AI-Based Screen Addiction Risk Prediction in Toddlers Using Behavioral Data

Kalyani Lasankar
Shri Shivaji College of Engineering and Technology Akola

Abstract—**Excessive screen exposure in early childhood has become a growing pediatric and public-health concern, especially because toddlers are in a rapid period of language, socioemotional, and executive-function development. The WHO recommends no screen time for children younger than 1 year and no more than 1 hour per day for children aged 1 to 4 years, within broader 24-hour movement guidance for children under 5. Recent studies report that screen exposure above this threshold is associated with poorer language development, behavioral difficulties, and weaker caregiver-child interaction quality in toddlers and preschoolers. This paper proposes a master's-level research framework for predicting screen addiction risk in toddlers using behavioral data collected from parents, caregivers, and digital-use logs. The framework combines structured behavioral features, duration-pattern features, and context-aware indicators such as co-viewing, device type, bedtime use, and interruption resistance, then applies supervised machine learning for early risk stratification. A feature pipeline and evaluation design are presented for binary and ordinal risk classification using logistic regression, random forest, gradient boosting, and calibrated neural models, with AUROC, F1-score, and confusion matrices used for assessment. The paper further discusses ethical safeguards, clinical limitations, and translational implications for pediatric screening tools, early-intervention apps, and parental monitoring systems.pmc.ncbi.nlm.nih+8**

***Index Terms*—Toddler screen use; screen addiction risk; behavioral data; pediatric digital health; machine learning; early intervention; risk prediction; parental monitoring.**

I. INTRODUCTION

Background

Toddlers are increasingly exposed to smartphones, tablets, and on-demand video platforms during a developmental period when sleep, language input, and caregiver interaction are especially important. WHO guidance for children under 5 sets strict limits on sedentary screen exposure and emphasizes physical activity and sleep as part of healthy development. Population-level evidence shows that many young children exceed these limits, with only about one-third of children aged 2 to 5 years meeting screen-time recommendations in global studies. This creates a strong rationale for computational screening methods that can identify children at elevated risk before harmful patterns become entrenched.journals.plos+3

Motivation

Traditional pediatric screening often relies on retrospective parental report and clinician observation, which can miss early warning signals such as escalating usage frequency, bedtime screen use, and difficulty disengaging from devices. Recent studies indicate that higher screen exposure in toddlers is associated with poorer language outcomes and reduced conversational engagement. Machine learning is increasingly used in pediatric risk prediction because it can integrate heterogeneous behavioral and contextual variables and produce calibrated risk estimates for early intervention. A toddler-focused screen-addiction risk model could support family counseling, digital-wellbeing apps, and primary-care triage.pubmed.ncbi.nlm.nih+3

Problem Statement

The problem is to predict whether a toddler is at high risk of problematic or addiction-like **s**creen use using only behavioral data available in home and childcare settings. The challenge is that toddler screen addiction is not a formal diagnostic entity in major classification systems, so the target must be operationalized through measurable behavioral markers such as excessive duration, frequent escalation, distress on removal, bedtime use, and reduced engagement in non-screen activities. The research objective is to build a validated AI model that can distinguish low-, moderate-, and high-risk screen-use profiles with clinically meaningful performance.who+2

II. LITERATURE REVIEW

WHO guidance remains the primary international reference for screen exposure in children under 5, recommending no screen time for infants and a maximum of 1 hour per day for older toddlers and preschoolers. Singapore's pediatric guidance and related local summaries align closely with these limits, reinforcing their relevance in applied child-health settings. Recent evidence shows that toddlers with higher mobile-device screen time are more likely to have poorer language development, and caregiver reading can partly buffer some negative effects. Broader reviews from PubMed-indexed literature report associations between high screen use and reduced physical activity, poorer sleep, attention difficulties, and emotional or social challenges.kkh+5

A growing literature links screen exposure to developmental and behavioral outcomes in young children. Studies of preschoolers report that screen duration and intensity are associated with delayed milestones, poorer vocabulary, and behavioral problems. More recent work also suggests that screen use can reduce the amount of adult words and back-and-forth conversation toddlers hear at home, affecting the language environment directly. These findings support the use of behavioral markers, not only exposure duration, as predictive features.scribd+2

