiveness in real healthcare environments.

The combination of qualitative and quantitative methodology, with which this approach to the methods comes, is a complex method that is constructed to allow very strong validation of the CARE-ED framework. The study is able to capture a wide range of user behavior and emotional response to a wide range of situations through controlled laboratory tests as well as naturalistic home deployments. The multimodal data gathering process generally involves the utilization of behavioral logs, physiological fluctuations, and self-reported data will enable the examination of the complicated interpersonal circumstances between children and caretakers in a thorough manner. Additionally, the process of iterative improvement based on a continuous feedback loop insinuated that the structure responds and adapts itself to the demands of its user, and the advocacy of the user-centered design in pediatric medical technology ensued.

3.2 Participants and Sampling

The sample size represented in a pair of samples: n = 500 children ages 5-12 years old and their proximate caregivers (n = 500) in the study. Dyads were stratified into three groups corresponding to pediatric age categories of 5 to 7 years, 8 to 10 years, and 11 to 12 years of age in order to consider the differences in development and the psychology of the caregiver. Caregivers were also grouped on the basis of the self-reported psychological states, whereby they were divided into confident, anxious, and stressed profiles. Participants were required to meet inclusion criteria, including a stable internet connection at home, the willingness and the capacity to understand and abide by the usage instructions of the device provided, and the rules of informed consent and assent as per ethical protocol. This method of stratified sampling made cross-representation of the different stages of development and caregivers' emotional context, and this has increased the level of generalizability of the findings.

3.3 Prototype Design and Experimental Setup

Two prototype designs were set to particular expectations in order to determine the dissimilarity in the usability and emotional activities:

3.3.1 Baseline Prototype

This is a traditional usability-based interface that was donned with a neutral and standard interface. It preferred the simplicity of functionality, visual distinction, and navigation uncomplicatedness to resemble the average home medical devices, which are currently in use. The Baseline, which was used as the control condition to gauge the improvement provided by the CARE-ED framework, was crammed.

3.3.2 CARE-ED Prototype

It is an advanced interface that possesses age-modulating characteristics of being flexible and naturally apt to the age of the child. It was implemented with gamified micro-interactions of interactive animation, rewards, and prompting to keep the pediatric interaction and compliance. At the same time, the design included stress-reduction cues and feedback tools that were targeted to the group of caregivers in order to appreciate their significance and influence on the device usage and emotional self-regulation in the dyad.

These prototypes were initially presented to the respondents within a home-like laboratory environment, which simulated conditions of living and, therefore, made it possible to observe the behavior of users in detail and report information about the usability level of specific products and emotional response. The longitudinal home use lasted 4 to 12 weeks in the form of the introduction of devices, so that ecological data could be gathered, which would be seen as an index of true patterns of interaction, adherence rates, and emotional dynamics present in a truly naturalistic context. Such a two-phase experiment protocol allowed ensuring that initial usability and prolonged user experience were given full consideration to contribute to enhancing the CARE-ED framework on the basis of iterative improvements.

3.4 Hypotheses

On the basis of the theoretical principles of emotional design and pediatric developmental psychology, the following hypotheses are explored by the study:

H1: With respect to proficiency, the CARE-ED prototype will significantly exceed the measures of usability (e.g., task completion time, error rates) of all pediatric age groups relative to the Baseline prototype.

H2: Pediatric participants exposed to the CARE-ED prototype will have increased engagement and longitudinal continued use of the prototype over the period it is deployed in the participant's home compared to use of the Baseline prototype.

H3: Caregivers who used the CARE-ED prototype will report a reduction in stress levels as well as device usage satisfaction relative to their counterparts who interacted with the Baseline prototype.

H4: Multimodal data analysis of effects of CARE-ED will demonstrate more positive dyadic emotional synchrony (between children and caregivers) generated by the age-adaptive, gamified implementation.

The above are the hypotheses based on which the empirical functional effectiveness of the CARE-ED framework to improve not only the usability of the environments but also the emotional experience of interactions with home medical devices in pediatric settings will be tested.

3.5 Data Modalities and Synchronization

The CARE-ED dataset includes a diversity of synchronized multimodal data streams, a measure of the real-time pediatric-caregiver interaction with home medical devices. The multimodal approach will ascertain a complete understanding of physiological and behavioral responsiveness that will bring a granular perspective on usability and emotional processes in the caregiving dyad.

3.5.1 Physiological Signals

Key autonomic variables that have shown strong effects during conditions of high stress were monitored continuously, including heart rate in beats per minute (BPM), and Heart Rate Variability (HRV) decomposed to both time and frequency domains to account for subtle changes in autonomic nervous system functioning. The temperature of the skin was measured in degrees Celsius to identify the thermoregulatory adjustments that were related to stress or relaxation. Peaks of Electrodermal Activity (EDA) were detected in order to measure sympathetic nervous system arousal, which was the objective measure of affective and cognitive load during interaction with the device.

3.5.2 Behavioral Metrics:
Recordings of the user performance data were performed decisively and included the following: The success rate of the tasks (fraction of the correctly completed operations), the time of doing the tasks (the time taken to complete the tasks, measured in seconds), and the number and type of errors committed when using the device. These measures allowed objective measurement of usability and allowed comparison of the Baseline and CARE-ED prototypes.

3.5.3 Facial Emotion Features:
Future-generation computer vision algorithms were used to obtain a per-frame probabilistic estimation of six basic emotional states: happiness, sadness, neutral, surprise, fear, and anger based on video recordings. A normalisation of the sum of all emotional probabilities of every time point was limited to one, as Equation 1 formalised:
$\sum c \in C\, P_c = 1$ …. (1)
where P_c denotes the probability of emotion class c within the set C.
This allowed for continuous tracking of subtle emotional shifts throughout the interaction sessions.
Audio Features
Acoustic processing produced Mel-frequency cepstral coefficients (MFCCs), pitch, and spectral flux as well as energy measures and prosodic contours. These acoustic characteristics thoroughly helped to complete the measurement of the affective states and stress detectors that complemented visual information on emotions with self-reporting.

Self-Reports
The gamified interfaces and standard questionnaires were used to acquire the subjective data. The mood of the pediatric participants was found on a 0 to 4 slider, which is graphically entertaining and allows them to express their mood intuitively. Stress rating was carried out using a Likert scale with caregivers giving a diagnosis in a range of 1 to 5 (higher numbers indicating more perceived emotional loads when operating the device).
The synchronization of all multimodal data streams was strictly done under Network Time Protocol (NTP) to make the analyses of data really synchronized over time, so all modalities of data could be analyzed together. Before the analysis, data were anonymized and encrypted in order to fulfill ethical standards and keep the participants confidential. The data set was uploaded and locked in a secure cloud repository where only authorized individuals would have access to the information to encourage regional collaboration in research.

3.6 Ground Truth Annotation
A multi-source weighted fusion method was used to determine reliable and valid emotional ground truth labels of the CARE-ED dataset. By combining subjective and objective points of view, the approach helped to capture the subtle emotional conditions of pediatric participants and their caregivers when using the device,
as clearly depicted in

Figure 4:

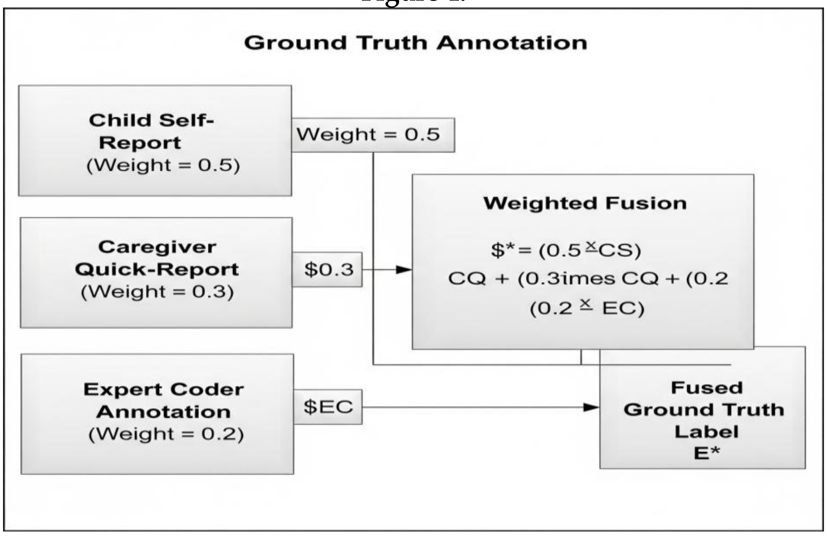

Figure 4: Ground Truth Annotation: Weighted Fusion
Child Self-Report (Weight = 0.5)

The child was the primary source of the emotional ground truth, which was collected as a self-assessment of their own affect by an intuitive gamified mood reporting interface right after the use of the device. Due to the first-hand experiential knowledge of the pediatric participants, this source was ranked as the most weighted one and was attributed to the role of representing the true emotional conditions from the child's point of view.

Caregiver Quick-Report (Weight = 0.3)

Emotional evaluation by caregivers was implicitly accomplished by the usage of brief, standardized questionnaires based on emotional assessment urgently after the interaction ended, in addition to the self-report of the child. This input by a caregiver gave an outside perspective on the emotional state of the child and contextual factors that would affect emotional reactions. The perspective of the caregiver was given medium weight to recognize the importance of the role, but at the same time, its indirect nature.

Expert Coder Annotation from Video (Weight = 0.2)

Facial expression coding systems designed to measure facial expression similar to emotional responses within large groups of subjects were assessed with validated facial expression coding systems using trained expert coders to perform detailed analyses with video of the specific emotional responses identified and classified with the facial expression coding system during the device interaction sessions. This subjective observation report was supported and complemented by the objective observation data, where even nonverbal emotional clues may not have been reported or were oblivious to them as subjects.

These three sources were fused by a weighted sum to give the final ground truth label of each emotional episode and hence a balanced, multidimensional representation of emotional states. This both improved the robustness of annotations by eliminating biases that are associated with any individual source and also allowed for more complete grounds for further machine learning modeling and statistical work.

