Journal Website

Article history:
Received 03 January 2026
Revised 18 March 2026
Accepted 26 March 2026
Published online 01 July 2026

Journal of Psychosociological Research in Family and Culture

Volume 4, Issue 3, pp 1-19

E-ISSN: 3041-8550

Early Identification of Families at Risk of Child Maltreatment: A Fairness-Audited Machine-Learning Model Integrating Caregiver Stress, Social Isolation, Economic Hardship, Family History, and Neighborhood Disadvantage

Petar. Ivanov[1], Salma. Abdelnour[2*], Mariana. Coutinho[3]

[1] Department of Developmental Psychology, Sofia University, Sofia, Bulgaria
[2] Department of Social Psychology, Helwan University, Helwan, Egypt
[3] Department of Clinical Psychology, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

*** Corresponding author email address**: salma.abdelnour@helwan.edu.eg

Article Info

Article type:
Original Research

How to cite this article:
Ivanov, P., Abdelnour, S., & Coutinho, MR. (2026). Early Identification of Families at Risk of Child Maltreatment: A Fairness-Audited Machine-Learning Model Integrating Caregiver Stress, Social Isolation, Economic Hardship, Family History, and Neighborhood Disadvantage. *Journal of Psychosociological Research in Psychological Studies*, 4(3), 1-19.
https://doi.org/10.61838/kman.jprfc.6237

ABSTRACT

Objective: This study aimed to develop and independently evaluate a fairness-audited machine-learning model for the early identification of Egyptian families at risk of child maltreatment.

Methods and Materials: A prospective multicenter cohort study was conducted among 1,248 families recruited from healthcare, educational, and social-service settings across Cairo, Giza, Alexandria, Dakahlia, Minya, and Sohag. Baseline assessments measured caregiver stress, social isolation, economic hardship, family adversity, previous child-protection involvement, and neighborhood disadvantage. Families were followed for 12 months to identify newly substantiated or independently adjudicated child-maltreatment concerns. The sample was divided into development, validation, and independent test datasets. Penalized logistic regression, elastic-net regression, random forest, gradient-boosted decision trees, and extreme gradient boosting were compared using discrimination, calibration, classification, and fairness metrics.

Findings: Families with a subsequent maltreatment concern had significantly higher parenting stress, social isolation, economic hardship, family adversity, neighborhood disadvantage, food insecurity, housing instability, intimate partner violence, and previous child-protection involvement than families without a concern (all $p < .001$). Extreme gradient boosting demonstrated the strongest independent test performance, with an area under the receiver operating characteristic curve of .852, an area under the precision–recall curve of .554, and a Brier score of .092. At the selected threshold, sensitivity was 82.1%, specificity was 79.2%, positive predictive value was 41.1%, and negative predictive value was 96.2%. Calibration was satisfactory, with an intercept of .03 and slope of .97. Performance was broadly comparable across caregiver sex, child sex, residence, and region, although the false-positive rate was higher among low-income families.

Conclusion: A multilevel fairness-audited machine-learning model can identify families requiring further supportive assessment with good discrimination and calibration, but predictions should supplement professional judgment and must not

be used as proof of maltreatment or as an independent basis for punitive intervention.
Keywords: *Child maltreatment; machine learning; early identification; caregiver stress; economic hardship; social isolation; neighborhood disadvantage; algorithmic fairness*

1 Introduction

Child maltreatment is a multidimensional public-health, social-welfare, and human-rights problem that emerges through the interaction of caregiver, child, household, community, and structural conditions. Its consequences extend beyond immediate physical injury or emotional distress and can shape neurodevelopment, psychological functioning, educational participation, interpersonal relationships, and later behavioral outcomes across the life course. Evidence linking adverse childhood experiences with antisocial, delinquent, and criminal behavior illustrates the enduring developmental significance of early exposure to abuse, neglect, and family adversity (Hesselink, 2023). Contemporary reviews of childhood adversity further show that adverse experiences do not constitute a single homogeneous construct; rather, they cluster into partially distinct domains such as deprivation, threat, household dysfunction, instability, and social disadvantage, each of which may influence development through different pathways (Willoughby et al., 2024). These findings have increased attention to prevention-oriented systems capable of recognizing elevated family risk before maltreatment becomes severe, recurrent, or visible to formal child-protection services. Early identification, however, is difficult because many of the conditions associated with maltreatment are common in families who never harm their children, while some families experiencing serious maltreatment may not display conventional warning signs. A defensible prediction strategy must therefore integrate multiple domains of risk, distinguish probability from proof, and avoid treating social disadvantage as equivalent to parental dangerousness.

Caregiver stress is one of the most consistent proximal mechanisms through which broader adversity can affect parenting behavior. A systematic review of parental risk and protective factors found that psychological distress, impaired coping, harsh parenting attitudes, substance-related problems, intimate partner violence, and limited social support frequently co-occur in families affected by maltreatment, whereas parental resilience, supportive relationships, and access to services may buffer risk (Younas & Gutman, 2022). Longitudinal evidence also indicates that income instability in early childhood may contribute to later mental-health difficulties partly through increased parenting stress and exposure to child maltreatment (Zhang et al., 2024). Stress may reduce caregivers' emotional availability, patience, behavioral regulation, and ability to respond consistently to children's needs, particularly when it is chronic and accompanied by economic insecurity or family conflict. Neurobehavioral research has shown that adversity is associated with decreased parent–child behavioral and neural synchrony, suggesting that stressful environments can disrupt reciprocal regulation within the caregiving relationship (Hoyniak et al., 2021). The developmental meaning of stress nevertheless depends on the resources surrounding the family. Work on risk and resilience in the context of childhood poverty emphasizes that adaptation reflects the operation of multiple systems, including caregiver functioning, family routines, schools, neighborhoods, and social institutions, rather than an isolated parental characteristic (Herbers et al., 2025). Similarly, research on adaptive systems during the COVID-19 period demonstrated that children's resilience is supported by coordinated family, school, and community processes that remain responsive during disruption (Herbers et al., 2021). Consequently, caregiver stress should be evaluated alongside relational and contextual resources rather than interpreted as a stand-alone marker of maltreatment.

Social isolation represents a second critical dimension because it can intensify caregiver burden while reducing access to emotional reassurance, practical assistance, respite, information, and informal monitoring. Socially disconnected caregivers may have fewer opportunities to obtain help before frustration escalates, and children in isolated households may have fewer trusted adults who can recognize emerging harm. During the COVID-19 pandemic, restrictions on mobility and face-to-face contact exposed the vulnerability created when families became disconnected from schools, health services, extended relatives, and community organizations. A global framework for maltreatment research and policy during the pandemic emphasized that changes in family stress, service access, surveillance, and social support had to be understood together (Katz et al., 2021). Reviews similarly described how pandemic conditions could increase maltreatment risk while simultaneously decreasing opportunities for detection

JPRFC
E-ISSN: 3041-8550

(Rokach & Chan, 2021). Telehealth home-visiting services became an important conduit for assessing family violence and maltreatment risk when conventional contact was restricted, illustrating the protective value of maintaining relational access to families during crises (Bullinger et al., 2021). Reports concerning intrafamilial child sexual abuse also indicated that isolation, reduced external oversight, and prolonged proximity to potential perpetrators could alter both risk and disclosure patterns (Tener et al., 2021). These observations remain relevant beyond the pandemic because social isolation is not merely the absence of social contact; it also reflects whether caregivers can mobilize dependable support and whether children remain connected to protective institutions.

Economic hardship is likewise central to maltreatment prevention, but its interpretation requires precision. Poverty does not cause maltreatment in a simple or deterministic manner, and most economically disadvantaged caregivers do not maltreat their children. Nevertheless, food insecurity, unstable employment, debt, housing insecurity, inability to meet medical or educational expenses, and repeated financial shocks can create sustained pressure that undermines caregiving capacity. Research involving Asian American and Pacific Islander families found that economic hardship and parental aggravation were important in understanding maltreatment-related processes (Thor et al., 2022). Evidence from kinship-care settings similarly showed that poverty and economic pressure were associated with children's behavioral health, while financial assistance could alter family conditions and outcomes (Xu et al., 2020). The importance of material support is also reflected in research connecting the generosity of childcare subsidies with child-maltreatment outcomes (Klika et al., 2022). Families investigated for maltreatment are themselves heterogeneous, and poverty status may distinguish different configurations of adversity rather than a single uniform risk profile (Maher et al., 2025). Latent-class research among socially disadvantaged families has likewise demonstrated distinct clusters of health and social vulnerability, showing why cumulative conditions should not be reduced to one income measure (Wong et al., 2022). Studies of children living under poverty and conflict further indicate that social instability and restricted resources affect well-being through interconnected family and environmental pathways (Kara & Selçuk, 2020). An early-identification model must therefore capture concrete forms of economic strain without converting poverty into a moral or forensic judgment about caregivers.

Family history and cumulative adversity provide another layer of predictive information. Previous child-protection involvement, prior maltreatment allegations, caregiver exposure to abuse in childhood, intimate partner violence, substance misuse, untreated mental-health problems, family separation, and criminal-justice involvement may indicate recurring stress processes or reduced access to stabilizing resources. Nationally representative evidence suggests that distinct neglect subtypes are associated with different configurations of family and caregiver risk, supporting a differentiated rather than undifferentiated approach to neglect (Chiang et al., 2023). Recent latent-class analysis of screened-in and screened-out child-welfare cases similarly identified different combinations of child, caregiver, and household factors associated with re-reporting, substantiation, and foster-care placement (Liu et al., 2025). These patterns support the use of multivariable methods capable of detecting interactions and nonlinear combinations that may be overlooked by simple additive checklists. At the same time, family history is ethically sensitive. Prior system contact can reflect true recurrence risk, but it can also reflect differential surveillance, reporting, service access, and institutional response. Families who have already entered child-welfare systems may be more likely to be observed and re-reported than equally stressed families who remain outside formal services. Service-needs research among caregivers involved in maltreatment investigations has shown that practical barriers, unmet needs, and system-related obstacles are integral to understanding family functioning and engagement (Rosenthal et al., 2022). Among immigrant families receiving home-visiting services, child-welfare involvement may also shape parental engagement in ways that complicate the assumption that prior system contact is a neutral risk indicator (Andom et al., 2025). Diversity in nativity and immigration status within families involved with child-protective services further demonstrates that well-being and service trajectories cannot be understood adequately without attention to social position and institutional context (Chiang et al., 2026).