In machine learning for child risk prediction, recent pediatric studies demonstrate that multilevel behavioral and contextual inputs can identify at-risk groups with strong accuracy. For example, supervised learning on child, parent, and contextual variables has achieved high classification performance for conduct-problem risk, showing the value of feature selection and ensemble models. Another child mental-health study used supervised and ensemble methods with cross-validation and calibration assessment to predict high-risk groups more effectively than traditional logistic regression. For screen-addiction prediction specifically, recent ACM-indexed and other digital-age behavior studies support the use of questionnaire-based and usage-pattern features for risk modeling. Taken together, the literature supports a toddler-specific, behavior-driven AI framework.acm+2

III. PROPOSED METHODOLOGY

The proposed study uses a supervised learning pipeline to predict toddler screen-addiction risk from behavioral data collected from parents, caregivers, and optional device logs. The design is cross-sectional for initial model development, with the possibility of longitudinal extension for temporal risk forecasting. The target label can be defined using a clinically informed composite score derived from duration, dysregulation, and impairment indicators, then binarized into low versus high risk or modeled as a three-class outcome.

The pipeline has five stages: data collection, preprocessing, feature extraction, model training, and validation. Candidate models include logistic regression as a baseline, random forest, XGBoost or LightGBM for nonlinear interactions, and a calibrated multilayer perceptron if the sample size is sufficient. This structure is consistent with recent pediatric prediction studies that show ensemble methods can outperform simpler parametric models when risk signals are nonlinear and heterogeneous.pmc.ncbi.nlm.nih+1

IV. DATA COLLECTION

Behavioral data should be collected from caregivers using structured questionnaires and optional passive logs. Core variables include daily screen duration, number of sessions, device type, time of first screen exposure, bedtime screen use, co-viewing frequency, content category, tantrums when screen use ends, and proportion of time spent in non-screen play. Parent-report tools should be supplemented by a brief developmental and household context survey, because family routines and supervision patterns are linked to screen outcomes.thekids+1

For an academic thesis, the dataset may be constructed from a prospective observational study in pediatric clinics or preschools. Ethical approval and informed consent are mandatory because the participants are minors. To improve robustness, the study should oversample diverse socioeconomic groups, since recent cohort evidence suggests screen exposure differs by family context and social resources. If device logs are available, they should be anonymized and aggregated into daily summaries.ucl.ac

Feature Extraction

Feature engineering should transform raw behavioral observations into predictive indicators. Useful feature groups include exposure intensity, temporal regularity, context of use, and dysregulation signals. Examples include mean daily minutes, standard deviation of daily exposure, late-evening exposure ratio, number of switch attempts, frequency of device requests, and proportion of unsupervised use.

A practical feature set is shown below.

Feature group	Examples	Rationale
Exposure intensity	Mean daily minutes, max daily minutes, days above threshold	Captures dose-response risk pubmed.ncbi.nlm.nih+1
Temporal pattern	Bedtime use, morning-first-use time, session fragmentation	Screens near sleep are developmentally concerning who
Behavioral dysregulation	Tantrums on removal, distress latency, refusal of alternative play	Reflects addiction-like dependence
Interaction context	Co-viewing, caregiver mediation, content type	Moderates developmental effect pubmed.ncbi.nlm.nih+1
Developmental environment	Reading frequency, outdoor play, sleep regularity	Related to buffering and confounding pubmed.ncbi.nlm.nih+1

These features are aligned with the literature indicating that context matters as much as duration, especially for language and socioemotional outcomes.pubmed.ncbi.nlm.nih+1

Machine Learning Model

The main model should be a supervised classifier trained on structured behavioral features. Logistic regression is suitable as a transparent baseline; random forest and gradient boosting are recommended for the main comparison because they handle nonlinear interactions and mixed feature types well. If the dataset is large enough, the final system can use calibrated probabilities rather than only hard labels so that pediatricians and parents receive interpretable risk scores.journals.plos+1

A recommended modeling sequence is:

1. Baseline logistic regression.
2. Random forest classifier.
3. Gradient-boosting classifier.
4. Calibrated ensemble with probability correction.

Feature importance should be reported using permutation importance or SHAP values. This is important because clinical adoption depends on interpretability and because parents must understand why a model flags risk.

Training & Validation

The dataset should be split into training, validation, and holdout test sets, for example 70/15/15, with stratification by risk class. Five-fold or ten-fold cross-validation should be used within the training set to reduce variance and support hyperparameter tuning. Nested cross-validation is preferable if the sample size permits, since it limits optimistic bias.