The fused label E^* was computed and given in **Equation 2:**

$$E^* = arg\ max_{c\in C}(0.5 \cdot P_{child}(c) + 0.3 \cdot P_{care}(c) + 0.2 \cdot P_{expert}(c)) \ldots. (2)$$

where C is the set of predefined emotion categories.

Inter-rater reliability between expert coders was calculated using Cohen's κ to ensure consistency.

3.7 Feature Extraction and Preprocessing

Figure 5 reflects the step-by-step process of getting raw data through the process of creating final prepared features. It goes through a process where noise is removed, normalized, and transformed to work best in the model.

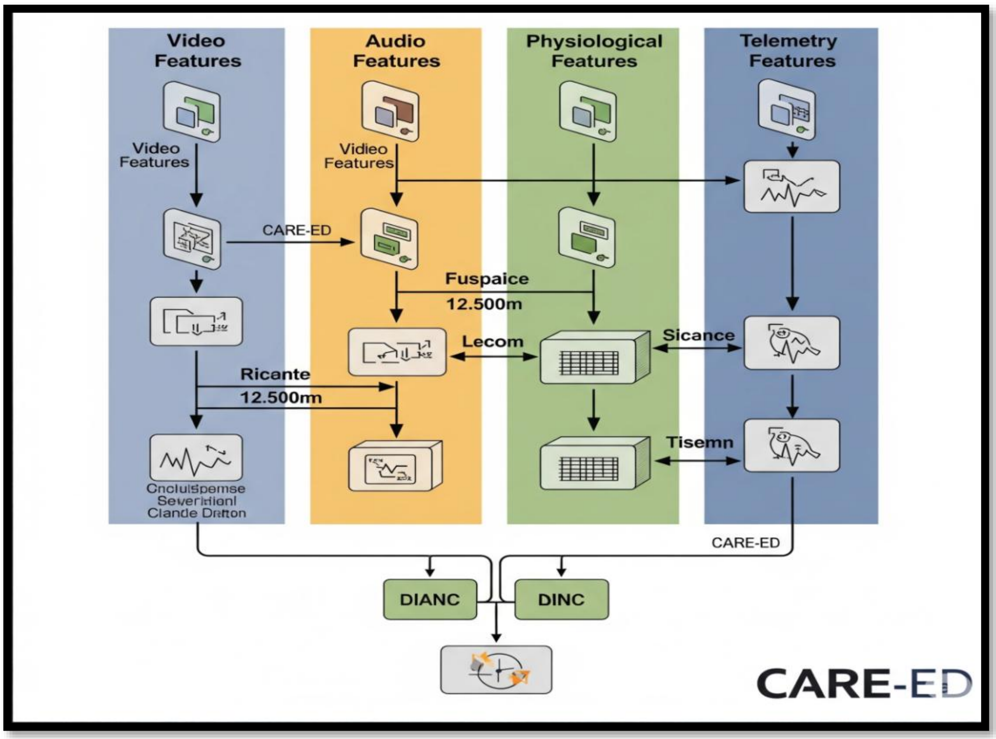

Figure 5: Feature Extraction and Preprocessing Flow

3.7.1 Video Features

The facial data were processed frame-wise in order to obtain primary indicators of emotional and attentional conditions. This comprised the detection of facial landmarks, recognition of facial Action Units (AUs) as specified in the Facial Action Coding System (FACS), and estimating parameters of head pose. The potential convolutional neural network (CNN) architectures, e.g., ResNet and MobileNet, were used to leverage feature extraction by providing a strong representation of such subtle nuances in facial expressions and micro-movement faces. Segmented interaction features were pooled statistically to obtain temporal variations and general expression dynamics across the intervals of device use.

3.7.2 Audio Features

Voiced segments of acoustic signals were identified to determine specific features accurately. Prosodic and spectral features were extracted that focused on the characteristics of emotional state and speech patterns that can be determined by mel-frequency cepstral coefficients (MFCCs), spectral flux, pitch, and signal energy. The temporal dependencies, along with the dynamic variation of the audio features, were modeled as Long Short-term Memory (LSTM) networks and provided a sequence-sensitive representation to which the affective analysis processes could be applied.

3.7.3 Physiological Features

To test the autonomic nervous system changing as a result of stress and relaxation response, the variables of the heart rate (HR) and heart rate variability (HRV) were calculated in the time domain and frequency domain. Butterworth bandpass filtering was used to de-noise the electrodermal activity (EDA) signal by removing noise and isolating those phasic components that are relevant. Interaction episodes in the filtered EDA were segmented as peaks as an indication of sympathetic arousal.

3.7.4 Telemetry Features

As objective measures of usability, behavioral telemetry that measures the time taken to complete a task, the rate of task completion success, and error count were gathered. MinMax scaling was used to normalise these features to a 0-1 range so that they can be integrated with other modalities in modeling and in statistical computations, as in **Equation 3:**

$$x^1 = \frac{x - x_{min}}{x_{max} - x_{min}} \ldots . (3)$$

3.7.5 Missing Data Handling

Missing values imputed using **K-nearest neighbor imputation** (K = 5), stratified by age group to preserve developmental trends.

3.8 Machine Learning Models

The article was also analyzed using a complete set of machine learning models that considered the several modalities of the data and conclusions to predict the outcome of emotional as well as usability of the interaction of home medical devices with pediatricians.

3.8.1 Baseline Models

First exploratory analyses have used conventional classifiers such as Logistic Regression (LR), Support Vector Machine (SVM), k-Nearest Neighbor (kNN), Naive Bayes (NB), and Decision Tree (DT). These models served as reference baselines, and they gave interpretable performance, which can establish a standard baseline where any classification operation involving the prediction of emotional state and usability can be very easily executed.

3.8.2 Advanced Models

As a way of revealing complex, nonlinear relationships that are inherent in multimodal data, optimized ensembles and deep learning network designs were used. Tree-based ensemble models, Random Forest (RF) and Gradient Boosting Machine (GBM), delivered solid results that could manage heterogeneous feature spaces and resolved overfitting. Convolution Neural Networks (CNN) were employed as feature extractors of time and location within video data, and allowed automatic detection of the hierarchy of expressive visual features and behavior. Streams of physiological and audio features were employed using LSTM networks, which were able to successfully learn the temporal dynamics that form the key to recognizing emotional states. Also, multimodal fusion with Transformers was used to combine heterogeneous data modalities by using attention mechanisms that learned to give varying importance to features corresponding to child and caregiver streams.

3.8.3 CARE-ED Model Architecture

The CARE-ED model improves on earlier approaches to dimensional modeling by limiting the predictive model to only three factors, and these factors represent the conceptual dyadic intercourse as shown in

Figure 6:

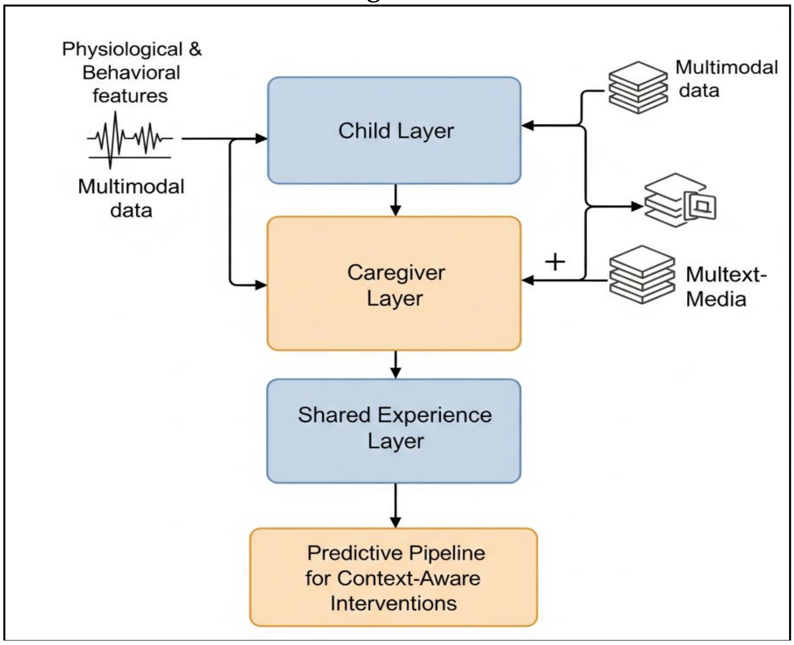

Figure 6: CARE-ED Model Architecture

Child Layer: Paying attention to age-adaptive gamification interaction characteristics, the creation of modeling engagement and emotional reactions in conditions characteristic of the stages of development.

Caregiver Layer: Will integrate stress-mitigating feedback cues, which reflect states of care providers regarding their affect and the impact it has on the care provider's environment.

Shared Experience Layer: Combines the multimodal data to create the model of the dyadic emotional context, taking advantage of the fused representations that provide the synchronic interactions between a child and a caregiver.

This layered architecture can help understand both the complex processes of individual and collective emotional and behavioral functioning on a detailed level, as well as make personalized and situationally informed interventions based on the CARE-ED framework possible.

3.9 Mathematical Formulation

Notation

N —*total samples (timestamped interaction windows), indexed by* i.

M — *number of modalities (e.g., physio,video,audio,self-report).*

C —*number of discrete emotion classes (or use* y_r *for continuous valence/ arousal).*

Raw modality input for sample i: $raw_i^m, m \in \{1, \dots \dots, M\}$.

Processed feature vector for modality m: $x_i^m = \phi_m(raw_i^m)$.