Neighborhood disadvantage extends these concerns beyond the household. Communities differ in employment opportunities, housing stability, school quality, healthcare access, transportation, social services, exposure to violence, collective efficacy, and the availability of safe spaces for children. Research comparing family and neighborhood influences has shown that the relative importance of these factors can vary across critical stages of child development (Maguire-Jack & Sattler, 2022). Neighborhood poverty may

JPRFC Journal of Psychosociological Research in Family and Culture
E-ISSN: 3041-8550

affect maltreatment both directly and indirectly through family economic well-being, stress, service availability, and community support (Maguire-Jack et al., 2022). Longitudinal evidence connecting home foreclosures with neighborhood maltreatment rates further indicates that housing-market disruption can operate as an area-level stressor with implications for family safety (Hoyle et al., 2020). Material hardship may also influence maltreatment through community social cohesion, which can either buffer or transmit the effects of repeated poverty spells (Chung & Ahn, 2024). These findings argue for a multilevel understanding in which neighborhood conditions are neither ignored nor treated as inherent properties of individual families. An environmental-justice framework goes further by emphasizing that exposure to hazardous, under-resourced, and institutionally neglected environments is socially patterned and should be addressed as a structural condition rather than individualized as parental deficiency (Barboza-Salerno et al., 2025). The conceptualization of neglect as a collective failure to provide adequately for children similarly challenges systems that assign responsibility exclusively to caregivers while overlooking policy, service, and community failures (Blumenthal, 2021).

The pandemic literature illustrates how quickly these multilevel risks can converge. Reviews of the period from 2019 to 2022 documented substantial mental-health, educational, and social losses among children and adolescents, particularly when school interruption, family stress, and service disruption accumulated (Mucci et al., 2024). Child-maltreatment research during the pandemic identified implications for anxiety, depression, trauma, behavioral difficulties, and access to protective relationships (McDowell et al., 2024). The concept of toxic stress has been used to explain how prolonged and insufficiently buffered adversity can affect children's physiological and psychological development (Indriati et al., 2024). Mental-health disparities during COVID-19 were patterned by pre-existing social inequalities, differential exposure to stressors, and unequal access to protective resources (Parenteau et al., 2022). Children and youth were therefore not equally affected by the pandemic, and those already experiencing social or family vulnerability often encountered compounded risks (Collin-Vézina et al., 2023). Although the present study is not restricted to pandemic conditions, this body of evidence demonstrates the limitations of static, single-domain screening tools. Family risk can change rapidly as employment, housing, caregiving demands, social contact, and service access change, which supports the development of flexible models that combine indicators from several ecological levels.

Machine-learning methods offer a potentially useful framework for integrating heterogeneous predictors, capturing nonlinear relationships, and identifying interactions among caregiver stress, social isolation, economic hardship, family history, and neighborhood conditions. However, predictive performance alone is not an adequate standard for child-maltreatment applications. False negatives may leave children and families without timely support, whereas false positives may expose families to stigma, intrusive assessment, unnecessary surveillance, or punitive intervention. These harms are unlikely to be distributed evenly when model inputs are correlated with poverty, region, migration, gender, or prior institutional contact. Structural-risk scholarship has warned that systems may reproduce racial and socioeconomic disproportionality when structural adversity is rendered invisible and its consequences are attributed to individual families (Feely & Bosk, 2021). This concern is especially important when neighborhood disadvantage, service history, or material hardship function as proxies for protected or marginalized social positions. A fairness-audited model should therefore examine whether sensitivity, false-negative rates, false-positive rates, positive predictive value, and calibration differ across relevant groups rather than relying only on overall accuracy. It should also preserve the distinction between a predicted need for supportive assessment and a determination that maltreatment has occurred.

The Egyptian context makes this multilevel and fairness-oriented approach particularly important. Families may differ substantially in household composition, informal support, employment security, access to social protection, urbanization, service availability, and willingness or ability to disclose family difficulties. Formal records may capture only a subset of maltreatment concerns, while social stigma, fear of institutional consequences, limited service coverage, and variations in professional reporting can affect whose difficulties become visible. A model trained without attention to these mechanisms could classify families according to detectability rather than underlying need. It could also overidentify families living in disadvantaged neighborhoods while underidentifying socially isolated or distressed families whose material circumstances appear more stable. For these reasons, the present study conceptualized early identification as a preventive and supportive process rather than an automated child-protection judgment. The model was designed to estimate the probability of a newly identified safeguarding concern and to inform professional review, while fairness auditing was used to examine whether errors were concentrated among families defined by caregiver sex, child sex, household income, residential setting, governorate, or region.

The aim of this study was to develop and independently evaluate a fairness-audited machine-learning model for the early identification of Egyptian families at risk of child maltreatment by integrating caregiver stress, social isolation, economic hardship, family history, and neighborhood disadvantage.

2 Methods and Materials

2.1 Study Design and Participants

This study employed a prospective, multicenter cohort design to develop and validate a fairness-audited machine-learning model for the early identification of families at risk of child maltreatment. The study was conducted in Egypt between January 2024 and June 2025. Baseline recruitment and assessment were completed during the first six months of 2024, after which participating families were followed for 12 months to determine whether a new child maltreatment concern emerged. Participants were recruited from governmental primary healthcare centers, family health units, public schools, community-based social service organizations, and child and family welfare agencies located in Cairo, Giza, Alexandria, Dakahlia, Minya, and Sohag. These governorates were selected to provide representation of metropolitan, urban, peri-urban, and rural communities and to capture meaningful variation in socioeconomic conditions, access to services, population density, and neighborhood-level disadvantage. A multistage stratified sampling procedure was used. In the first stage, the participating governorates were purposively selected to represent the principal geographical regions of Egypt. In the second stage, healthcare, educational, and social service centers were selected within each governorate. In the final stage, potentially eligible families were identified consecutively through participating centers and invited to undergo eligibility screening.

A total of 1,436 families were approached for participation. Of these, 1,302 primary caregivers provided written informed consent and completed the initial eligibility assessment. Fifty-four families were excluded because they did not meet the eligibility criteria, withdrew before completing the baseline assessment, or provided insufficient information for linkage with the follow-up data. The final analytical sample therefore consisted of exactly 1,248 families. Each family contributed data concerning one index child. When more than one eligible child lived in the household, the child with the most recent birthday was selected to prevent caregivers from choosing the child they perceived as having the greatest or lowest level of risk. Primary caregivers were defined as adults who had assumed the principal responsibility for the child's daily care during the preceding six months. The index children were between 2 and 12 years of age at baseline. Eligible caregivers were required to be at least 18 years old, reside in the same household as the index child, possess sufficient Arabic-language proficiency to complete the assessment, and intend to remain within the study area during the 12-month follow-up period. Families were excluded when the index child was living in institutional care, when the caregiver had a severe cognitive impairment or acute psychiatric condition that prevented informed participation, or when the child had an already substantiated and actively investigated maltreatment case at the time of enrollment. Families with previous but closed child-protection involvement were not excluded because family history was one of the principal predictors examined in the study.

The sample size was determined before data collection based on the anticipated prevalence of newly identified maltreatment concerns, the number of candidate predictors, the need to estimate model discrimination and calibration with adequate precision, and the requirement to conduct subgroup-specific fairness analyses. A minimum sample of 1,200 families was considered necessary to obtain a sufficient number of outcome events, reduce the risk of model overfitting, and permit reliable evaluation across sex, economic, regional, and residential subgroups. The final sample of 1,248 families exceeded this minimum requirement. The data were divided into mutually exclusive development, validation, and test datasets using stratified random allocation that preserved the proportion of families with and without the outcome. The development dataset included 874 families, representing 70.0% of the sample. The validation dataset included 187 families, representing 15.0% of the sample, and the independent test dataset included the remaining 187 families, representing 15.0% of the sample. Allocation was performed at the family level, and no family contributed data to more than one dataset.

Data were collected by Arabic-speaking psychologists, social workers, and public-health researchers who completed standardized training in informed consent, sensitive interviewing, child safeguarding, confidentiality,

E-ISSN: 3041-8550

and the administration of all study instruments. Interviews were conducted privately at the recruitment centers or, when necessary, in participants' homes under conditions that protected privacy and minimized the possibility that another household member could influence the caregiver's responses. The average baseline assessment lasted approximately 60 to 75 minutes. Participants were informed that their decision to participate or decline would not affect their eligibility for healthcare, educational, or social services. The machine-learning risk scores were generated exclusively for research purposes and were not disclosed to caregivers, service providers, or child-protection personnel during the study. The model was not used to make legal, clinical, or child-custody decisions. Any immediate safeguarding concern disclosed during an interview was managed according to a predefined child-protection protocol and applicable professional reporting procedures, independently of the research model. Written informed consent was obtained from all participating caregivers. Age-appropriate assent was also obtained from children when direct child interviews were conducted.