Evaluation metrics should include accuracy, precision, recall, F1-score, ROC-AUC, and calibration metrics such as Brier score or calibration slope. Because the positive class may be imbalanced, precision-recall AUC should also be reported. This approach follows modern predictive modeling practice in child health, where discrimination and calibration are both important.pmc.ncbi.nlm.nih+1

Dataset Description

A thesis-level paper should clearly document the dataset source, whether original, public, or synthetic. Because no single authoritative public dataset specifically targets toddler screen addiction risk, the recommended approach is to build a new dataset from caregiver surveys and behavioral logs, then anchor the protocol to published guidance and evidence. The table below describes a reproducible dataset specification that can be implemented in an ethics-approved study.jamanetwork+1

Dataset element	Description	Source URL / reference
Inclusion criteria	Children aged 12 to 48 months	WHO under-5 guidance who
Exposure variables	Daily screen minutes, session counts, bedtime use	Behavioral survey and logs
Context variables	Co-viewing, caregiver mediation, reading frequency	Toddler language study evidence pubmed.ncbi.nlm.nih+1
Outcome label	Low / moderate / high screen-addiction risk	Operational definition based on behavior
Reference threshold	No screen time under 1 year; max 1 hour/day for toddlers and preschoolers	WHO guidance who+1

External benchmark	Screen use associated with language and behavioral outcomes	PubMed studies pubmed.ncbi.nlm.nih+1

Source URLs that should be cited in the thesis include the WHO under-5 guideline page, the JAMA Pediatrics meta-analysis on guideline adherence, the PubMed study on toddler mobile-device screen time and language development, and the PubMed review on screen time and child development.pubmed.ncbi.nlm.nih+3

Experimental Setup

The experimental environment can be implemented in Python using scikit-learn, pandas, NumPy, and XGBoost or LightGBM. Data preprocessing should include missing-value imputation, one-hot encoding for categorical fields, and scaling where needed. The software stack should also support SHAP for interpretability and Matplotlib or Plotly for visualization.

Evaluation should be conducted on a held-out test set and reported with confidence intervals if bootstrap resampling is feasible. Suggested metrics include:

- Accuracy.
- ROC-AUC.
- Precision, recall, and F1-score.
- Confusion matrix.
- Calibration curve.
- Sensitivity at clinically relevant thresholds.

This evaluation style is consistent with recent machine learning work in pediatric risk prediction, where both ranking performance and calibration are central.journals.plos+1

V. RESULTS & DISCUSSION

Because the task here is a thesis-style research paper rather than an executed empirical study, the results below should be presented as a reporting template for the final dissertation once the dataset is collected. In an actual implementation, gradient boosting is expected to outperform logistic regression when risk depends on nonlinear interactions among screen duration, bedtime use, and caregiver mediation. A plausible result pattern would show moderate-to-high ROC-AUC, with confusion-matrix errors concentrated between adjacent risk classes rather than between extreme classes.pmc.ncbi.nlm.nih+1

A results table may be structured as follows in the final thesis:

Model	Accuracy	ROC-AUC	F1-score	Notes
Logistic regression	Reported after training	Reported after training	Reported after training	Transparent baseline

Random forest	Reported after training	Reported after training	Reported after training	Handles nonlinearities
Gradient boosting	Reported after training	Reported after training	Reported after training	Likely strongest performer
Calibrated ensemble	Reported after training	Reported after training	Reported after training	Best for deployment

The discussion should link stronger predicted risk to the literature showing that excessive or unsupervised screen use is associated with poorer language development, reduced conversation, and more behavioral difficulties. It should also note that interactive, co-viewed, and age-appropriate use may be less harmful than passive use, which reinforces the need for context-aware prediction rather than a simple minutes-per-day rule.pmc.ncbi.nlm.nih+3

Ethical Considerations & Limitations

Research on toddlers requires strict privacy protection, parent consent, and minimal data collection. Behavioral logs should be anonymized, encrypted, and retained only for the approved study period. Because risk prediction may influence parenting decisions, the model must avoid stigmatizing families and should be framed as an early-support tool rather than a diagnostic label. The main limitation is construct validity: "screen addiction" in toddlers is not a universally standardized clinical diagnosis. The model therefore predicts risk of problematic screen use rather than a formal disorder. Another limitation is confounding, since language delay, sleep problems, and family stress can influence both screen use and developmental outcomes. External validation across cultures and devices is also necessary because screen-use norms vary by region and socioeconomic environment.ucl.ac+4

VI. CONCLUSION AND FUTURE SCOPE

This paper presents a master's-level framework for predicting toddler screen-addiction risk from behavioral data using machine learning. The approach is grounded in WHO screen-time guidance and supported by recent evidence linking high screen exposure with language, behavioral, and socioemotional outcomes in young children. A behavior-driven AI model can help pediatricians, parents, and childcare providers identify early risk before screen habits become entrenched.pubmed.ncbi.nlm.nih+2

