

# 基于大数据的家庭暴力再犯预测模型构建和干预研究

宋昱萱 陈慕优 宗云丹 范琳婕 嵇文杰

江苏警官学院, 中国·江苏 南京 210031

**摘要:** 在社会转型与法治建设深化背景下, 家庭暴力再犯呈现复杂化、隐蔽化特征, 当前家暴防控面临风险评估经验化、数据壁垒突出、干预措施同质化、本土化适配不足等现实困境。本文以分级干预理论为支撑, 结合大数据技术发展背景, 论证了构建家暴再犯预测模型的理论与实践可行性, 选取 XGBoost 作为核心算法, 通过两轮问卷调查获取实证数据, 识别家暴者核心特征与再犯成因, 搭建“多源数据融合-特征工程-算法训练-风险分级”的模型构建逻辑, 构建显性与隐性结合的双维度特征体系, 设定低、中、高三级风险阈值并匹配差异化干预策略。在此基础上, 从模型与基层警务融合、跨部门数据共享协同、社会支持长效治理三个维度提出闭环干预路径, 推动家暴治理从“经验驱动”向“数据驱动”、从“事后处置”向“事前预警+事中精准干预”转型, 为提升家暴再犯防控的科学化、精准化水平, 推进警务治理现代化与法治社会建设提供技术方案与实践思路。

**关键词:** 大数据; 家庭暴力; 再犯预测; XGBoost 算法; 分级干预

## Construction and Intervention Research on Domestic Violence Recidivism Prediction Model Based on Big Data

Song Yuxuan, Chen Muyou, Zong Yundan, Fan Linjie, Ji Wenjie

Jiangsu Police Institute, China Jiangsu Nanjing 210031

**Abstract:** Against the backdrop of social transformation and deepening law-based governance, domestic violence recidivism has become increasingly complex and concealed. Current domestic violence prevention and control are confronted with practical dilemmas such as empirical risk assessment, prominent data barriers, homogeneous intervention measures, and inadequate local adaptation. Supported by the hierarchical intervention theory and in light of the development of big data technology, this paper demonstrates the theoretical and practical feasibility of constructing a domestic violence recidivism prediction model. With XGBoost as the core algorithm, empirical data are obtained through two rounds of questionnaire surveys to identify the core characteristics of perpetrators and the causes of recidivism. A model construction logic of "multi-source data fusion - feature engineering - algorithm training - risk classification" is established, and a dual-dimensional feature system combining explicit and implicit indicators is developed. Three-level risk thresholds (low, medium, and high) are set and matched with differentiated intervention strategies. On this basis, a closed-loop intervention path is proposed from three dimensions: integration of the model with grassroots policing, cross-departmental data sharing and collaboration, and long-term governance with social support. This research promotes the transformation of domestic violence governance from "experience-driven" to "data-driven", and from "post-incident disposal" to "pre-incident early warning plus accurate in-incident intervention", providing technical solutions and practical ideas for enhancing the scientific and precise prevention and control of domestic violence recidivism, as well as advancing the modernization of policing governance and the construction of a law-based society.

**Keywords:** Big data; Domestic violence; Recidivism prediction; XGBoost algorithm; Hierarchical intervention

## 0 引言

在我国当前社会转型与法治建设不断深化的背景下, 家庭暴力及家暴再犯的发生趋势呈现出复杂化、隐蔽化态势。数据显示, 在婚姻生活中女性遭受过配偶身体暴力和精神暴力的比例为 8.6%, 比 2010 年下降了 5.2 个百分点<sup>[1]</sup>。2016 年以来, 各级人民法院共签发人身安全保护

令 1.5 万余份, 签发率由 2016 年的 52% 提升至 2022 年的 77.6%, 人身安全保护令的适用越来越广泛, 逐渐成为家庭暴力受害人可用、会用、有用的自卫武器。2022 年, 全国法律援助机构共办理涉及虐待、遗弃、家庭暴力的法律援助案件 8900 余件, 为妇女儿童提供法律咨询超过 206 万人次。2016 年至 2022 年, 妇联组织通过 12338 妇女维权

公益服务热线等渠道受理家庭暴力投诉 25.2 万余件次<sup>[2]</sup>。

“坚持男女平等基本国策，保障妇女儿童合法权益”被连续写入党的十八大、十九大、二十大报告<sup>[3]</sup>。《中华人民共和国反家庭暴力法》作为顶层设计的核心法律，确立了告诫书制度、人身安全保护令、强制报告义务等关键机制。同时，《中国妇女发展纲要（2021—2030年）》和《“十四五”平安中国建设规划》进一步将预防和制止家庭暴力纳入国家社会治理体系，推动建立多部门联动干预机制、家暴危险评估系统和施暴者矫治项目试点，国家“十五五”规划明确提出保障妇女儿童权益、完善社会治理，为遏制家暴及再犯趋势提供了坚实的政策支持与制度引导。