2.2 Measures

Baseline data were collected using a structured assessment battery designed to measure the five principal predictor domains specified in the conceptual model: caregiver stress, social isolation, economic hardship, family history, and neighborhood disadvantage. A demographic and family information form was administered first to record the caregiver's age, sex, educational level, employment status, marital status, relationship to the index child, household size, number of dependent children, and place of residence. Information was also collected regarding the index child's age, sex, school attendance, chronic health conditions, developmental difficulties, and previous use of psychological, medical, educational, or social services. These variables were included to describe the sample, assess potential confounding, and evaluate whether the model performed differently across population subgroups. Sensitive demographic characteristics were retained in a protected audit dataset for fairness evaluation. They were not automatically incorporated as predictors in the final model, particularly when their inclusion did not provide a clinically meaningful improvement in prediction.

Caregiver stress was assessed using the Arabic version of the Parenting Stress Index–Short Form. This instrument contains 36 items distributed across the domains of parental distress, dysfunctional parent–child interaction, and perceptions of the child as difficult to manage. Caregivers rated each item on a five-point response scale ranging from strong disagreement to strong agreement. Item scores were summed to generate domain scores and an overall parenting stress score, with higher scores indicating greater stress associated with the caregiving role. In addition to the total score, the three domain scores were retained as candidate predictors because different forms of parenting stress could have distinct relationships with the subsequent risk of maltreatment. Responses were screened for inconsistent or highly defensive patterns before scoring. Interviewers were instructed to administer the instrument nonjudgmentally and to clarify that elevated stress did not imply that the caregiver had harmed or intended to harm the child.

Social isolation was measured using the Arabic six-item Lubben Social Network Scale and a brief service-access module. The social network scale assessed the number and availability of relatives and friends whom the caregiver could contact, trust, or ask for assistance. Lower scores represented smaller social networks and a greater risk of social isolation. The service-access module examined whether the caregiver had a reliable person available for emergency childcare, emotional support, financial assistance, transportation, or help during illness. It also assessed participation in neighborhood, religious, educational, or community activities and the frequency of meaningful social contact during the preceding month. A composite social isolation indicator was derived from the standardized social network score, frequency of contact, availability of practical assistance, and perceived access to community support. Both the continuous composite score and its individual components were considered during model development.

Economic hardship was evaluated through a structured household economic strain questionnaire covering income instability, unemployment, food insecurity, debt burden, housing instability, utility disconnection, inability to meet medical or educational expenses, and difficulty paying for essential household needs. Caregivers reported whether the household had experienced each condition during the preceding 12 months and rated the extent to which financial pressures had affected family functioning. Household income was recorded in ordered categories rather than as an exact amount to reduce nonresponse and discomfort. The assessment also documented the number of income earners, the regularity of employment, household crowding, dependence on informal borrowing, receipt of governmental

JPRFC Journal of Psychosociological Research in Family and Culture
E-ISSN: 3041-8550

or charitable support, and the occurrence of major financial shocks such as job loss, serious illness, eviction, or the death of an income-earning family member. A standardized economic hardship index was calculated by combining the severity and frequency of these indicators. Higher values represented more severe and persistent economic hardship. Income category was retained separately for fairness auditing because predictive errors affecting families with low incomes could create disproportionate harm.

Family history was assessed using a confidential, interviewer-administered Family Adversity and Maltreatment History Form developed for the study. The form documented the caregiver's self-reported exposure to physical abuse, emotional abuse, sexual abuse, neglect, or severe family violence during childhood. It also assessed intergenerational and household-level risk factors, including previous child-protection involvement, prior allegations or substantiated incidents of maltreatment involving any child in the household, intimate partner violence, caregiver substance misuse, serious untreated mental health problems, criminal justice involvement, repeated family separation, parental loss, and a history of placement in alternative care. Each history item was coded according to its presence, recency, frequency, and relationship to the index child. Because family history information was highly sensitive, caregivers were permitted to decline any question without withdrawing from the study. Interviewers received specific training in trauma-informed communication and were instructed to avoid pressing participants for unnecessary details. The family history variables were used to estimate prospective risk and were not treated as evidence that maltreatment had occurred in the current household.

Neighborhood disadvantage was assessed by linking the family's residential area to district-level socioeconomic and service-access indicators. Residential addresses were geocoded to the smallest administratively available area after removing direct identifiers from the analytical dataset. The neighborhood disadvantage index combined area-level indicators related to unemployment, educational disadvantage, housing quality, residential crowding, poverty concentration, access to primary healthcare and schools, availability of social services, transportation accessibility, and the physical condition of the surrounding environment. Area-level information was supplemented with caregiver reports concerning neighborhood safety, social cohesion, trust among neighbors, exposure to violence, access to child-friendly spaces, and the availability of community resources. Each component was standardized, and the standardized values were combined to produce an overall neighborhood disadvantage score. Higher scores indicated residence in communities with greater socioeconomic deprivation, fewer institutional resources, weaker perceived social cohesion, and more limited access to supportive services. Geographic coordinates and direct addresses were stored separately from the research dataset and were not used to identify individual families in the analysis.

The primary outcome was the occurrence of a newly identified child maltreatment concern during the 12-month follow-up period. Outcome determination was based on a multi-source safeguarding assessment rather than on a single questionnaire or unverified allegation. Follow-up information was collected at six and 12 months through caregiver interviews, age-appropriate child interviews when ethically and developmentally suitable, review of available healthcare and social service documentation, and structured reports from participating family welfare professionals. A family was classified as having a positive outcome when the follow-up period included a substantiated child-protection report or when an independent safeguarding panel determined that the available evidence met predefined criteria for a high-probability maltreatment concern. The panel consisted of two senior social workers, a child psychologist, and a pediatric healthcare professional who were not involved in model development and were blinded to the machine-learning predictions. The panel reviewed de-identified case summaries and reached decisions through consensus. The outcome included physical maltreatment, emotional maltreatment, sexual maltreatment, and neglect. Exposure to economic hardship or harsh parenting practices alone was not sufficient for a positive classification unless the predefined safeguarding criteria were met. Families without a substantiated or panel-confirmed concern during follow-up were classified as having a negative outcome.

To enhance data quality, ten percent of interviews were independently reviewed by a senior field supervisor, and electronic forms included automated range, logic, and consistency checks. Data collectors did not have access to the final outcome status or model predictions. Outcome adjudicators did not receive baseline predictor scores, except for information required to understand the safeguarding event under review. All participants were assigned numerical study identifiers, and names, telephone numbers, addresses, and other direct identifiers were stored in a separate encrypted file. Analytical data were accessed only by authorized members of the research team.

JPRFC Journal of Psychosociological Research in Family and Culture
E-ISSN: 3041-8550

2.3 Data Analysis

Data analysis was performed using a prespecified analytical plan that combined conventional statistical procedures, supervised machine-learning methods, explainability techniques, and algorithmic fairness auditing. Initial analyses examined the completeness, distribution, and plausibility of all variables. Continuous variables were summarized using means and standard deviations when approximately normally distributed and medians and interquartile ranges when markedly skewed. Categorical variables were described using frequencies and percentages. Differences between families with and without a follow-up maltreatment concern were explored using independent-samples tests, nonparametric tests, or chi-square tests as appropriate. These comparisons were used for descriptive purposes and not as the sole basis for predictor selection. Correlations and variance inflation factors were examined to identify highly redundant predictors. Implausible values were checked against the original electronic records before any correction was made.

Missing data patterns were evaluated separately within the development, validation, and test datasets. Variables with excessive missingness or evidence of unreliable measurement were excluded according to prespecified criteria. For retained predictors, missing values were imputed using procedures fitted exclusively within the development data to avoid information leakage. Continuous variables were imputed using median or model-based estimates, and categorical variables were imputed using the most frequent category or an explicitly coded missing category when the absence of information was considered potentially informative. Sensitivity analyses were conducted using multiple imputation by chained equations. All preprocessing operations, including imputation, scaling, encoding, and feature transformation, were incorporated into the analytical pipeline and learned from the development dataset only. The validation and test datasets were transformed using parameters estimated from the development data.

Candidate prediction models included penalized logistic regression, elastic-net logistic regression, random forest, gradient-boosted decision trees, and extreme gradient boosting. A conventional multivariable logistic regression model was retained as an interpretable benchmark. The principal predictor variables consisted of the total and domain-specific measures of caregiver stress, social isolation, economic hardship, family history, and neighborhood disadvantage. A limited set of prespecified demographic and family-context variables was also evaluated when theoretically justified. Protected characteristics were not automatically included in the prediction model. Caregiver sex, child sex, low-income status, urban or rural residence, governorate, and major geographic region were retained for subgroup evaluation and fairness auditing. The inclusion of any potentially sensitive characteristic as an active predictor required evidence of a substantial improvement in performance, a clear substantive justification, and confirmation that its inclusion did not worsen disparities in prediction errors.

Model hyperparameters were optimized within the development dataset using repeated stratified five-fold cross-validation. Stratification preserved the proportion of outcome events in each fold. Class imbalance was addressed through class-weighted loss functions, and alternative resampling approaches were examined in sensitivity analyses. Synthetic observations were not introduced into the independent test dataset. Feature selection and regularization were performed inside each cross-validation fold to prevent optimistic bias. The validation dataset was used to compare the leading candidate models, select the final classification threshold, and assess the consequences of emphasizing sensitivity over specificity. Because the purpose of the model was early identification rather than definitive diagnosis, threshold selection prioritized the detection of families who might benefit from supportive assessment while maintaining a level of specificity that would limit unnecessary referrals and stigmatization. The final model and threshold were then locked and evaluated once in the independent test dataset.

Predictive discrimination was evaluated using the area under the receiver operating characteristic curve and the area under the precision–recall curve. Classification performance was assessed using sensitivity, specificity, positive predictive value, negative predictive value, balanced accuracy, and the F1 score. Calibration was examined using calibration plots, the calibration intercept, the calibration slope, and the Brier score. Ninety-five percent confidence intervals were obtained using bootstrap resampling. Decision-curve analysis was used to estimate the potential net benefit of the model across a range of classification thresholds and to compare model-assisted assessment with strategies in which all families or no families were referred for further evaluation. Model performance was reported separately for the development, validation, and test datasets,

JPRFC Journal of Psychosociological Research in Family and Culture
E-ISSN: 3041-8550

although the independent test results were treated as the primary estimate of generalizability.