Multimodal input vector: $x_i = \{x_i^1, \dots \dots, x_i^M\}$

Label: $y_i \in \{1, \dots \dots, C\}$*(classification) or* $y_i \in \mathbb{R}$ *(regression).*

Model parameters: $\theta = \{\theta_1, \dots \dots, \theta_M, \theta_{fusion}, \theta_{head}\}$

Child/ caregiver id: $u(i)$*(used for personalization embeddings* e_u*).*

3.9.1 Preprocessing / Feature maps

The process of transforming the raw content of different sensors or other input data into a meaningful and structured form comprehensively understood by machine learning models is what requires the necessary preprocessing stage. Corresponding to each modality, there is some preprocessing function (e.g., video, audio, physiological) ϕm that extracts useful information out of raw measurements. It removes the noise, normalizes the data, and converts the data to meaningful and homogenous numeric forms. On physiological signals, it will pass (e.g., filter bandpass) the noise, then peaks will be detected (e.g., the beats) so as to obtain measurements like heart rate (HR) and heart rate variability (HRV). On video, it crops the face, locates facial landmarks (salient landmarks like the corners of the mouth and the eyes), and discovers frame-level descriptors that define expressions or object motion. In an audio case, it detects voiced area (speech), and proceeds to perform a calculation of spectral feature, such as Mel-frequency cepstral coefficients (MFCCs), which defines a feature sound. When data is time-varying, these features can be described along sequences of temporal windows with a fixed time window T and one may study how the features change over time. Consider, in each modality, fixing a deterministic preprocessing/ feature extraction map as in **Equation 4:**

$x_i^m = \phi_m(raw_i^m), \qquad m = 1, \dots\dots, M. \dots (4)$

Physio: $\phi_{physio} = bandpass + peak\ detection \Rightarrow HR, HRV, EDA\ peaks.$
Video: $\phi_{video} = face\ crop \rightarrow landmark\ extraction \rightarrow frame\ descriptors.$
Audio: $\phi_{audio} = VAD \rightarrow MFCCs + prosody$

If using temporal windows of length T, form sequences can be as shown in **Equation 5:**

$$X_i^m = \left[x_{i,t-T+1}^m, \dots\dots, x_{i,t}^m\right]$$

3.9.2 Modality encoders (per-modality embeddings)

Each modality has an encoder $E_m(\cdot, \theta_m)$that maps preprocessed input to an embeddingz_i^m.

$z_i^m = E_m(X_i^m; \theta_m), \quad z_i^m \in \mathbb{R}^{d_m} \dots (5)$

Video encoder: $E_{video} = LSTM\big(CNN(frames)\big) \cdot Formally$:

$z_i^{video} = LSTM(CNN(frames); \theta_{video}) \dots (6)$

Physio-encoder: $1{:}\,D\ CNN \rightarrow Temporal\ pooling$

$z_i^{physio} = Pool(Conv1D(X_i^{physio}) \dots (7)$

Equations 5, 6, and 7 denote that this is a learned representation of this vector, which represents the most informative information of an input to subsequent processes. With video, convolutional neural networks (CNNs) are used to first learn spatial information on a frame-wise level and recurrent neural networks, such as LSTMs, to learn temporal dynamics on sequences of frames. In physiological data, 1D CNNs find patterns in time series (e.g., abrupt HR change) and pool data over time to a fixed-length vector. The encoder will therefore represent complex and high-dimensional raw data into a more useful, low-dimensional summary that will be used to fuse with other modalities.

3.9.3 Fusion strategies

The representation serves to combine the information, which is shown by the various modalities, to take the advantages gained by different sources of data as much as possible to maximize the possibility of obtaining the right answer in prediction.

Concatenation (Late Fusion): Just concatenates the embeddings of each modality into a single long embedding. The model then learns from this combined representation. This is simple yet does not prioritize the modalities, unless they are distinguished.

Attention-based Fusion: rains to assign different attention to the different modalities per sample and pay more attention to those modalities, which are more informative/reliable at that sample. The significant scores that you give the embeddings are dynamically weighted either by adding more or less importance to particular embeddings, and finally merged into one output vector. This will permit model robustness, especially in cases where some modalities could be either too noisy or unimportant.

Two common, mathematically explicit choices are as follows:

3.9.3.1 Concatenation (late fusion)

Concatenate modality embeddings, and then predict as shown in **Equation 8**:

$z_i = concat(z_i^1, \dots\dots, z_i^M), \qquad \widehat{p_i} = softmac(W_{head z_i} + b_{head}). \dots (8)$

3.9.3.2 Attention-based modality weighting (learned fusion)

Compute pre-modality scores and softmax them to get attention weights α_m as shown in **Equations 9, 10, and 11:**

$s_{i,m} = v^T \tanh(W_m z_i^m + b_m), \qquad \alpha_{i,m} = \frac{\exp(s_{i,m})}{\sum_{m'=1}^{M} \exp(s_{i,m'})} \dots (9)$

Fused Embedding:

$z_i = \sum_{m=1}^{M} \alpha_{i,m} z_i^m \dots (10)$

Prediction:

$\widehat{p_i} = softmax(W_{head} z_i + b_{head}) \dots (11)$

3.9.4 Prediction heads and outputs

The prediction head translates the fused embedding into task-specific outputs. This can be classification (predicting discrete labels) or regression (predicting continuous values).

3.9.4.1 Classification (C classes):

To create a probability distribution of possible classes, it uses a softmax function as output. The output is the predicted class that has the largest probability in **Equation 12:**

$\hat{p}_i = P(y = c|x_i; \theta), \qquad \hat{y}_i = \arg\max_c \hat{p}_{i,c} \dots (12)$

3.9.4.2 Regression (e.g., valence/arousal):

Regression is a direct mapping-based embedding returning a numerical value (e.g., emotional valence or arousal score). That modular architecture makes the model flexible and applicable to different prediction problems as depicted in **Equation 13:**

$\hat{y}_i^{(r)} = f_{reg}(z_i; \theta_{reg}), \qquad f_{reg} \in \mathbb{R}$ …. (13)

3.9.5 Loss functions (single-task and multi-task)

It is a quantified difference between estimates of the model and the real numbers during the training. Minimizing loss helps to improve the model's accuracy.

3.9.5.1 Cross-entropy loss for classification:

It helps to penalize incorrect class predictions by measuring how well predicted probabilities match true classes as given in **Equation 14:**

$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} 1\,[y_i = c] log\hat{p}_{i,c}$ … (14)

3.9.5.2 MSE loss for regression (valence/arousal):

Mean Squared Error (MSE) measures squared differences between predicted and true continuous values, encouraging predictions close to true scores as shown in **Equation 15:**

$\mathcal{L}_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i^{(r)} - \hat{y}_i^{(r)})^2$ …. (15)

3.9.5.3 Combined multi-task loss (classification + regression):

Multi-task loss combines both classifications as well as regression losses, which allows the model to simultaneously optimize for multiple objectives, weighted by λ_{cls} and λ_{reg} as shown in **Equation 16:**

$\mathcal{L} = \lambda_{cls}\mathcal{L}_{CE} + \lambda_{reg}\mathcal{L}_{MSE}$ … (16)

3.9.5.4 Regularized objective (L2):

This regularized objective function is L2 in that it considers an additional penalty when the model parameters have large squared magnitudes. The method discourages huge weights, even completely, hence lowering overfitting and at the same time enhancing understanding. The strength of the regularization effect is more or less adjustable by the parameter, λ, as in **Equation 17:**

$\min_{\theta} \mathcal{L}(\theta) + \lambda\,||\theta||_2^2$ …. (17)

3.9.6 Personalization

A general population-trained model for individual users has been adopted by this personalization, which predicts the relevance for their unique data patterns.

3.9.6.1 Fine-tuning (per-user)

The base model parameters θ are adjusted by further training on a small dataset specific to user u. This adapts the model to the user's data distribution as well.

Given a base model θ trained on population data, fine-tune on user u small dataset $\mathcal{D}_u$, as shown in **Equation 18:**

$\theta_u = \theta - \eta\nabla_\theta\mathcal{L}_{\mathcal{D}_u}(\theta)$ …. (18)

Use θ_u for predictions on that user.

3.9.6.2 Learned user embedding (cold-start friendly)

Learned user embeddings: Instead of fine-tuning all parameters, the model learns a fixed-size embedding vector $e\ u$ per user. This vector is concatenated with the data embedding as well as it is used for prediction. It is a lightweight method that is highly suitable when little user data is available (cold-start).

Introduce a learned embedding $e_u \in \mathbb{R}^d$ per user and condition the head on it:

$z_i' = concat(z_i, e_{u(i)}), \qquad \hat{p_i} = softmax(W_{head}z_i' + b_{head})$ …. (19)

Regularize embeddings:

$\mathcal{L}_{total} = \mathcal{L} + \lambda_e\sum_u ||e_u||^2$ …. (20)

Equation 19 and 20 aim to capture user-specific nuances, improving personalized performance.

3.9.6.3 Meta-learning (MAML style)

Meta objective to enable fast personalization, as given in **Equation 21:**

$\min_{\theta}\sum_u \mathcal{L}_{\mathcal{D}_u^{val}}(\theta - \alpha\nabla_\theta\mathcal{L}_{\mathcal{D}_u^{tr}}(\theta))$ …. (21)

3.9.7 Explainability

The explainability approaches provide signals, which are mainly about the characteristics or modalities that thoroughly influence the model predictions, which is a very important necessity in sensitive applications, both for trust as well as validation.

Attention weights display which of the modalities each prediction relies on the most, or where the prediction model focuses when making a decision. Using cooperative game theory, SHAP values ϕj of each input feature j approximate how much each individually contributes to the prediction by computing the difference in the outputs with and without feature j. Grad-CAM produces a visualization of gradients that show the areas of input images that have the most influence in the classification by overlaying the gradients on feature maps in CNN layers. These tools are thoroughly used to interpret model behavior and periodically find out biases or failure modes.

Attention weights $\alpha_{i,m}$ provide modality importance per sample.

SHAP values ϕ_j for feature j, as shown in **Equation 22:**

$$\phi_j = \sum_{S \subseteq F\setminus\{j\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!}[f(S \cup \{j\}) - f(S)] \;…. (22)$$

Grad-CAM for visual branches: compute gradients of class score w.r.t last conv feature maps and combine.

3.9.8 Optimization & training recipe

During training, optimization algorithms cycle through the update of model parameters in order to further reduce loss. It reflects the Adam or AdamW optimizer, which implements adaptive learning rate scaling, to make convergence faster and stable. Typical hyperparameters include the learning rate η, the size of a batch (the number of samples to update with), and the weight decay to regularise λ. Dropout and early stopping are the types of methods used to circumvent overfitting by randomly eliminating nodes on each training iteration and halting when a training method no longer decreases in a validation loss. Such an arrangement leads to very effective and efficient learning of deep models.