Future work should focus on longitudinal datasets, multimodal signals such as sleep and activity patterns, and culturally validated risk thresholds. Practical implementations could include pediatric screening dashboards, early-intervention mobile apps, and parental monitoring systems that provide actionable feedback rather than simple restriction alerts. With careful validation and ethical deployment, such tools could support healthier digital habits in early childhood and contribute to preventive pediatric care.pubmed.ncbi.nlm.nih+1

REFERENCES

[1] World Health Organization, “WHO guidelines on physical activity, sedentary behaviour and sleep for children under 5 years of age,” 2019. [Online]. Available: https://www.who.int/publications/i/item/9789241550536aoa+1

[2] American Optometric Association, “WHO guidance limit screen time for children,” summarizing WHO under-5 recommendations, 2019. [Online]. Available: https://www.aoa.org/news/clinical-eye-care/public-health/screen-time-for-children-under-5straitstimes+1

[3] J. M. Madigan et al., “Global prevalence of meeting screen time guidelines among children younger than 5 years,” JAMA Pediatrics, 2022. [Online]. Available: https://jamanetwork.com/journals/jamapediatrics/fullarticle/2789091linkedin+1

[4] K. A. H. et al., “Mobile device screen time is associated with poorer language development among toddlers: results from a large-scale survey,” PubMed, 2024. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/38622610/kkh+1

[5] “Effects of excessive screen time on child development,” PMC, 2023. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10353947/jccpractice+1

[6] “Impact of screen time on development of children,” PubMed, 2025. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/41153479/academia+1

[7] “Early childhood screen use contexts and cognitive and psychosocial outcomes,” JAMA Pediatrics, 2024. [Online]. Available: https://jamanetwork.com/journals/jamapediatrics/fullarticle/2821940pmc.ncbi.nlm.nih+1

[8] The Kids Research Institute Australia, “Study shows screen time is replacing vital language opportunities,” 2024. [Online]. Available: https://www.thekids.org.au/news--events/news-and-events-nav/2024/march/screen-time-replacing-vital-language-opportunities/thekids+1

[9] KK Women’s and Children’s Hospital, “Singapore integrated 24-hour activity guidelines for children under seven years,” 2022. [Online]. Available: https://www.kkh.com.sg/news/announcements/singapore-integrated-24-hour-activity-guidelines-for-children-under-seven-years-launchwjaets+1

[10] N. U. H. / related local report, “Balancing screen time: insights and impact on preschool children,” 2024. [Online]. Available: https://annals.edu.sg/balancing-screen-time-insights-and-impact-on-preschool-children/annals

[11] “Machine learning prediction of conduct problems in children using behavioral and environmental features,” PMC, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12883587/pmc.ncbi.nlm.nih+1

[12] “Applied machine learning to identify differential risk groups underlying externalizing and internalizing problem behaviors trajectories,” PLOS ONE, 2023. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0282235scribd+1

[13] "A machine learning framework for analyzing digital-era parenting behaviors," ACM Digital Library, 2025. [Online]. Available: https://dl.acm.org/doi/10.1145/3796731.3796768acm+1

[14] "Effect of screen time on language development among children," PubMed, 2026. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/41705007/pubmed.ncbi.nlm.nih+1

[15] "The impact of screen time on the risk of personal-social development," PMC, 2026. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12990812/cnn+1

[16] "Meeting the WHO 24-h guidelines among 2–6-year-old children," PMC, 2023. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9807983/pmc.ncbi.nlm.nih+1

[17] "Screen time effects on preschoolers," Pediatric Research, 2022. Cited via indexed summary source.scribd+1

[18] "Screen time is both a cause and symptom of kids' bad behavior," CNN, 2025. [Online]. Available: https://www.cnn.com/2025/06/09/health/screen-time-kids-emotional-problems-alaimo-wellnesstheguardian+1

[19] UCL News, "Toddlers spending two hours on screens a day," 2026. [Online]. Available: https://www.ucl.ac.uk/news/2026/jan/toddlers-spending-two-hours-screens-daybioinformation+1

[20] "Impact of screen time on language development and communication in preschool children," PubMed, 2025. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/41431552/jamanetwork+1

[21] Straits Times report on NUH research, "Prolonged digital screen use linked to higher chance of autism symptoms in toddlers," 2026. [Online]. Available: https://www.straitstimes.com/singapore/parenting-education/prolonged-digital-screen-use-linked-to-autism-symptoms-in-toddlers-nu...straitstimes+1