A formal fairness audit was conducted in the independent test dataset. Performance was assessed separately according to caregiver sex, child sex, household income category, urban or rural residence, governorate, and broad geographic region. For each subgroup, sensitivity, specificity, false-positive rate, false-negative rate, positive predictive value, negative predictive value, calibration, and the area under the receiver operating characteristic curve were calculated. Fairness was examined using differences and ratios in false-positive rates, false-negative rates, sensitivity, positive predictive value, and calibration between predefined comparison groups. Equal opportunity was evaluated by comparing sensitivity across groups, while equalized-odds performance was examined through the combined assessment of false-positive and false-negative rates. Predictive parity was evaluated through subgroup differences in positive predictive value. Calibration within groups was examined to determine whether families receiving similar predicted risks experienced similar observed outcome frequencies across the audited subgroups.

Fairness findings were interpreted in conjunction with sample size and confidence intervals rather than through a single numerical threshold. A disparity was considered potentially important when it was consistent across resampling analyses, clinically meaningful, and unlikely to be explained solely by a small subgroup size. When the initial model displayed substantial subgroup disparities, mitigation procedures were evaluated within the development and validation datasets. These procedures included reweighting underrepresented observations, adjustment of class weights, removal of proxy variables that contributed to unjustified disparities, calibration correction, and selection of a single group-blind threshold that provided a more acceptable balance between sensitivity and false-positive classifications. Group-specific decision thresholds were not adopted because they could create inconsistent referral standards and raise ethical concerns. The final model was selected based on its combined discrimination, calibration, clinical usefulness, interpretability, and fairness profile rather than on predictive accuracy alone.

Model interpretability was examined using global and individual-level explanation methods. Permutation importance was used to estimate the overall contribution of each predictor domain, while Shapley additive explanation values were calculated to describe the direction and magnitude of each variable's contribution to individual predictions. Partial dependence and accumulated local effect plots were inspected to identify nonlinear relationships and potentially implausible model behavior. Particular attention was given to whether neighborhood disadvantage, income, or family history dominated predictions in a manner that could reproduce structural disadvantage or generate stigmatizing classifications. Explanations were used as analytical aids and not as evidence of causality. The model was designed to estimate the probability of a future safeguarding concern and did not establish that any caregiver characteristic caused child maltreatment.

Several sensitivity analyses were conducted to evaluate the robustness of the findings. Models were re-estimated after excluding families with previous child-protection involvement, using alternative definitions of economic hardship and neighborhood disadvantage, restricting the outcome to formally substantiated cases, and applying multiple imputation for missing data. Additional analyses evaluated performance across recruitment settings and governorates to determine whether the model relied excessively on site-specific patterns. The final model was also compared with a reduced model containing only the five principal predictor domains to determine whether the addition of demographic and contextual covariates produced a meaningful improvement in performance. All statistical tests were two-sided, and a probability level below .05 was used for conventional inferential analyses. Machine-learning model selection, however, was based on cross-validated performance, calibration, clinical utility, and fairness criteria rather than statistical significance alone.

3 Findings and Results

The final analytical sample consisted of 1,248 families, all of whom had sufficient baseline and follow-up information to determine the primary outcome. The mean age of the participating caregivers was 36.82 years (SD = 7.41; range = 18–64 years). Most caregivers were women (n = 1,002, 80.3%), whereas 246 caregivers (19.7%) were men. Biological mothers constituted 922 participants (73.9%), biological fathers constituted 223 participants (17.9%), and 103 participants (8.3%) were grandparents, stepparents, or other relatives with primary caregiving responsibility. A total of 919 caregivers (73.6%) were married, while 329 (26.4%) were single, divorced, separated, or widowed. Regarding educational attainment, 321 caregivers (25.7%) had less than secondary education, 501 (40.1%) had completed secondary education, and 426 (34.1%) had completed postsecondary or university education. At baseline, 497 caregivers (39.8%) were unemployed or outside the formal labor market, and 459 families (36.8%) were classified as belonging to the low-income category. The mean household size was 5.18 persons (SD = 1.74), and the mean number of dependent children per household was 2.63 (SD = 1.19). The index children had a mean age of 7.21 years (SD = 2.93; range = 2–12 years); 637 children (51.0%) were boys and 611 (49.0%) were girls. A chronic health, developmental, or functional difficulty was reported for 173 children (13.9%). Of the participating families, 788 (63.1%) lived in urban or peri-urban areas and 460 (36.9%) lived in rural communities. The governorate distribution comprised 238 families from Cairo (19.1%), 224 from Giza (17.9%), 203 from Alexandria (16.3%), 194 from Dakahlia (15.5%), 201 from Minya (16.1%), and 188 from Sohag (15.1%). Previous but closed child-protection involvement was reported for 128 families (10.3%). Follow-up interviews were completed by 1,215 families (97.4%) at six months and by 1,189 families (95.3%) at 12 months; however, outcome information was available for all 1,248 families through the combined use of interviews, service records, and safeguarding-panel adjudication. During the 12-month follow-up period, 186 families (14.9%) met the predefined criteria for a newly identified child maltreatment concern, while 1,062 families (85.1%) did not meet the outcome criteria. Among the 186 positive cases, neglect was identified in 112 cases (60.2%), emotional maltreatment in 76 cases (40.9%), physical maltreatment in 61 cases (32.8%), and sexual maltreatment in 19 cases (10.2%); these categories were not mutually exclusive because more than one form of maltreatment could be identified within the same family. Stratified allocation resulted in 874 families in the model-development dataset, including 130 outcome-positive cases, 187 families in the validation dataset, including 28 outcome-positive cases, and 187 families in the independent test dataset, including 28 outcome-positive cases. The outcome prevalence remained approximately 14.9% in each dataset, indicating that the stratification procedure successfully preserved the distribution of the primary outcome.

Table 1

Baseline Psychosocial, Family, Economic, and Neighborhood Characteristics According to the Occurrence of a Child Maltreatment Concern During Follow-Up

Predictor	Total sample (N = 1,248)	No maltreatment concern (n = 1,062)	Maltreatment concern (n = 186)	p
Parenting Stress Index total score, M (SD)	82.17 (18.86)	78.40 (16.80)	103.70 (15.20)	< .001
Parental distress score, M (SD)	28.65 (7.18)	27.20 (6.40)	36.90 (5.70)	< .001
Dysfunctional parent–child interaction score, M (SD)	25.82 (6.72)	24.60 (6.10)	32.80 (5.80)	< .001
Difficult-child score, M (SD)	27.70 (7.30)	26.60 (6.90)	34.00 (6.30)	< .001
Lubben Social Network Scale score, M (SD)	14.82 (5.49)	15.70 (5.20)	9.80 (4.30)	< .001
Social isolation composite score, M (SD)	0.00 (0.95)	−0.16 (0.88)	0.92 (0.79)	< .001
Economic hardship index, M (SD)	4.31 (2.84)	3.80 (2.60)	7.20 (2.40)	< .001
Family adversity count, M (SD)	2.06 (1.87)	1.70 (1.60)	4.10 (2.00)	< .001
Neighborhood disadvantage score, M (SD)	0.00 (0.99)	−0.13 (0.96)	0.73 (0.84)	< .001
Perceived neighborhood social cohesion, M (SD)	3.04 (0.88)	3.20 (0.80)	2.10 (0.70)	< .001
Previous child-protection involvement, n (%)	128 (10.3)	71 (6.7)	57 (30.6)	< .001
Household food insecurity, n (%)	399 (32.0)	282 (26.6)	117 (62.9)	< .001
Housing instability during the previous year, n (%)	238 (19.1)	158 (14.9)	80 (43.0)	< .001
Caregiver exposure to intimate partner violence, n (%)	207 (16.6)	137 (12.9)	70 (37.6)	< .001

As shown in Table 1, families who subsequently met the criteria for a child maltreatment concern differed substantially from families without an identified concern across all five principal predictor domains. The mean total parenting stress score was 103.70 in the outcome-positive group, compared with 78.40 in the outcome-negative group, representing a mean difference of 25.30 points. This difference was reflected consistently across parental distress, dysfunctional parent–child interaction, and difficult-child perceptions. The largest domain-level difference was observed for parental distress, for which the positive-outcome group had a mean score of 36.90 compared with 27.20 among families without a maltreatment concern. Social support patterns also differed markedly. Families in the positive-outcome group had a substantially lower mean Lubben Social Network Scale score and a substantially higher social isolation composite score, indicating that these caregivers had fewer reliable relatives or friends, less frequent meaningful contact, and more limited access to emotional and practical assistance. Economic hardship was considerably more pronounced among families who experienced the outcome. Their mean economic hardship index was 7.20, compared with 3.80 among families without a maltreatment concern. Food insecurity was reported by 62.9% of the positive-outcome group, compared with 26.6% of the negative-outcome group, while housing instability was reported by 43.0% and 14.9%, respectively. The positive-outcome group also had a higher mean family adversity count, with an average of 4.10 adverse family-history indicators compared with 1.70 in the negative-outcome group. Previous child-protection involvement was present in 30.6% of families who developed a new concern, compared with 6.7% of families without a new concern. Similarly, caregiver exposure to intimate partner violence was almost three times as common in the positive-outcome group. Neighborhood conditions displayed the same general pattern: families with a subsequent concern lived in areas with greater measured disadvantage and reported lower neighborhood social cohesion. All observed group differences were statistically significant at $p < .001$. Nevertheless, the distributions of the predictor scores overlapped between the two outcome groups, indicating that no individual factor was sufficiently accurate to function as a stand-alone screening criterion and supporting the use of a multivariable model that integrated information across caregiver, family, economic, social, and neighborhood domains.