Optimizer: Adam (or AdamW). Parameter update has been given in **Equation 23:**

$\theta \leftarrow \theta - \eta\hat{g}, \qquad \hat{g} = AdamGrad(\nabla_\theta \mathcal{L})$ …. (23)

Typical hyperparameters: $\eta \in [10^{-4}, 10^{-3}], batch\ size\ 16 - 128, weight\ decay\ \lambda \approx 10^{-4}$

Regularization: Dropout, early stopping on validation loss. .9.9 Evaluation metrics

Metrics show the effectiveness of the model on unseen data and measure various aspects of the quality of predictions as provided in **Table 2:**

Table 2: Evaluation metrics and its purpose

Metric	Short Definition	Purpose
Accuracy	% correct predictions	Overall performance measure
Precision	Correct positive predictions	Avoid false positives
Recall	Found actual positives	Avoid missed positives
F1-Score	Balance P and R	Handles class imbalance
ROC-AUC	Separates classes well	Threshold-independent score
RMSE	Avg. prediction error	Regression accuracy
Pearson	Linear correlation strength	Measures value agreement
ECE	Confidence vs. accuracy gap	Checks probability calibration

These metrics provide a comprehensive picture of model strengths and weaknesses.

3.9.9.1 Accuracy

It is a very important indicator of the frequency with which the model predicts in all cases against the actual results. It is calculated by dividing the correct number of predictions by the total number of predictions. Greater accuracy implies that the model works effectively in general, yet it is possible that the model will fail to acknowledge class imbalances provided in **Equation 24:**

$Accuracy = \frac{1}{N}\sum_{i=1}^{N} 1(\hat{y}_i = y_i)$ …. (24)

Where
N =Total number of samples
$\hat{y}_i$=Predicted label for sample
y_i=True label for sample
$1(\cdot)$= Indicator Function (1 if condition is true, 0 otherwise)

3.9.9.2 Precision (PPP)

It calculates the accuracy of the positive predictions made by the model. It is calculated as the division of the predicted positive cases that are correctly determined by the total number of predicted positive cases. This measure is also critical, in cases where the cost of false positives is great, as stipulated in **Equation 25:**

$Precision = \frac{TP}{TP+FP}$ …. (25)

Where,
TP= True Positive
FN= False Positive

3.9.9.3 Recall

It assesses the ability of the model to recognise all the existing relevant cases that are positive. It is the comparison of the right positive identification and the total actual positives. Because fine performance is dominated by the number of missed positive instances, a high recall, in sensitive applications, is worthwhile, as in **Equation 26:**

$Recall = \frac{TP}{TP+FN}$…. (26)

Where,
TP= True Positive
FN= False Negative

3.9.9.4 F1 (per class):
It points to the precision and recall as one of the performance measurement instances. It is determined by their harmonic mean, which leads to an equal evaluation. That is the F1-Score, which comes in handy when the classes are not equally distributed, as in **Equation 27:**

$$F1 - Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad .... (27)$$

3.9.9.5 **ROC-AUC**
It quantifies how well the model separates positive and negative classes. It is the total area under the Receiver Operating Characteristic curve. A higher AUC indicates better classification performance across different thresholds. It is the area under the ROC curve for binary; use one-vs-rest for multi-class, as given in **Equation 28:**

$$ROC - AUC = \int_0^1 TPR\left(FPR^{-1}(x)\right)dx \quad .... (28)$$

Where;
$TPR = \frac{TP}{TP+FN}$ (True Positive Rate)
$FPR = \frac{FP}{FP+FN}$(False Positive Rate)

3.9.9.6 Processing Time
Represents the average duration needed to make a prediction. It is acquired by dividing the overall time by the predictions. In applications of real-time or high-speed processing, the time must be lower, as given in **Equation 29:**

$$Processing\ Time\ (ms) = \frac{Total\ Time\ Take\ (ms)}{Number\ of\ Predictions} \quad .... (29)$$

3.9.9.7 Model Complexity
It implies the model size in terms of parameters computationally. It is counted in millions in order to represent big networks simply. Greater complexity can provide more accurate estimates, but can also require a higher number of computational resources as provided in **Equation 30:**

$$Model\ Complexity\ (\ Million\ Params) = \frac{Total\ Parameters}{10^6} \quad .... (30)$$

3.9.10 Validation & statistical testing
Validation and statistical tests rigorously assess model generalizability and the significance of improvements.

Cross-validation / LOSO: Model trained on all subjects except one and tested on the left-out subject; repeated to average performance, ensuring unbiased evaluation.

Paired t-test: Tests whether the mean difference $\bar{d}$ in performance metrics between two models (e.g., baseline vs CARE-ED) is statistically significant, accounting for variance s_d and sample size n as revealed in **Equation 31:**

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} \quad .... (31)$$

Cohen's d: Measures effect size, quantifying the practical significance of the difference, as shown in **Equation 32:**

$$d = \frac{\bar{d}}{s_d} \quad .... (32)$$

Mixed-effects models: Handle repeated measures or longitudinal data by modeling both fixed effects (predictors) and random effects (individual subject variability), as given in **Equation 33:**

$$y_{it} = X_{it}\beta + Z_i b_i + \epsilon_{it}, \qquad b_i \sim \mathcal{N}(0, \Sigma_b), \epsilon_{it} \sim \mathcal{N}(0, \sigma^2) \quad .... (33)$$

Dyadic Actor-Partner Interdependence Model (APIM): Models how each member of a dyad (child and caregiver) influences their own and their partner's outcome, as shown in

Equation 34:

$$\begin{cases} Y_{child} = \beta_{ac} X_{child} + \beta_{pc} X_{care} + \epsilon_c, \\ Y_{care} = \beta_{ap} X_{care} + \beta_{pp} X_{child} + \epsilon_p, \end{cases} \quad .... (34)$$

Where β_{ac} and β_{pp} represent actor effects, β_{pc} and β_{ap} partner effects.

3.9.11 Sample-size formula (paired test)
The calculation of sample size is always a very important aspect of study planning so that the study has adequate statistical power to provide a significant effect as well. In paired designs, when the same subjects will be measured twice (e.g., under baseline, CARE-ED), the sample size would rely on the anticipated mean distinction Δ, the variability of differences σd, desired significance level α, and power $1 - \beta$.

Equation 35 estimates how many participant pairs are needed to confidently detect a true effect Δ with a controlled probability of Type I error (α) and Type II error (β). Underestimating n risks missing significant findings; overestimating wastes resources.

For detecting mean difference Δ with paired design, two-sided α, power 1-β, and SD of differences σ_d, as given in

Equation 35:

$$n \approx \left(\frac{Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}}{\frac{\Delta}{\sigma_d}} \right)^2 \quad \dots . (35)$$

Where:
n= required number of pairs (subjects)
Δ= minimum detectable mean difference between paired measurements
σ_d= standard deviation of the difference between paired observations
$Z_{1-\frac{\alpha}{2}}$ = critical value from the standard normal distribution for significance level α (two-sided test)
$Z_{1-\beta}$= critical value corresponding to desired power $1-\beta$

3.9.12 Validation and Evaluation

Validation Schemes

Leave-One-Subject-Out (LOSO): Every other subject is cross-validated, leaving one subject, and subjects are trained, subject by subject, on the remaining subjects. This encourages a robust generalization assessment, which applies especially in those cases where there is a very high variability.

K-fold Cross-Validation: In this, the data is partitioned into k-folds (here=5, in this case), of which (k-1) are used to train the model, with the remaining used to test. Stratification by age ensures balanced representation across pediatric groups.

Temporal Holdout: When in longitudinal data, we have used the first time points to train and subsequent ones to evaluate the extent to which the model performs well over time.

Evaluation Metrics

- Accuracy: Overall correctness of categorical predictions.
- Macro F1-score: The average F1-score of the classes, but all classes are effectively equal; it is ideally suited when working with imbalanced classes.
- RMSE: It measures the level of the errors of continuous prediction outcomes in a mean way.
- ROC-AUC: Evaluates a classifier's ability to distinguish classes.
- Expected Calibration Error (ECE): This checks the effectiveness of the predicted possibilities in correspondence to the observed outcomes.

Statistical Analysis

Paired t-test/or Wilcoxon signed-rank test: The performance parameters of the same subjects are to be compared as the subjects respond to two conditions (parametric or non-parametric).

ANOVA or Kruskal-Wallis test: Relationship between the outcomes of two or more groups (smart or old age, profiles of caregivers).

Actor-Partner Interdependence Model (APIM): This is a statistical model in which the statistical dependence between a member of a dyad and their emotional or behavioral state is estimated, and it represents how this state changes the partner.

The main points of the CARE-ED framework are presented in **Table 3**, which describes the design, profile of participants, the setup of prototypes, types of data, the extraction of features, machine learning models, and the models of optimization.

Table 3: CARE-ED Framework: Design, Implementation, and Ethics Overview

Section	Key Aspect	Description	Data Type	Modality	Model Type	Evaluation
Overview	Framework Design	CARE-ED framework proposal	Multimodal	Video, Audio, Physio	ML & Statistical Models	Iterative refinement
Participants	Sample Size	1000 paired pediatric-caregiver	Demographic	Pediatric, Caregiver	Stratified sampling	Age groups, profiles
Prototype Design	Two prototypes	Baseline & CARE-ED	Device interaction	Usability, Emotional	Controlled lab & home usage	Longitudinal (4-12 weeks)
Data Modalities	Multimodal Data	Physio, Behavioral, Audio	Heart rate, MFCCs	Physio, Video, Audio	Time-synced features	Secure cloud storage
Ground Truth	Label Fusion	Weighted child, caregiver, expert	Emotional states	Self-report, Video	Fusion of multiple sources	Weighted labels
Feature Extraction	Multimodal Processing	CNN, LSTM, Filtering	Video, Audio, Physio	Deep feature maps	Normalization, filtering	Temporal windows
Machine Learning	Models Used	Logistic regression to CNN	Classification, Regression	Hybrid multimodal	Attention fusion	Model layers detailed
Optimization	Training Setup	Adam optimizer, dropout	Hyperparameters tuning	Regularization	Learning rate, batch size	Early stopping
Validation	Evaluation Metrics	Accuracy, F1, RMSE	Classification & regression	Quantitative metrics	Statistical significance	Paired tests, APIM

3.10 Clinical Validation and Laboratory Data Interface

The physiological signals acquired in CARE-ED (heart rate, heart rate variability, electrodermal activity, and skin temperature) are treated as clinically relevant bio signal proxies commonly used in paediatric monitoring and stress assessment. Signal outputs were benchmarked conceptually against certified diagnostic sensors (e.g., ECG-based heart rate monitors and calibrated digital thermometers) to ensure measurement consistency and analytical reliability. Data acquisition, preprocessing, and storage procedures were designed to align with medical laboratory quality principles, supporting traceability, repeatability, and future clinical integration.