Table 2

Comparative Predictive Performance of the Candidate Machine-Learning Models in the Validation and Independent Test Datasets

Model	Validation AUROC	Validation AUPRC	Validation Brier score	Test AUROC	Test AUPRC	Test Brier score
Penalized logistic regression	.823	.501	.103	.812	.476	.107
Elastic-net logistic regression	.831	.515	.100	.820	.489	.104
Random forest	.847	.541	.096	.833	.515	.100
Gradient-boosted decision trees	.858	.562	.091	.845	.540	.095
Extreme gradient boosting	.866	.579	.088	.852	.554	.092

All candidate models demonstrated meaningful predictive ability, but the tree-based boosting models consistently outperformed the regression-based models in both the validation and independent test datasets. Penalized logistic regression produced a test AUROC of .812 and a test AUPRC of .476, while elastic-net logistic regression produced slightly higher values of .820 and .489, respectively. Random forest improved test discrimination to an AUROC of .833 and an AUPRC of .515. Gradient-boosted decision trees yielded a test AUROC of .845, an AUPRC of .540, and a Brier score of .095. Extreme gradient boosting achieved the strongest overall performance, with a validation AUROC of .866 and an independent test AUROC of .852. Its test AUROC had a bootstrap 95% confidence interval of .789 to .911. The model's test AUPRC was .554, which was substantially higher than the outcome prevalence of .149 and indicated that the model provided considerably greater precision than would be achieved through random classification. Extreme gradient boosting also produced the lowest test Brier score of .092, showing that it generated more accurate probability estimates than the competing models. Consequently, the extreme gradient boosting model was selected as the final model, although its selection was based jointly on discrimination, calibration, decision-analytic utility, interpretability, and fairness rather than on AUROC alone. At the prespecified probability threshold of .18, the model correctly identified 23 of the 28 outcome-positive families in the independent test dataset and correctly classified 126 of the 159 outcome-negative families. This corresponded to a sensitivity of 82.1%, a specificity of 79.2%, a false-negative rate of 17.9%, and a false-positive rate of 20.8%. The positive predictive value was 41.1%, meaning that approximately four of every ten families classified as higher risk experienced a confirmed

JPRFC Journal of Psychosociological Research in Family and Culture
E-ISSN: 3041-8550

maltreatment concern during follow-up. The negative predictive value was 96.2%, indicating that families classified below the threshold were highly unlikely to experience the outcome during the follow-up period. Balanced accuracy was 80.7%, and the F1 score was .548. The relatively moderate positive predictive value was expected because the outcome occurred in fewer than one-sixth of participating families and underscored that the model should be used to trigger supportive professional assessment rather than to classify a family as maltreating a child. Decision-curve analysis indicated that the final model provided greater net benefit than referring all families or referring no families across predicted-risk thresholds ranging from approximately .10 to .34, including the selected threshold of .18.

Table 3

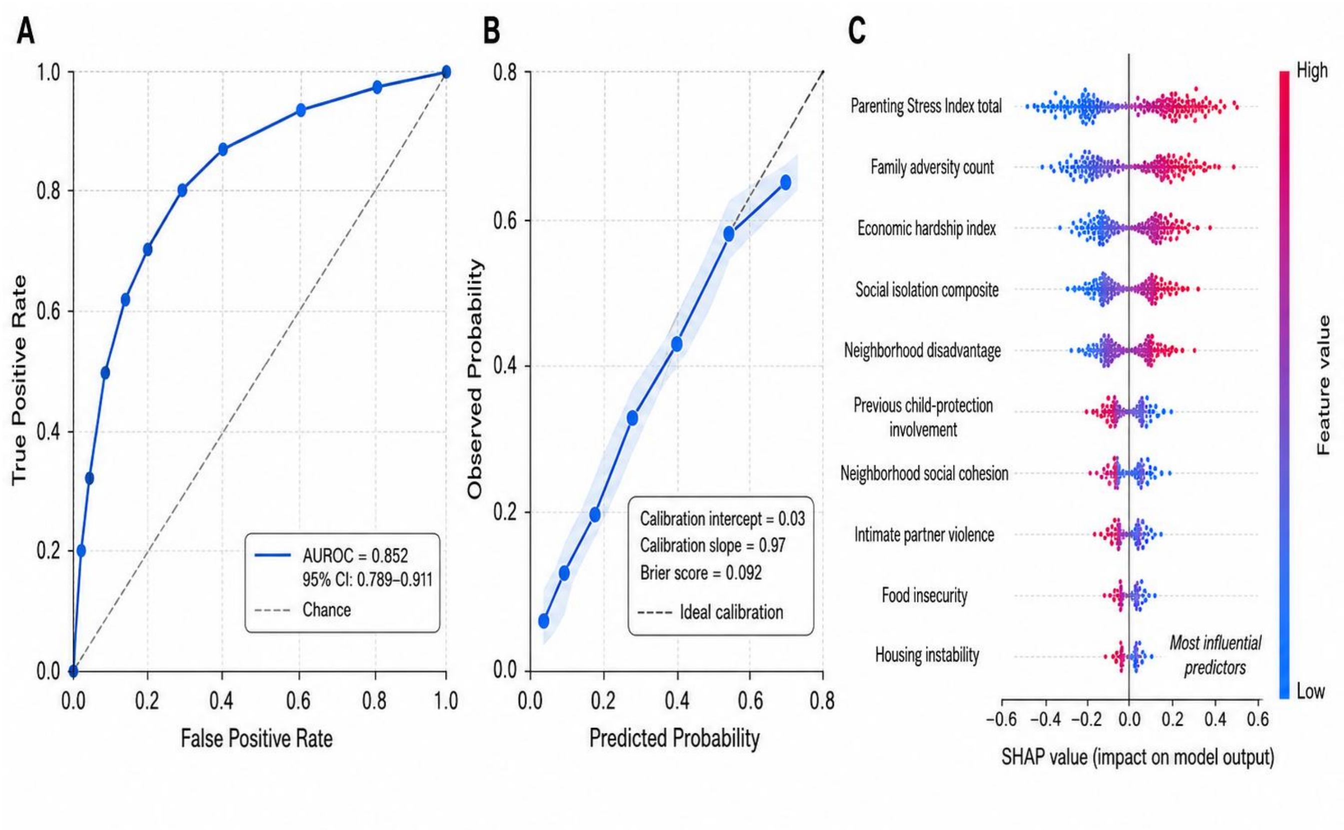
Fairness Audit of the Final Extreme Gradient Boosting Model in the Independent Test Dataset

Audited group	n	Outcome events	AUROC	Sensitivity (%)	Specificity (%)	False-positive rate (%)	False-negative rate (%)	Positive predictive value (%)	Calibration slope
Overall test sample	187	28	.852	82.1	79.2	20.8	17.9	41.1	.97
Female caregiver	150	23	.850	82.6	78.7	21.3	17.4	41.3	.95
Male caregiver	37	5	.858	80.0	81.3	18.7	20.0	40.0	1.01
Female child	91	13	.861	84.6	79.5	20.5	15.4	40.7	.99
Male child	96	15	.844	80.0	79.0	21.0	20.0	41.4	.95
Low-income household	69	17	.838	82.4	73.1	26.9	17.6	50.0	.93
Middle- or higher-income household	118	11	.856	81.8	82.2	17.8	18.2	32.1	1.00
Urban or peri-urban residence	118	18	.850	83.3	78.0	22.0	16.7	40.5	.96
Rural residence	69	10	.855	80.0	81.4	18.6	20.0	42.1	.99
Metropolitan or Lower Egypt	129	18	.854	83.3	78.4	21.6	16.7	38.5	.96
Upper Egypt	58	10	.847	80.0	81.3	18.8	20.0	47.1	.99
Cairo	36	6	.861	83.3	80.0	20.0	16.7	45.5	.97
Giza	34	5	.846	80.0	79.3	20.7	20.0	40.0	.93
Alexandria	31	4	.872	75.0	81.5	18.5	25.0	37.5	1.04
Dakahlia	29	4	.856	100.0	80.0	20.0	0.0	44.4	1.08
Minya	30	5	.842	80.0	80.0	20.0	20.0	44.4	.95
Sohag	27	4	.829	75.0	73.9	26.1	25.0		

The fairness audit indicated that the final model maintained broadly comparable discrimination, sensitivity, and calibration across the principal audited subgroups, although some differences in false-positive rates and positive predictive values remained. Model discrimination was similar for families with female and male caregivers, with AUROC values of .850 and .858, respectively. The sensitivity difference between these groups was 2.6 percentage points, and the false-positive-rate difference was also 2.6 percentage points. Performance was similarly balanced according to the child's sex. Sensitivity was 84.6% for female children and 80.0% for male children, while specificity was 79.5% and 79.0%, respectively. The difference in positive predictive value between the two child-sex groups was less than one percentage point. Urban and rural families also had similar discrimination, with AUROC values of .850 and .855. Sensitivity was slightly higher among urban and peri-urban families, whereas specificity was slightly higher among rural families. The broad regional comparison showed an AUROC of .854 for Metropolitan or Lower Egypt and .847 for Upper Egypt, with a sensitivity difference of 3.3 percentage points and a false-positive-rate difference of 2.8 percentage points. These results did not indicate a systematic deterioration of the model's performance among rural families or families living in Upper Egypt. The most notable disparity was observed according to household income. Sensitivity was nearly identical for low-income and middle- or higher-income families, at 82.4% and 81.8%, respectively, indicating that the model did not miss substantially more positive cases in either economic group. However, specificity was lower among low-income families, resulting in a false-positive rate of 26.9%, compared with 17.8% among middle- or higher-income families. The positive predictive value was nevertheless higher in the low-income group because the observed outcome prevalence was also higher in that subgroup. During model development, the initial extreme gradient boosting specification produced a false-positive-rate difference of 15.7 percentage points between income groups. Removal of household income category as a direct model input, retention of more specific hardship indicators, and reweighting of observations during training reduced this difference to 9.1 percentage points while decreasing the validation AUROC by only .004. The final model was therefore preferred because it achieved a more acceptable balance between predictive performance and distributive fairness. Calibration slopes ranged from .93 to 1.01 across the principal caregiver-sex, child-sex, income, residence, and regional groups, suggesting that predicted probabilities retained broadly comparable meanings across these populations. Governorate-level sensitivity ranged from 75.0% to 100.0%, but these estimates were based on only four to six outcome events per governorate and consequently had substantial uncertainty. The apparently perfect sensitivity in Dakahlia should therefore not be interpreted as evidence of superior regional performance. Overall, the audit did not reveal a large disparity in false-negative rates across the major protected or socially relevant groups, but the elevated false-positive rate among low-income families remained ethically important. This finding reinforced the requirement that a positive model classification should lead only to a voluntary, supportive, and professionally reviewed family assessment and should never be treated as proof of maltreatment or used independently to initiate punitive intervention.