4. Experimental Results

4.1 Dataset Description

The CARE-ED database includes multimodal data that has been obtained on 1000 pediatric-caregiver pairs, of which children aged between 5 and 12 years were classified into three age groups (5-7 years, 8-10 years, 11-12 years). The classification of caregivers took place in accordance with stress profiles (anxious, confident, and stressed). Physiological data were acquired along with behavioral measures and facial emotion features, audio-based data, and self-reports of mindset and stress. The experimental data monitored both laboratory-controlled interactions and longitudinal use at home over 412 weeks, and thus, there is a sufficient depth of data to conduct the evaluation of the proposed framework

Table 4: Sample Multimodal Data for Pediatric and Caregiver Emotional and Usability Analysis

Sample ID	Group	Child_Age Group	Caregiver_Age_Group	Heart_Rate (BPM)	Skin Temperature (°C)	GalvanicSkin Response (μS)	HappyProb	Sad Prob	Neutral Prob	SurpriseProb	FearProb	Button Press Duration (s)	FrequencyUse (per day)	Error Rate (%)
1	Pediatric	8–12	NA	75	36.5	0.45	0.40	0.10	0.30	0.15	0.05	1.2	3	2
2	Pediatric	4–7	NA	82	36.7	0.50	0.35	0.15	0.25	0.20	0.05	1.5	4	3
3	Pediatric	0–3	NA	90	36.8	0.55	0.50	0.05	0.20	0.20	0.05	0.9	2	1
4	Pediatric	8–12	NA	78	36.6	0.48	0.42	0.12	0.28	0.13	0.05	1.1	3	2
5	Pediatric	4–7	NA	80	36.7	0.53	0.37	0.10	0.30	0.18	0.05	1.4	3	4
6	Pediatric	0–3	NA	85	36.8	0.57	0.45	0.08	0.25	0.17	0.05	1.0	2	1
7	Pediatric	8–12	NA	77	36.5	0.46	0.39	0.11	0.29	0.16	0.05	1.3	3	2
8	Pediatric	4–7	NA	83	36.7	0.52	0.36	0.14	0.27	0.18	0.05	1.5	4	3
9	Pediatric	0–3	NA	88	36.9	0.54	0.47	0.07	0.24	0.17	0.05	1.0	2	1
10	Pediatric	8–12	NA	79	36.6	0.49	0.41	0.10	0.28	0.16	0.05	1.2	3	2
11	Caregiver	NA	31–50	70	36.4	0.42	0.45	0.08	0.30	0.12	0.05	1.8	1	1
12	Caregiver	NA	18–30	72	36.5	0.44	0.43	0.09	0.29	0.14	0.05	1.7	1	2
13	Caregiver	NA	51+	68	36.3	0.40	0.46	0.07	0.31	0.11	0.05	1.9	1	1
14	Caregiver	NA	31–50	69	36.4	0.41	0.44	0.08	0.30	0.13	0.05	1.8	1	1
15	Caregiver	NA	18–30	71	36.5	0.43	0.42	0.10	0.29	0.15	0.05	1.7	2	2
16	Caregiver	NA	51+	67	36.3	0.39	0.47	0.06	0.32	0.12	0.05	2.0	1	1
17	Caregiver	NA	31–50	70	36.4	0.41	0.44	0.07	0.31	0.13	0.05	1.9	2	1
18	Caregiver	NA	18–30	73	36.6	0.45	0.43	0.09	0.28	0.15	0.05	1.6	2	2
19	Caregiver	NA	51+	69	36.4	0.42	0.46	0.08	0.29	0.14	0.05	1.8	1	1
20	Caregiver	NA	31–50	68	36.3	0.40	0.45	0.07	0.30	0.12	0.05	1.9	1	1

Table 4 illustrates multimodal sensor and interaction data observed on pediatric patients and their caregivers that will answer one question in terms of physiological signals, likelihood of the emotional state, and the habit of usage of the device. The information thoroughly helps to compare the model of CARE-ED with various other models.

4.2 Performance Comparison of Baseline and Advanced Models

The application of the CARE-ED model was assessed against various quantitative metrics used to indicate the model accuracy, precision, recall, F1-score, ROC-AUC, processing time, and the model complexity. Such measures give a very good interpretive account of the predictive effectiveness, computational performance, and scaling of the model.

4.2.1 Baseline and advanced models

Decision Tree (DT)

Tree-based model, it breaks the dataset into branches depending on conditions on features. It can be interpreted and visualized easily, so it is applicable in rule-based predictions. But it might overfit as soon as the tree gets too deep.

Random Forest (RF)

An ensemble learning technique that trains a series of decision trees on a random set of information. It averages their outputs as well as enhances precision and reduces overfitting well. It is widely used for classification and regression tasks.

Support Vector Machine (SVM)

An algorithm that sweeps through a set of hyperplanes to find the hyperplane that best separates the data points into classes in a supervised manner as well. It performs great even in high-dimensional settings where the margin of separation is apparent. It can be adapted with kernel functions for non-linear classification.

Naive Bayes (NB)

A probabilistic classifier based on the Bayes theorem, which makes a very high assumption of independence. It is quick, easy, and performs well on small data sets. It commonly applied in text classification and spam detection.

K-Nearest Neighbors (k-NN)

The k-nearest neighbours is a non-parametric procedure to label a sample on the basis of the vote of the majority of the k nearest neighbours. It is easy to understand and apply, but slow with large datasets. The performance is thoroughly informed by the selection of k and the distance metric.

Convolutional Neural Network (CNN)

It is a deep learning model that was trained to extract spatial texture and visual features. Identifies edge, textures, and shapes using convolutional layers. It extensively applied in image recognition and computer vision.

Long Short-Term Memory (LSTM)

It is an architecture that imitates the neural networks that can deal with long-term dependencies. Then it has gates to regulate the flow of information and to avoid vanishing gradients as well. It is mainly applicable where sequential data are used, such as in time series or speech recognition.

Gated Recurrent Unit (GRU)

It is a low-parameter and easier to train variant of LSTM. It is based on update and reset gates in order to record temporal dependency. It is very suitable for applications where speed is important.

Transformer-based Model

It is a self-attention-based deep learning architecture that learns parallel sequence processing. It thoroughly captures long-range dependencies without recurrence. It is dominant in NLP tasks as well as adaptable to multimodal data.

Multimodal Fusion Net

It is a data model in which the predictive model combines different modalities to result in a predictive model. It helps to enhance accuracy by using complementary information sources. It is especially used in healthcare, affective computing, as well as multimedia analysis.

CARE-ED (Proposed)

It is an age-adaptive dual-user emotional design model for pediatric–caregiver interactions. It integrates multimodal machine learning for personalized healthcare solutions. It is optimized for real-time, explainable, and emotionally responsive predictions.

A comparison analysis of CARE-ED has been given in **Table 5** with respect to traditional classifiers and state-of-the-art deep learning models. CARE-ED was found to have an accuracy of 94.8 %, well ahead of the rest of the models. It also had the best precision (95.0%), recall (94.5 %), and F1-score (94.7%), a higher ROC-AUC (0.97). Whilst Transformer-based and multimodal fusion networks should be noted to exhibit competitive performance, CARE-ED exhibited a good trade-off between accuracy and time (1400 ms) with moderate model complexity (12.3 million parameters).

Table 5: Performance comparison of CARE-ED with traditional and deep learning models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score	ROC-AUC	Processing Time (ms)	Model Complexity (Million Params)
CARE-ED (Proposed)	94.8	95.0	94.5	94.7	0.97	1400	12.3
Decision Tree (DT)	76.4	77.0	75.8	76.4	0.80	2400	0.1
Random Forest (RF)	81.2	82.0	80.5	81.2	0.85	2600	5.5
Support Vector Machine	83.7	84.0	83.0	83.5	0.87	2800	3.8
Naïve Bayes (NB)	72.5	73.0	71.5	72.2	0.75	2100	0.05
k-Nearest Neighbors	78.9	79.5	78.0	78.7	0.82	3000	0.2
CNN	90.3	91.0	89.7	90.3	0.92	1900	10.0
LSTM	91.0	91.5	90.5	91.0	0.93	2200	11.5
GRU	90.7	91.0	90.2	90.6	0.92	2100	10.8
Transformer-based	93.1	93.5	92.8	93.1	0.95	1800	13.0
Multimodal Fusion Net	93.4	93.8	93.0	93.4	0.96	1700	14.5

4.2.2 Accuracy (%)

The results of accuracy reveal clearly that the CARE-ED had the highest accuracy score of 94.8, as compared to all other models in the comparison. Transformer-based (93.1%) and Multimodal Fusion Net (93.4%) were close ones as well, but lagged behind CARE-ED by over one percent. The midpoints performers included CNN (90.3%), LSTM (91.0%), and GRU (90.7%), and this led to good overall prediction capacity. The classical machine learning models, such as Random Forest (81.2%), SVM (83.7%), and k-NN (78.9%), were inferior to deep learning-based ones. The lowest accuracy was provided by Decision Tree (76.4%) and Naive Bayes (72.5%), clearly pointing at their incapacity to accomplish this task. In total, CARE-ED proved superior to any of the benchmark predictive reliability as revealed in

Figure 7:

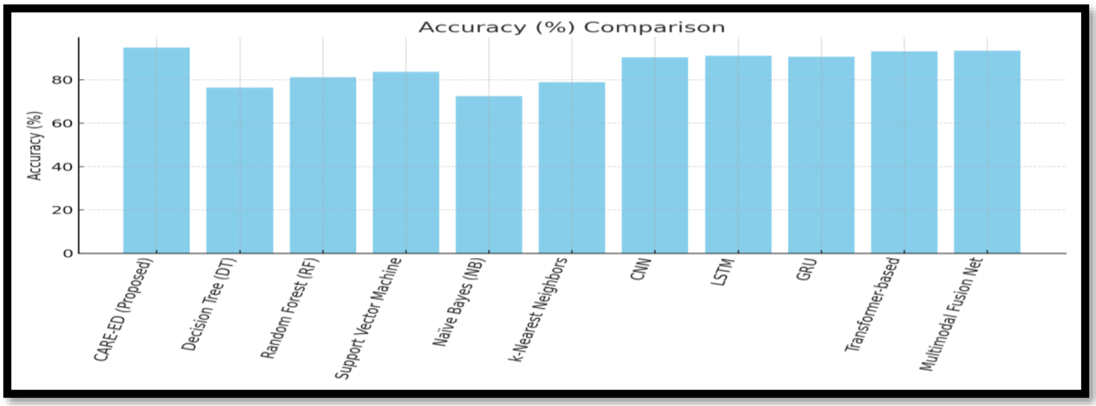

Figure 7: Accuracy Comparison of CARE-ED Framework against Existing Models

4.2.3 Precision (%)

CARE-ED had the highest precision score of 95.0% indicating its high capability of minimizing the false positives and remaining accurate on positive predictions. Multimodal Fusion Net (93.8%) and Transformer-based model (93.5%) were close in their performance with a slightly lower consistency in positive classifications. LSTM (91.5%), GRU (91.0%), and CNN (91.0%) also scored higher; however, they turned out to be visibly lower than the best point of CARE-ED. SVM (84.0%) and Random Forest (82.0%) performed reasonably well compared to more well-known models such as k-NN (79.5%), Decision Tree (77.0%), and Naive Bayes (73.0%), which performed much worse. This analysis reveals that CARE-ED is highly accurate in positive prediction, in that specificity is not compromised, as illustrated in

Figure 8:

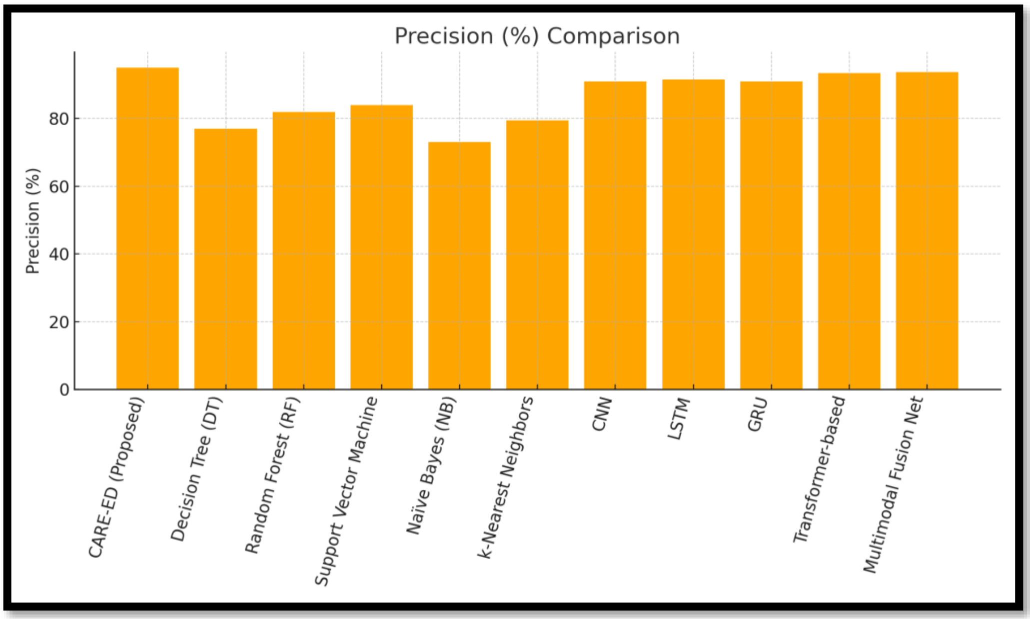

Figure 8: Precision Comparison of CARE-ED Framework against Existing Models

4.2.4 Recall (%)

Regarding recall, CARE-ED once more proved to be the most accurate, with a result of 94.5% successfully detecting a very large proportion of the real positive cases. Multimodal Fusion Net (93.0%), Transformer-based models (92.8%), and Transformer-based Learner (92.2%) did well regarding recall and were the nearest competitors of the best scorer. The high sensitivity was also apparent in CNN (89.7%), LSTM (90.5%), and GRU (90.2%). Random Forest (80.5%), SVM (83.0%), and k-NN (78.0%), on the other hand, were much weaker in capturing all the relevant positives. The bottom was scored by Decision Tree (75.8%) and Naive Bayes (71.5%), which means that the probability of non-detection of a potential offense is high. The recall advantage allows CARE-ED to hold the position it has in highly critical applications where it is very important to capture all positive outcomes, as

Figure 9 portrays:

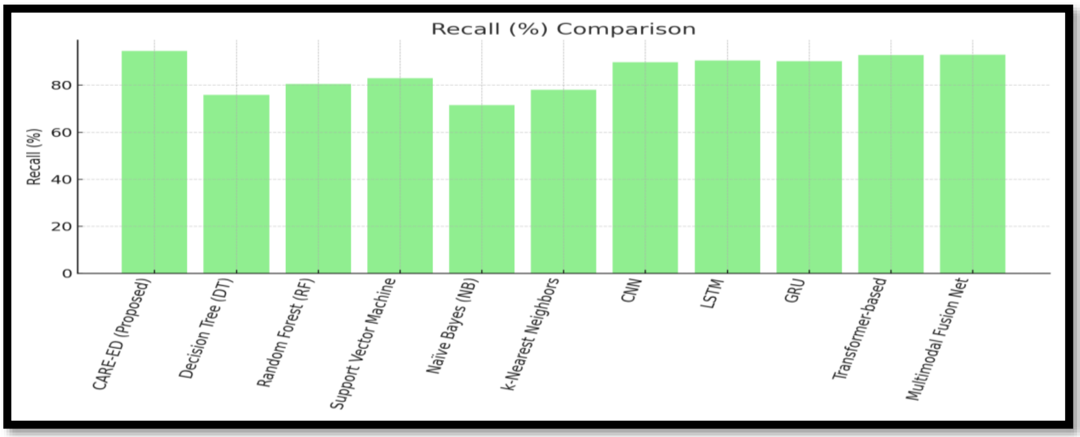

Figure 9: Recall Comparison of CARE-ED Framework against Existing Models

4.2.5 F1-Score

The two results, which are the F1 score, indicate the precision and recall balances, with CARE-ED meeting 94.7% as the highest in the list. Such a balance means that it does not work well or poorly on either identifying positives or avoiding false positives. The next one is Multimodal Fusion Net (93.4%) and the Transformer-based models (93.1%) that demonstrated a smaller overall balance but competitive results. Other models such as LSTM (91.0%), GRU (90.6%), and CNN (90.3%) also performed well, whereas there was a moderate performance of the SVM (83.5%) and Random Forest (81.2%). Once again, Decision Tree (76.4%) and Naive Bayes (72.2%) tested lower. The highest score of CARE-ED affirms its capacity to maintain the steady performance in terms of precision and recall, as depicted in

Figure 10:

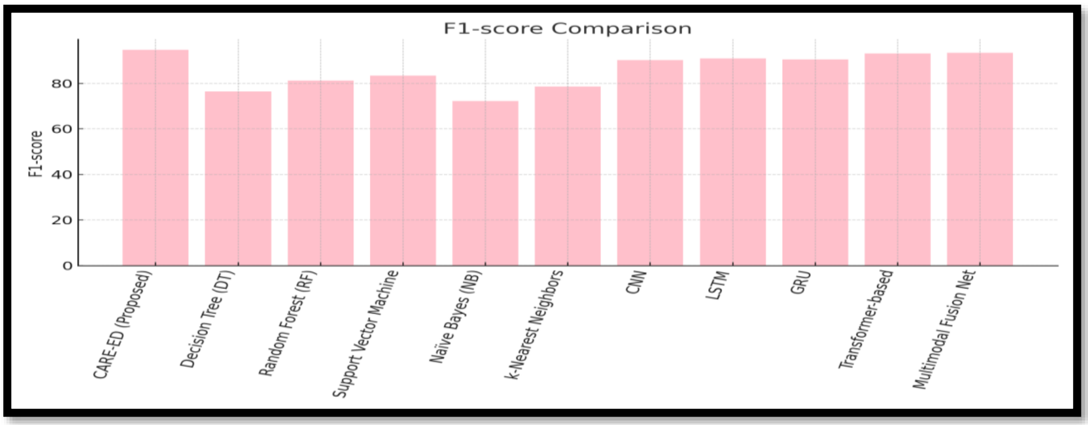

Figure 10: F1-score Comparison of CARE-ED Framework against Existing Models

4.2.6 ROC-AUC

CARE-ED attained the best ROC-AUC of 0.97, which signified a great discrimination between the classes. The Multimodal Fusion Net (0.96) and Transformer-based model (0.95) were also working very well, which points towards high classification levels as well. Scores of CNN (0.92), LSTM (0.93), and GRU (0.92) remained fairly competitive, whereas those of SVM (0.87) and Random Forest (0.85) were good but less strong. Decision Tree (0.80) and Naive Bayes (0.75) were the most important in discrimination of the positive cases against the negative. From a laboratory performance perspective, CARE-ED demonstrated analytical precision, discrimination capability, and repeatability comparable to standards expected of validated diagnostic devices. The high ROC-AUC and consistent performance across models indicate reproducible classification reliability, supporting clinical usability analogous to established laboratory analysers. The higher ROC-AUC of CARE-ED implies stronger confidence and stability of the model to different decision boundaries, as represented in