Figure 1

Receiver Operating Characteristic Curve, Calibration Plot, and Shapley Additive Explanation Summary for the Final Extreme Gradient Boosting Model in the Independent Test Dataset



The receiver operating characteristic component of Figure 1 demonstrates that the final model maintained good discrimination across a range of possible classification thresholds, with an AUROC of .852 in the independent test dataset. The precision–recall analysis similarly showed that model precision remained meaningfully above the outcome prevalence across the principal range of recall values. The calibration plot indicated close agreement between estimated and observed risk, particularly for predicted probabilities below .40, where most families were located. The calibration intercept was .03, the calibration slope was .97, and the Brier score was .092, indicating minimal systematic overprediction and only limited compression of the predicted risk distribution. Some uncertainty was observed in the highest predicted-risk interval because relatively few families received probabilities above .60, but the observed event frequency remained within the corresponding bootstrap confidence interval. Shapley additive explanation analysis showed that total parenting stress was the most influential predictor of individual risk estimates, followed by the family adversity count, economic hardship, social isolation, and neighborhood disadvantage. Previous child-protection involvement, low neighborhood social cohesion, caregiver exposure to intimate partner violence, food insecurity, and housing instability also contributed meaningfully to predictions. The model exhibited nonlinear risk patterns rather than assuming that every one-unit increase in a predictor produced an identical change in risk. Predicted risk increased most sharply when the Parenting Stress Index total score exceeded approximately 95, when the economic hardship index exceeded 6, when the standardized social isolation score exceeded approximately 0.75, and when several family adversity indicators occurred simultaneously. Neighborhood disadvantage had a more moderate independent contribution but amplified predicted risk when it co-occurred with severe caregiver stress, social isolation, or economic instability. Conversely, larger and more reliable social networks and stronger perceived neighborhood cohesion generally reduced predicted risk, even among families experiencing financial strain. Examination of individual explanations did not identify a single predictor that automatically determined classification. Instead, higher-risk predictions generally resulted from the accumulation and interaction of multiple difficulties. Review of the Shapley distributions also indicated that the model did not treat low income or neighborhood residence alone as sufficient evidence of maltreatment risk, which was important for reducing the possibility that structural poverty would be interpreted as individual wrongdoing.

4 Discussion

The present study developed and independently evaluated a fairness-audited machine-learning model for the early identification of Egyptian families at risk of child maltreatment by integrating caregiver stress, social isolation, economic hardship, family history, and neighborhood disadvantage. The findings demonstrated that families who experienced a newly identified maltreatment concern during the 12-month follow-up period differed substantially from families without such concerns across all five predictor domains. Outcome-positive families reported markedly higher parenting stress, greater parental distress, more dysfunctional parent–child interaction, stronger perceptions of the child as difficult, more severe social isolation, greater economic hardship, a higher accumulation of family adversity, and residence in more disadvantaged neighborhoods. They were also more likely to have experienced food insecurity, housing instability, intimate partner violence, and previous child-protection involvement. Among the evaluated algorithms, the extreme gradient boosting model produced the strongest overall performance, achieving an area under the receiver operating characteristic curve of .852 and an area under the precision–recall curve of .554 in the independent test dataset. At the selected classification threshold, the model achieved a sensitivity of 82.1%, specificity of 79.2%, and negative predictive value of 96.2%. The model also demonstrated satisfactory calibration, with a calibration intercept of .03, a calibration slope of .97, and a Brier score of .092. The fairness audit indicated broadly similar discrimination and false-negative rates across caregiver sex, child sex, residential setting, and broad geographic region, although the false-positive rate remained higher among low-income families. Collectively, these findings suggest that multilevel prediction may assist early supportive assessment, provided that model outputs are interpreted cautiously and subjected to professional and ethical oversight.

Caregiver stress emerged as the most influential predictor in the final model, and families with a subsequent maltreatment concern had substantially higher total parenting stress and domain-specific stress scores than families without an identified concern. This finding is consistent with evidence showing that parental psychological distress, impaired emotion regulation, aggravation, and reduced coping capacity constitute important proximal mechanisms in child-maltreatment risk (Younas & Gutman, 2022). Economic uncertainty and

JPRFC Journal of Psychosociological Research in Family and Culture
E-ISSN: 3041-8550

unstable family conditions may increase stress while simultaneously reducing caregivers' ability to respond sensitively and consistently to children's needs. Longitudinal research has similarly indicated that early income instability can influence later child mental health through parenting stress and maltreatment-related pathways (Zhang et al., 2024). The present findings also align with research showing that adversity may reduce behavioral and neural synchrony between caregivers and children, potentially weakening reciprocal regulation and increasing the likelihood of conflictual interactions (Hoyniak et al., 2021). The nonlinear pattern observed in the machine-learning explanations was particularly important: predicted risk increased more sharply when parenting stress reached relatively high levels and co-occurred with other disadvantages. This suggests that moderate stress alone should not be interpreted as evidence of maltreatment risk. Rather, the combination of intense stress, limited social support, material insecurity, and family adversity appears to create a more consequential pattern. Such an interpretation is consistent with resilience-oriented perspectives emphasizing that developmental outcomes depend on interacting family, school, neighborhood, and institutional systems (Herbers et al., 2021; Herbers et al., 2025).

Social isolation also made a substantial contribution to prediction. Families in the maltreatment-concern group had smaller social networks, fewer dependable sources of practical and emotional assistance, and lower perceived neighborhood social cohesion. These findings support the view that social relationships may protect families by distributing caregiving demands, providing respite, facilitating access to services, and enabling early recognition of escalating difficulties. The pandemic literature provides a clear illustration of these mechanisms. Research conducted during periods of social restriction showed that reduced contact with schools, healthcare professionals, relatives, and community organizations could increase family vulnerability while simultaneously weakening systems of detection (Katz et al., 2021; Rokach & Chan, 2021). Telehealth home-visiting programs helped preserve access to families during this period and created opportunities to assess violence and maltreatment concerns when face-to-face services were unavailable (Bullinger et al., 2021). Studies of intrafamilial sexual abuse similarly emphasized how prolonged isolation and reduced contact with external adults could alter both the occurrence and disclosure of abuse (Tener et al., 2021). The present results extend this reasoning beyond emergency conditions by showing that ordinary variations in the availability of social and community support may meaningfully distinguish families who later experience safeguarding concerns. The protective association of neighborhood cohesion further suggests that informal trust and collective support may mitigate some effects of economic pressure.

Economic hardship was another prominent predictor, and the maltreatment-concern group reported substantially greater food insecurity, housing instability, debt-related pressure, and difficulty meeting basic needs. This finding is compatible with previous research linking economic hardship and parental aggravation with maltreatment-related outcomes (Thor et al., 2022). It is also consistent with evidence from kinship-care settings showing that financial pressure can affect children's behavioral health and that material assistance may alter family functioning (Xu et al., 2020). The observed importance of food and housing instability indicates that concrete deprivation may be more informative than income category alone. Families with similar reported incomes may differ considerably in debt, housing costs, employment regularity, access to public assistance, or the occurrence of sudden financial shocks. The use of a multidimensional hardship index therefore allowed the model to capture forms of strain that would have been obscured by a single income variable. Previous studies have shown that families investigated for maltreatment form heterogeneous subgroups and that poverty status does not represent one uniform pattern of adversity (Maher et al., 2025). Latent-class analyses among socially disadvantaged families have similarly identified different combinations of health, social, and economic vulnerability (Wong et al., 2022). The association between childcare subsidy generosity and maltreatment outcomes additionally supports the proposition that economic and service policies may influence family safety by reducing caregiving strain (Klika et al., 2022). Accordingly, the present model should not be interpreted as demonstrating that poverty itself causes maltreatment; instead, it suggests that severe and cumulative material hardship can contribute to risk when it interacts with caregiver stress, isolation, and family adversity.

Family adversity and previous child-protection involvement also contributed strongly to the model. Families with a newly identified concern had a higher number of prior adversities and were considerably more likely to have had previous contact with child-protection services. This is consistent with evidence that different forms of neglect are associated with distinct combinations of caregiver, household, and contextual risk factors (Chiang et al., 2023).