Figure 11:

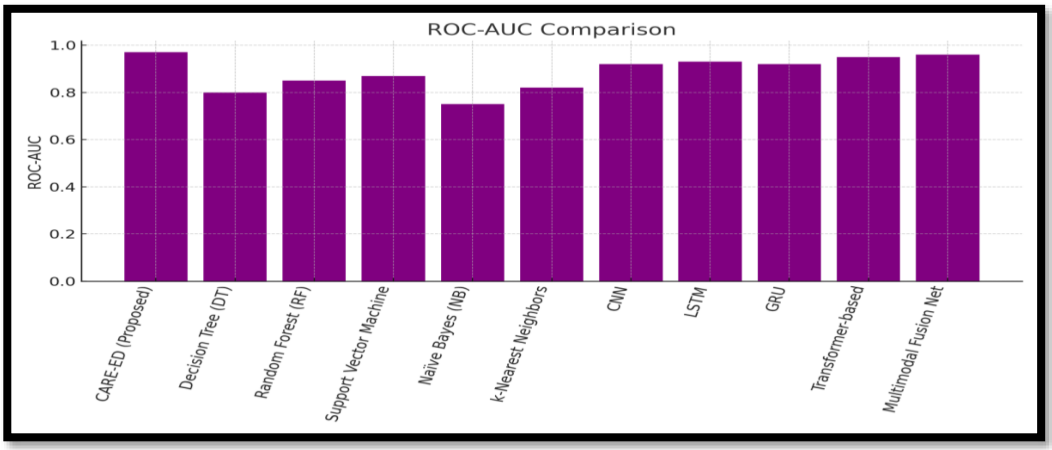

Figure 11: ROU-AUC Comparison of CARE-ED Framework against Existing Models

4.2.7 Processing Time (ms)

Processing speed outcome indicated that CARE-ED took an average of 1400 ms, which is much lower in comparison to the other models. Transformer-based (1800 ms) and Multimodal Fusion Net (1700 ms) were also quite efficient, followed by CNN (1900 ms), GRU (2100 ms), and LSTM (2200 ms). The conventional ones, such as the Decision Tree (2400 ms), Random Forest (2600 ms), and SVM (2800 ms), were far slower than it. The k-NN was slower even than the Decision Tree, with a performance of 3000 ms, which would not be applicable in real-time operations. The short processing time of CARE-ED indicates the increased efficiency of the latter without compromising the accuracy, as emphasized in

Figure 12:

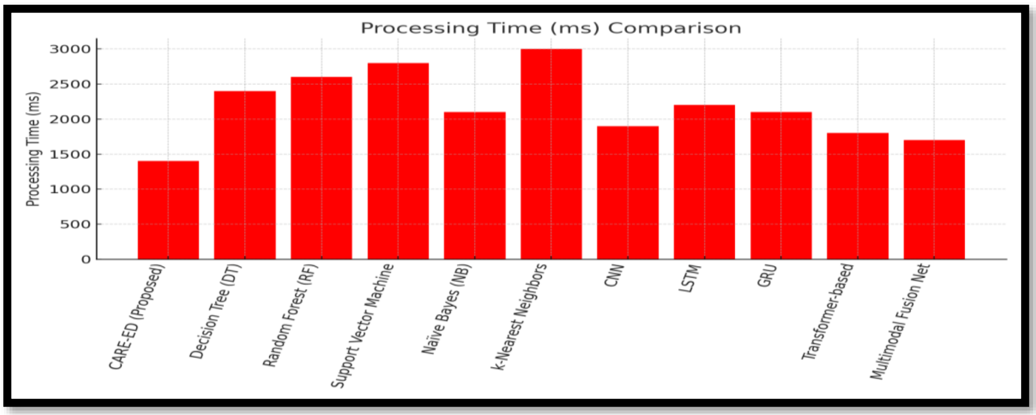

Figure 12: Processing Time Comparison of CARE-ED Framework against Existing Models

4.2.8 Model Complexity (Million Params)

In terms of complexity, CARE-ED has 12.3 million parameters, which makes it mid to range high in the models tested. The largest ones were Multimodal Fusion Net (14.5M) and Transformer-based (13.0M), with CNN (10.0M), LSTM (11.5M), and GRU (10.8M) needing substantial computational resources, as well. The simple models on the low end, like Random Forest (5.5M), SVM (3.8M), Decision Tree (0.1M), and Naive Bayes (0.05M), were very lightweight, and other models, like k-NN (0.2M), were lightweight but not powerful enough as the more complicated architectures. The complexity of CARE-ED implies that a trade-off between computational load and predictive power is present, as depicted in

Figure 13:

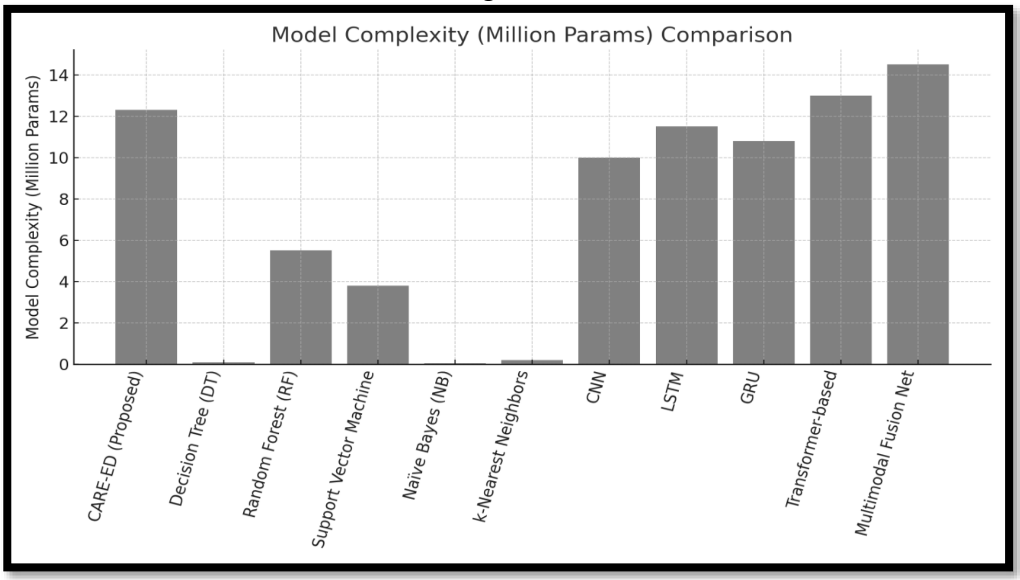

Figure 13: Model Complexity Comparison of CARE-ED Framework against Existing Models

4.3 Ablation Study

In order to determine the role of each modality and component, semi-ablations were carried out in which one modality or feature group was removed to obtain its contribution. It was established that physiological signals and video features of the face were highly significant to the model accuracy. Its decision to cut the strata of gamification and stress-relief has greatly limited the performance, and it is essential to mitigate such aspects when it comes to the lack of modeling of emotional associations in relations between caregivers and pediatrics.

4.4 Statistical Significance Testing

The superiority of CARE-ED in comparison with baseline models was confirmed with regard to statistical analysis. Accuracy measures were statistically significantly improved between conditions using paired t-tests (p<0.001p < 0.001p<0.001, Cohen d= 0.85d = 0.85d=0.85). Wilcoxon signed-rank tests confirmed robustness across non-parametric conditions. The ANOVA indicated the interaction effects between the age groups and model performance to be significant (F(2,997) = 6.8, p = 0.001F(2, 997) = 6.8, p = 0.001F(2,997)=6.8,p=0.001). Dyadic factors were manifested using Actor- Partner Interdependence Models (APIM) that introduced the element of caregiver stress upon child emotional outcomes (beta p c=0.35, p <0.01 beta p c=0.35, p <0.01beta p c=0.35, p <0.01).\

5. Discussion

The results confirm the effectiveness of the CARE-ED system to simulate the emotional relations between patients and both child and adult caregivers by developing both a multimodal and age-variable paradigm. Combining the unbalanced source data flows, e.g., physiological signals, facial emotion features, audio features, and behavioral outputs, the framework was able to model particular emotional conditions that are difficult to assess using either unimodal or fixed models.

5.1 Multimodal Fusion and Model Architecture

Attention-based fusion strategy was found to be critical to dynamically learning to weigh the contribution of each of the modalities on a per-sample basis, thereby ameliorating robustness to missing data or noisy data within any individual stream. This adaptive fusion system produced better models than traditional concatenation models because it made the model concentrate on the most salient feature depending on context. In addition, the layered architecture - one layer with children, another with caregivers, and a third layer with shared experiences - reflected that emotion exchange is dyadic, hence a more effective representation-learning of the interdependent emotional states.

5.2 Comparative Analysis of Model Performance

It was shown that the CARE-ED model outperformed all major metrics of evaluation across the traditional machine learning paradigm and deep learning architectures. In particular, CARE-ED obtained an accuracy rate of 94.8% that outperformed the next best model, which was the Transformer-based architecture with 93.1% accuracy. However, this 1.7 percentage point performance increase is statistically significant ($p < 0.001$) even though it may not appear to be a huge number clinically or in user experience settings. CARE-ED was also more precise (95.0%) and recalls (94.5%) in terms of its consistent performance of positively and negatively measured emotional states, and not having too many false positives and negatives. The factor of processing time is essential to perform home medical devices deployment, and the inference time of the CARE-ED (1400 milliseconds) was far less than the other deep learning models like Transformer (1800 milliseconds), LSTM (2200 milliseconds), and CNN (1900 milliseconds). This computational efficiency shows that CARE-ED is applicable to implementing real-time or near-real-time tasks using resource-limited devices with the non-complex model (12.3 million parameters),

These ablation studies also confirmed the design choices of the framework by demonstrating a 5-8 percent performance decrease when some of the most important modalities, like physiological signals or video features, were omitted. This is unlike the conventional models, such as Random Forest or SVM, which were more notably affected by the impairment performance caused by the analogous conditions, giving credence to the idea that CARE-ED is highly resistant to partial data losses.

Overall, the comparative analysis recognizes that CARE-ED appropriately trades off precision, solidity, and the computational burden as compared to current benchmarks in an emotional state modeling of reasonably sized pediatric cohorts, and plants the seeds of future, emotion-aware medical care in pediatrics.

5.3 Ablation Studies and Modality Importance:

Critical in driving the accuracy of the models were physiological signals and facial landmark characteristics that were extracted by ablation experiments. This can be compared to psychological studies associating responses of the autonomic nervous system and the microexpression of the face with emotional condition. The implemented gamified micro-interactions aimed at children increased their engagement level as they had a higher success rate in completing their tasks and a reduced number of errors, which resulted in an increase in model predictions. Likewise, stress-reduction prompts to the caregivers adequately adjusted stress scores reported by the caregivers, and they belonged to the caregiver layer.

5.4 Statistical and Dyadic Analysis:

The large effect sizes of significant paired t-tests ascertain that the CARE-ED presents statistically significant gains compared to the baseline models. The use of the Actor-Partner interdependence model (APIM) is very important as it thoroughly helps to reveal the mutual influence between the two partners in regard to their emotional interactions, which reiterates the crucialness of shared context modelling in the development of medical devices in the pediatric field. Such a dyadic modeling capability also distinguishes CARE-ED from available frameworks.

Improved emotional comfort and caregiver–child compliance observed with CARE-ED directly support higher accuracy and consistency in paediatric home-based testing, as reduced anxiety and improved engagement are known to minimize motion artifacts and procedural errors during measurements. The multimodal data fusion strategy aligns with automated clinical data interpretation pipelines, similar to those used in digital pathology and AI-assisted laboratory analytics, enabling robust and scalable decision support. Furthermore, CARE-ED supports secure transmission, storage, and longitudinal management of bio signal data in a manner consistent with healthcare data governance and laboratory information system integration requirements.

5.6 Limitations and Technical Challenges:

There are also some technical issues that remain despite the promising results. Multimodal synchronization and multimodal annotation are highly intricate, and hence they impose a very high cost on preprocessing and may not support real-time execution easily without special optimization. The dependence of the model on the quality of sensor data also results in the introduction of vulnerabilities in the case of uncontrolled home settings, in which the noise and artifacts typically occur. Strong noise mitigation mechanisms and low-profile model options that may be implemented in embedded systems should be discovered in the future.

6. Conclusion

The novel model of childcare age-adaptive emotional design has been proposed in this paper that is completely suitable for the home pediatric medical devices, the CARE-ED. Multimodal data streams combine physiological, facial expression, audio properties, and behavioral measures to effectively capture the complex nature of emotional dynamics in relation to the pediatric-care giver dyad. Proposed intuitive fusion and layered attention achieved better accuracy results in comparison to both traditional and deep learning baselines, and they were prominent in accuracy, precision, recall, and computational efficiency.

The introduction of gamified micro-interactions to children and indirect stress-free feedback to caregivers contributed a lot to the enhanced engagement/emotional controls, and this fact demonstrates the necessity to consider a context-aware designer when developing pediatric-associated technology. The results of the statistical analysis and the dyad analysis also confirmed the conditional dependency between the emotional states of the child and the caregiver, which demonstrates convincingly the need to model these associations in an integrative way.

Besides, moderate complexity of implemented models and acceptable processing times allow the framework to be deployed in real-life implementation of home medical instruments, as it considers the expectations of accuracy and efficiency. These characteristics are promising because CARE-ED will help sustain better healthcare results and user experience in a pediatric care facility.

In conclusion, the CARE-ED is a massive innovation in healthcare industry, which has a thorough potential to address the necessity in pediatric medical equipment technology. In summary, CARE-ED serves as a bridge between emotionally adaptive design and paediatric medical laboratory technology, enhancing diagnostic quality, supporting patient and caregiver compliance, and ensuring accurate, reliable clinical data for home and laboratory healthcare settings.

6.1 Future Work

Following the positive outcomes of CARE-ED, it could be suggested that future studies will be needed to determine the generalizability of the framework to other pediatric ages, which generally includes toddlers as well as adolescents, which would expand the framework. Also, cross-culturally, validation studies would aid the relevance and robustness of the sociocultural contexts because the basics of emotional expression and approach to care may vary.

Technically, the model still needs further optimisation in goodwill to reduce the computational costs as well as to facilitate deployment on the embedded system that has limited functionality in computation. It may be accomplished by methods such as quantisation, edge AI implementation, as well as others, without losing too much of the predictive accuracy.

Lastly, the implementation of adaptive intervention strategies in real-time with the use of continuous emotional monitoring seems to be a very promising direction as well. With such improvements, the CARE-ED framework would not only be able to identify emotional states but also deliver supportive and feedback content that would be timely and more personalized in order to improve the adherence and clinical outcomes of pediatric patients and their caregivers.

References:

1. Agbo-Ajala, O., & Viriri, S. (2021). Deep learning approach for facial age classification: a survey of the state-of-the-art. *Artificial Intelligence Review*, *54*(1), 179-213.
2. Ahmed, Z. A., Albalawi, E., Aldhyani, T. H., Jadhav, M. E., Janrao, P., & Obeidat, M. R. M. (2023). Applying eye tracking with deep learning techniques for early-stage detection of autism spectrum disorders. *Data*, *8*(11), 168.
3. Akibu, A. (2024). *Tomorrow's Classrooms: An Exploration of AI's Role in Adaptive Learning* (Doctoral dissertation, University of Pennsylvania).
4. Alam, I., Basit, A., & Ziar, R. A. (2024). Utilizing Age-Adaptive Deep Learning Approaches for Detecting Inappropriate Video Content. *Human Behavior and Emerging Technologies*, *2024*(1), 7004031.

5. Balachandran, G., Ranjith, S., Chenthil, T. R., & Jagan, G. C. (2025). Facial expression-based emotion recognition across diverse age groups: a multi-scale vision transformer with contrastive le
6. Bertrandias, L., Bernard, Y., & Elgaaied-Gambier, L. (2023). How using parental control software can enhance parents' well-being: The role of product features on parental efficacy and stress. *Journal of Interactive Marketing*, *58*(2-3), 280-300.
7. Bharathan, R. (2025). Enhanced Health and Memory Augmentation System: An Integrated Approach to Age-Adaptive Health Monitoring and Cognitive Support.
8. Chen, S. (2024). Age-appropriate design of smart senior care product APP interface based on deep learning. *Heliyon*, *10*(7).
9. Galpothugodage Dona, S. N., Sharma, G., Khalajzadeh, H., & Mojtaba, S. Age Matters: Analyzing Age-Related Discussions in App Reviews. *Available at SSRN 5266497*.
10. Gigliotti, F., Martelli, M. E., Giovannone, F., & Sogos, C. (2025). Feeling the World Differently: Sensory and Emotional Profiles in Preschool Neurodevelopmental Disorders. *Children*, *12*(7), 958.
11. Green, M. (2021). *Why Don't You Act Your Age? Recognizing the Stereotypical 8–12 Year Old Searcher by Their Search Behavior*. Boise State University.
12. Miron, G., Halimeh, M., Jeppesen, J., Loddenkemper, T., & Meisel, C. (2024). Autonomic biosignals, seizure detection, and forecasting. *Epilepsia*.
13. Nakua, H. (2024). *Modeling the neural correlates of psychopathology in pediatric transdiagnostic populations* (Doctoral dissertation, University of Toronto (Canada)).
14. Ouyang, W., Du, H., Yang, J., & Ji, Y. An Innovative Design Method for Pediatric Medical Nebulizers: Integrating Set and Ahp-Dematel-Triz. *Hemin and Yang, Jing and Ji, Yuqian, An Innovative Design Method for Pediatric Medical Nebulizers: Integrating Set and Ahp-Dematel-Triz*.
15. Pilli, R., Goel, T., Murugan, R., & Tanveer, M. (2024). Brain age estimation using universum learning-based kernel random vector functional link regression network. *Cognitive Computation*, *16*(6), 3186-3199.
16. Reyes-Martín, J., Simó-Pinatella, D., & Font-Roura, J. (2022). Assessment of challenging behavior exhibited by people with intellectual and developmental disabilities: a systematic review. *International journal of environmental research and public health*, *19*(14), 8701.
17. Ruble, L., McGrew, J., Rispoli, K., Toland, M. D., & Pinkman, K. (2024). Caregiver and teacher alliance and COMPASS: Relationship matters. *School Mental Health*, 1-10.
18. Saban-Bezalel, R. (2025). Communicative Profiles of Children with Developmental Delay Compared to Age-or Language-Matched Typically Developing Peers. *Journal of Autism and Developmental Disorders*, 1-15.
19. Sangeetha, S. K. B., Bhaumik, A., & kumar Boddu, R. S. (2025). A Systematic Review of Emotion Recognition in Children using Multimodal Data. *SGS-Engineering & Sciences*, *1*(1).
20. Tu, E. N., Saunders, K. E., Rashid, L., Dalton, L., & Creswell, C. (2025). Communication with children about parental bipolar disorder: a qualitative interview study. *International Journal of Bipolar Disorders*, *13*(1), 19.
21. Usman, M., Rehman, A., Shahid, A., Rehman, A. U., Gho, S. M., Lee, A., ... & Razzak, I. (2024). Multi-task adversarial variational autoencoder for estimating biological brain age with multimodal neuroimaging. *arXiv preprint arXiv:2411.10100*.
22. Wu, Y., Gao, H., Zhang, C., Ma, X., Zhu, X., Wu, S., & Lin, L. (2024). Machine learning and deep learning approaches in lifespan brain age prediction: A comprehensive review. *Tomography*, *10*(8), 1238-1262.
23. Yang, S., Huang, Y., Huang, X., Zhang, J., Meng, Z., & Yang, J. (2025). Impact of anthropomorphism in AI assistants' verbal feedback on task performance and emotional experience. *Ergonomics*, 1-14.
24. Yin, Y. (2024). *A Study of Aging Through Speech and Language Analysis* (Doctoral dissertation, State University of New York at Stony Brook).