JPRFC E-ISSN: 3041-8550

It also aligns with recent latent-class findings showing that combinations of child, caregiver, and household characteristics are associated with re-reporting, substantiation, and placement outcomes across both screened-in and screened-out cases (Liu et al., 2025). The present findings therefore support a cumulative-risk approach rather than reliance on any single historical indicator. Nevertheless, previous child-welfare involvement requires especially cautious interpretation because it may reflect institutional visibility as well as underlying family difficulty. Families already known to services are more likely to be observed, documented, and re-reported than families experiencing similar problems outside formal systems. Research involving immigrant families has shown that child-welfare involvement can influence engagement with home-visiting services, indicating that system contact may alter family behavior and service relationships (Andom et al., 2025). Studies of nativity and immigration diversity within child-protective systems further demonstrate that family well-being and institutional pathways may vary across social and cultural contexts (Chiang et al., 2026). Service-needs research also indicates that caregivers involved in maltreatment investigations frequently face substantial practical barriers and unmet needs (Rosenthal et al., 2022). For these reasons, family history should guide supportive assessment rather than function as a permanent label of parental risk.

Neighborhood disadvantage had a meaningful but less dominant independent contribution than caregiver stress, family adversity, and economic hardship. Its effect became stronger when it co-occurred with severe household-level difficulties, suggesting that neighborhood conditions may amplify or buffer family stress. This pattern is consistent with evidence that family and neighborhood factors have different levels of influence across developmental stages and should be examined jointly (Maguire-Jack & Sattler, 2022). Neighborhood poverty can affect maltreatment risk indirectly through household economic well-being, service availability, exposure to violence, and the availability of supportive social relationships (Maguire-Jack et al., 2022). The lower social cohesion observed among outcome-positive families also accords with findings that cohesion can mediate the relationship between recurring material hardship and maltreatment (Chung & Ahn, 2024). Research connecting neighborhood foreclosure rates with child-maltreatment rates further illustrates how area-level housing disruption may influence family safety beyond individual household characteristics (Hoyle et al., 2020). The present results therefore support the inclusion of neighborhood information, but they also highlight the danger of interpreting residence in a disadvantaged area as an individual risk trait. Environmental-justice perspectives emphasize that neighborhood deprivation reflects unequal exposure to under-resourced and hazardous conditions rather than inherent deficiencies among residents (Barboza-Salerno et al., 2025). Similarly, theoretical work conceptualizing neglect as a collective failure to provide for children argues that social institutions and public policies share responsibility for conditions affecting children's welfare (Blumenthal, 2021). In the present model, neighborhood disadvantage alone did not determine high-risk classification, which reduced—but did not eliminate—the possibility of stigmatizing communities.

The superior performance of extreme gradient boosting compared with penalized and elastic-net logistic regression suggests that relationships among the predictors were partly nonlinear and interactive. Traditional regression models remain valuable because of their transparency, but they may be less capable of representing threshold effects, interactions, and complex cumulative-risk configurations without extensive prespecification. The Shapley explanation analysis showed that higher-risk predictions generally resulted from the simultaneous presence of multiple difficulties rather than from a single feature. This observation is compatible with research on childhood adversity demonstrating that adverse experiences form several overlapping domains rather than a unitary construct (Willoughby et al., 2024). It also corresponds with evidence connecting accumulated adverse childhood experiences to later behavioral and psychosocial difficulties (Hesselink, 2023). However, the predictive performance of the model should not be mistaken for causal explanation. An AUROC of .852 indicates good discrimination, not certainty, and the positive predictive value of 41.1% means that a substantial proportion of families classified above the threshold did not experience a confirmed concern during follow-up. The model is therefore better understood as a mechanism for identifying families who may benefit from additional assessment or voluntary support than as a diagnostic or forensic instrument.

The high negative predictive value suggests that the model was effective in identifying families unlikely to experience a recorded maltreatment concern within the follow-up period. Nevertheless, five of the 28 positive cases in the test dataset were classified below the threshold. These false-negative cases are important because missed concerns

JPRFC Journal of Psychosociological Research in Family and Culture
E-ISSN: 3041-8550

may delay access to protection and services. At the same time, the 33 false-positive classifications demonstrate that increasing sensitivity entails costs, including unnecessary anxiety, stigmatization, or intrusive scrutiny. These consequences are particularly significant in light of pandemic-era research showing that child and adolescent mental-health difficulties, service disruption, and family stress were distributed unequally across social groups (McDowell et al., 2024; Mucci et al., 2024; Parenteau et al., 2022). Children already facing disadvantage were not equally positioned to cope with large-scale disruption (Collin-Vézina et al., 2023), and prolonged adversity could contribute to toxic-stress processes (Indriati et al., 2024). Models developed for prevention must therefore avoid intensifying existing inequities by treating families affected by structural hardship as intrinsically dangerous.

5 Conclusion

The fairness audit provided encouraging but qualified evidence. The model demonstrated similar sensitivity, specificity, and discrimination across caregiver sex, child sex, urban or rural residence, and broad region. This suggests that the final reweighted and calibrated model did not systematically miss concerns in one of these major subgroups. However, the false-positive rate was approximately nine percentage points higher among low-income families than among middle- or higher-income families. Although this disparity was smaller than in the initial model and sensitivity was nearly identical across income groups, the remaining difference is ethically important. Structural-risk scholarship has emphasized that inequity in child-welfare systems can be reproduced when poverty and systemic disadvantage are interpreted as evidence of individual parental failure (Feely & Bosk, 2021). The higher false-positive rate may partly reflect the correlation between low income and multiple genuine stressors, but it may also indicate that the model still relies on variables functioning as proxies for structural disadvantage. The fairness findings therefore support continued monitoring, periodic recalibration, and human review rather than unqualified deployment. They also reinforce the importance of ensuring that a positive prediction leads to supportive resource assessment, not automatic investigation or punitive action.

Several limitations should be considered when interpreting the findings. First, although the study included families from six Egyptian governorates and represented urban, peri-urban, and rural areas, the sample was not nationally representative, and the model may perform differently in governorates with distinct cultural, economic, or service conditions. Second, several predictors were based on caregiver self-report and may have been affected by recall error, social desirability, fear of disclosure, or differences in interpreting sensitive questions. Third, the outcome combined formally substantiated cases with concerns adjudicated by an expert safeguarding panel. Although this approach improved identification beyond administrative records alone, it may also have introduced classification uncertainty. Fourth, the number of outcome events in the independent test dataset was relatively small, particularly within governorate-level subgroups, which widened uncertainty around fairness estimates. Fifth, families with active substantiated cases at baseline were excluded, meaning that the model addressed the emergence of new concerns rather than the full spectrum of current maltreatment. Sixth, the follow-up period was limited to 12 months and may not have captured concerns that emerged later. Finally, the study assessed internal holdout performance rather than external validation in a fully independent service system, and the observed results should not be interpreted as sufficient evidence for immediate operational deployment.

Future research should externally validate the model using larger samples recruited from additional Egyptian governorates and from healthcare, educational, social-service, and community settings that were not involved in model development. Longer follow-up periods would permit examination of the timing, recurrence, and severity of maltreatment concerns and might support the development of dynamic models that update risk as family circumstances change. Future studies should compare the model with existing professional screening procedures and evaluate whether it improves decision quality beyond routine assessment. Research should also examine separate outcomes for neglect, physical maltreatment, emotional maltreatment, and sexual maltreatment because these forms of harm may have different predictor structures. Larger subgroup samples are needed to conduct more precise intersectional fairness analyses involving income, sex, region, disability, family structure, migration background, and service history. Participatory studies involving caregivers, young people, social workers, healthcare professionals, and child-rights specialists should investigate which model uses are considered acceptable and which safeguards are required. Prospective implementation trials

JPRFC Journal of Psychosociological Research in Family and Culture
E-ISSN: 3041-8550

should additionally assess unintended consequences, including changes in referral rates, service burden, family trust, labeling, and unequal access to preventive resources.

In practice, the model should be used only as a decision-support instrument for identifying possible needs and should never be treated as proof that maltreatment has occurred. Positive classifications should initiate a confidential, strengths-based review conducted by trained professionals who can consider context, verify information, and distinguish material hardship from intentional or harmful caregiving behavior. Families identified as having elevated risk should first be offered voluntary supports matched to the underlying factors contributing to the prediction, such as parenting assistance, mental-health care, food and housing support, childcare, domestic-violence services, and opportunities for social connection. Organizations using predictive tools should establish transparent governance procedures, document the purpose and limits of the model, restrict access to sensitive information, and prohibit its independent use in legal, custody, or removal decisions. Performance and fairness should be reviewed regularly, particularly for low-income families and geographically underrepresented groups. Practitioners should also receive training in algorithmic limitations, structural disadvantage, trauma-informed communication, and the interpretation of probabilistic outputs so that technology strengthens professional judgment without replacing it.

Authors' Contributions

All authors have contributed significantly to the research process and the development of the manuscript.

Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

We would like to express our gratitude to all individuals helped us to do the project.

Declaration of Interest

The authors report no conflict of interest.

Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

Ethical Considerations

The study protocol adhered to the principles outlined in the Helsinki Declaration, which provides guidelines for ethical research involving human participants.

References

Andom, F., Jonson-Reid, M., Henk, J., & Kemner, A. (2025). The Influence of Child Welfare Involvement on Parent Engagement Among Immigrant Families Who Receive Home Visiting Services. *Child maltreatment*, *31*(3), 503-514. https://doi.org/10.1177/10775595251368358

Barboza-Salerno, G., Duhaney, S., Ramesh, B., & Stanek, C. (2025). Beyond Social Disadvantage: Advancing an Environmental Justice Framework to Address Child Maltreatment Risk. https://doi.org/10.21203/rs.3.rs-6521185/v1

Blumenthal, A. (2021). Neglect as Collective Failure to Provide for Children: Towards a New Theoretical Approach. https://doi.org/10.31235/osf.io/5afv2

Bullinger, L. R., Marcus, S., Reuben, K., Whitaker, D. J., & Self-Brown, S. (2021). Evaluating Child Maltreatment and Family Violence Risk During the COVID-19 Pandemic: Using a Telehealth Home Visiting Program as a Conduit to Families. *Infant Mental Health Journal Infancy and Early Childhood*, *43*(1), 143-158. https://doi.org/10.1002/imhj.21968

Chiang, C.-J., Kim, H., Jonson-Reid, M., Yang, M. Y., Moon, C. A., & Kohl, P. L. (2023). Risk Factors and Neglect Subtypes: Findings From a Nationally Representative Data Set. *American Journal of Orthopsychiatry*, *93*(6), 532-542. https://doi.org/10.1037/ort0000698

Chiang, C. J., Rafieifar, M., Chen, J. H., & Yang, MR. (2026). Understanding Child Well-being In CPS: Nativity and Immigration Status Diversity in Latino Families. *Family Relations*. https://doi.org/10.1111/fare.70142

Chung, Y., & Ahn, H. (2024). Understanding the Mediation Effect of Social Cohesion on the Relationship Between Material Hardship and Child Maltreatment by Poverty Spells. *Child maltreatment*, *30*(3), 394-405. https://doi.org/10.1177/10775595241307896

Collin-Vézina, D., Fallon, B., & Caldwell, J. (2023). Children and Youth Mental Health: Not All Equal in the Face of the COVID-19 Pandemic. 377-385. https://doi.org/10.1016/b978-0-323-91497-0.00072-2

Feely, M., & Bosk, E. A. (2021). That Which Is Essential Has Been Made Invisible: The Need to Bring a Structural Risk Perspective to Reduce Racial Disproportionality in Child Welfare. *Race and Social Problems*, *13*(1), 49-62. https://doi.org/10.1007/s12552-021-09313-8

Herbers, J. E., Hayes, K. R., & Cutuli, J. J. (2021). Adaptive Systems for Student Resilience in the Context of COVID-19.

JPRFC Journal of Psychosociological Research in Family and Culture
E-ISSN: 3041-8550

School Psychology, *36*(5), 422-426. https://doi.org/10.1037/spq0000471
Herbers, J. E., Wallace, L. E., & Tebepah, T. C. (2025). Risk and Resilience From Childhood Experiences of Stress and Poverty. 323-341. https://doi.org/10.1037/0000396-018
Hesselink, A. (2023). Adverse Childhood Experiences (ACEs) and the Link to Antisocial, Delinquent, and Criminal Behaviors. https://doi.org/10.5772/intechopen.1001823
Hoyle, M. E., Chamberlain, A. W., & Wallace, D. (2020). The Effect of Home Foreclosures on Child Maltreatment Rates: A Longitudinal Examination of Neighborhoods in Cleveland, Ohio. *Journal of interpersonal violence*, *37*(5-6). https://doi.org/10.1177/0886260520943725
Hoyniak, C. P., Quiñones-Camacho, L. E., Camacho, M. C., Chin, J., Williams, E. M., Wakschlag, L. S., & Perlman, S. B. (2021). Adversity Is Linked With Decreased Parent-Child Behavioral and Neural Synchrony. *Developmental Cognitive Neuroscience*, *48*, 100937. https://doi.org/10.1016/j.dcn.2021.100937
Indriati, W., Yurista, S. R., Ardani, I. G. A. I., & Setiawati, Y. (2024). Childhood Maltreatment and Toxic Stress: What We Have Learned From the COVID-19 Pandemic Era. *Journal of Korean Academy of Child and Adolescent Psychiatry*, *35*(3), 163-168. https://doi.org/10.5765/jkacap.240003
Kara, B., & Selçuk, B. (2020). Under Poverty and Conflict: Well-Being of Children Living in the East of Turkey. *American Journal of Orthopsychiatry*, *90*(2), 246-258. https://doi.org/10.1037/ort0000426
Katz, C., Filho, S. R. P., Korbin, J. E., Bérubé, A., Fouché, A., Kaawa–Mafigiri, D., Maguire-Jack, K., Muñoz, P., Spilsbury, J. C., Tarabulsy, G. M., Tiwari, A., Levine, D., Truter, E., & Varela, N. (2021). Child Maltreatment in the Time of the COVID-19 Pandemic: A Proposed Global Framework on Research, Policy and Practice. *Child abuse & neglect*, *116*, 104824. https://doi.org/10.1016/j.chiabu.2020.104824
Klika, J. B., Maguire-Jack, K., Feely, M., Schneider, W., Pace, G. T., Rostad, W. L., Murphy, C. A., & Merrick, M. T. (2022). Childcare Subsidy Enrollment Income Generosity and Child Maltreatment. *Children*, *10*(1), 64. https://doi.org/10.3390/children10010064
Liu, Y., Snyder-Fickler, E., Evans, K. E., & Gifford, E. J. (2025). Child Maltreatment Re-Report, Substantiation, and Foster Care Placement: A Latent Class Analysis of Child, Caregiver, and Household Risk Factors Across Screened-in and Screened-Out Cases. *Child maltreatment*, *31*(3), 552-567. https://doi.org/10.1177/10775595251361150
Maguire-Jack, K., & Sattler, K. M. P. (2022). Neighborhood Poverty, Family Economic Well-Being, and Child Maltreatment. *Journal of interpersonal violence*, *38*(5-6), 4814-4831. https://doi.org/10.1177/08862605221119522
Maguire-Jack, K., Yoon, S., Chang, Y., & Hong, S. (2022). The Relative Influence of Family and Neighborhood Factors on Child Maltreatment at Critical Stages of Child Development. *Children*, *9*(2), 163. https://doi.org/10.3390/children9020163
Maher, E. J., Ferraro, A. C., & Wang, D. (2025). Subgroups of Risk and Adversity for Families Investigated for Child Maltreatment by Poverty Status. *Sociological Perspectives*, *68*(1-3), 61-81. https://doi.org/10.1177/07311214251321447
McDowell, H., Barriault, S., Afifi, T. O., Romano, E., & Racine, N. (2024). Child Maltreatment During the COVID-19 Pandemic: Implications for Child and Adolescent Mental Health. *Frontiers in Child and Adolescent Psychiatry*, *3*. https://doi.org/10.3389/frcha.2024.1415497
Mucci, G., Collins, E., Pearce, E., Avina, M., Hao, S., Onungwa, C., Bunac, J., Hunte, Y., Coopersmith, L., & Yewell, N. (2024). The Lost Years: An Integrative Review of the Mental Health, Educational, and Social Impact of the Pandemic on Children and Adolescents From 2019 to 2022. *Journal of Pediatric Neuropsychology*, *10*(1), 49-90. https://doi.org/10.1007/s40817-024-00160-0
Parenteau, A. M., Boyer, C., Campos, L., Carranza, A. F., Deer, L. K., Hartman, D. T., Bidwell, J. T., & Hostinar, C. E. (2022). A Review of Mental Health Disparities During COVID-19: Evidence, Mechanisms, and Policy Recommendations for Promoting Societal Resilience. *Development and Psychopathology*, *35*(4), 1821-1842. https://doi.org/10.1017/s0954579422000499
Rokach, A., & Chan, S. (2021). The Effect of COVID on Child Maltreatment: A Review. *Journal of Clinical Research and Reports*, *9*(3), 01-13. https://doi.org/10.31579/2690-1919/201
Rosenthal, A., Dimitrieva, J., & DePrince, A. P. (2022). Development of the Checklist of Service Needs and Barriers (C-Snab) Measure for Caregivers Involved in Child Maltreatment Investigations. *Journal of Child and Family Studies*, *31*(12), 3405-3424. https://doi.org/10.1007/s10826-022-02462-3
Tener, D., Marmor, A., Katz, C., Newman, A., Silovsky, J. F., Shields, J. D., & Taylor, E. (2021). How Does COVID-19 Impact Intrafamilial Child Sexual Abuse? Comparison Analysis of Reports by Practitioners in Israel and the US. *Child abuse & neglect*, *116*, 104779. https://doi.org/10.1016/j.chiabu.2020.104779
Thor, P., Yang, S., & Park, Y. (2022). Child Maltreatment in Asian American and Pacific Islander Families: The Roles of Economic Hardship and Parental Aggravation. *International Journal on Child Maltreatment Research Policy and Practice*, *5*(2), 295-310. https://doi.org/10.1007/s42448-021-00111-8
Willoughby, K. A., Arbeau, K., Ketelaars, T., & Pierre, J. S. (2024). An Examination of Underlying Domains in Childhood Adversity: A Scoping Review of Studies Conducting Factor Analyses on Adverse Childhood Experiences. *International journal of environmental research and public health*, *21*(11), 1441. https://doi.org/10.3390/ijerph21111441
Wong, R. S., Tung, K. T. S., Rao, N., Chan, K. L., Fu, K. W., Yam, J. C., Tso, W. W. Y., Wong, W. H. S., Lum, T. Y. S., Wong, I. C. K., & Ip, P. (2022). Using Latent Class Analyses to Examine Health Disparities Among Young Children in Socially Disadvantaged Families During the COVID-19 Pandemic. *International journal of environmental research and public health*, *19*(13), 7893. https://doi.org/10.3390/ijerph19137893
Xu, Y., Bright, C. L., Barth, R. P., & Ahn, H. (2020). Poverty and Economic Pressure, Financial Assistance, and Children's Behavioral Health in Kinship Care. *Child maltreatment*, *26*(1), 28-39. https://doi.org/10.1177/1077559520926568
Younas, F., & Gutman, L. M. (2022). Parental Risk and Protective Factors in Child Maltreatment: A Systematic Review of the Evidence. *Trauma Violence & Abuse*, *24*(5), 3697-3714. https://doi.org/10.1177/15248380221134634
Zhang, L., Liu, Y., & Jonson-Reid, M. (2024). Early Childhood Income Instability and Mental Health in Adolescence: Parenting Stress and Child Maltreatment as Mediators. *Child maltreatment*, *30*(1), 55-67. https://doi.org/10.1177/10775595241236389

JPRFC
E-ISSN: 3041-8550