因此，本文通过对国内家庭暴力现状进行梳理与分析，深入剖析其原因，发掘目前的干预措施存在的问题，构建基于大数据的家庭暴力再犯预测模型，为未来的家庭暴力干预、警务治理现代化与法治社会建设提供思路。

## 1 本模型构建与干预研究的理论支持与可行性分析

### 1.1 从分级干预理论映射本项目的模型实践

分级干预（Tiered Intervention）源自犯罪学中的风险需求响应模型（Risk-Need-Responsivity Model, RNR），强调根据犯罪风险等级和个体需求匹配差异化的干预措施，以实现矫正资源的最优配置<sup>[4]</sup>。在家庭暴力防治领域，分级干预的核心在于通过科学、量化的风险评估，将施暴者划分为不同再犯风险等级，并针对性地实施警告教育、日常监督、专业心理矫治、人身安全保护令执行等分层响应策略，既避免对低风险个案的过度干预造成公共资源浪费，也防止对高风险个案的干预不足引发恶性事件，实现“精准识别、精准施策、精准防控”，从而避免“一刀切”式执法的低效与资源浪费。本项目将分级干预理念贯穿于“风险识别—等级划分—措施匹配”的全流程，形成了“低—中—高”三级风险对应的干预策略体系。

### 1.2 以大数据为基础的技术支撑与实践验证

#### 1.2.1 大数据赋能警务治理的技术成熟度

当前，我国公安执法正经历从“经验驱动”向“数据驱动”的深刻转型。学者赵子雄指出，“大数据、人工智能等多技术领域的跨界融合正在以前所未见的力量重塑社会治理格局，数据、算法、模型已成为公安执法的核心生产要素”<sup>[5]</sup>。从技术维度来看，机器学习算法能够从海量多源数据中自动提取非线性关联，不但成功突破传统统计模型对线性假设的限制，而且也显著提升预测精度<sup>[6]</sup>。从技

术应用深度来看，犯罪预测研究已从宏观趋势预测迈向微观个体行为预测。孙菲菲等指出，通过对少数人群的重点关注预测其是否再犯罪，可达到预防、减少犯罪的目的<sup>[7]</sup>。而家庭暴力作为一种具有显著反复性、升级性的特殊违法犯罪行为，其案件数据、施暴者行为数据具有明显的可挖掘性，为大数据技术在该领域的应用提供了良好的场景基础，也让依托大数据构建家暴再犯预测模型具备了成熟的技术条件。

#### 1.2.2 国内公安院校与科研团队的先行探索

大数据在犯罪预测中的应用已在国内学术界与实务界取得初步成效，有力验证了本项目的可行性。例如，中国人民公安大学研究团队基于《以基层治理共同体干预家庭暴力研究——以公安机关为视角》的调研数据，构建分析模型对“多次报警但未立案”的家庭暴力案件进行再犯风险分析，成功识别出“施暴者前科次数”“双方分居状态”等显著预测因子，为基层派出所开展重点人员管控提供了量化依据<sup>[8]</sup>。该研究表明，即使采用相对简洁的统计模型，也能从警务记录中挖掘出有效的风险信号，印证了公安数据用于再犯预测的潜力。

另一典型案例来自相关研究团队，他们参考《社会性别主流化视野下警察干预家庭暴力的思考》的研究框架，尝试融合多源数据与传统报案记录，构建多维度评估体系。研究发现，结合多元数据特征后，模型对高风险个案的识别召回率显著提升，有效弥补了传统数据中“隐形受害者”信息缺失的短板<sup>[9]</sup>。这一探索展示了大数据技术突破单一数据源局限、补充传统评估维度的能力，也为本项目未来接入更多社会数据（如社区网格员随访记录、妇联求助热线）提供了方法论参考。

上述实践充分证明，依托公安院校的技术积累与跨学科研究经验，结合现有警务大数据资源，构建面向实战的家暴再犯预测模型不仅是可行的，更是提升家暴防治精准化水平的必然路径。本项目将在前人研究基础上，引入更先进的 XGBoost 算法，优化模型的特征工程与算法训练流程，力求实现家暴再犯防控从“经验研判”向“数据驱动”的实质性跨越，推动家暴治理工作的科学化、精准化发展。

#### 1.2.3 XGBoost 算法及其在再犯预测领域的适用性

XGBoost 是一种优化的分布式梯度提升库，由 Tianqi Chen 等人开发，基于梯度提升决策树框架实现<sup>[10]</sup>。该算法通过引入正则化项防止过拟合，利用二阶导数信息加速收敛，支持并行计算和稀疏数据处理。XGBoost 在再犯预测领域的适用性已得到多项研究验证。韩国学者

WonHa-ram 等提出基于 XGBoost 的再犯预测模型, 在 13,002 条样本数据集上准确率达 69.4%, 显著优于逻辑回归、决策树、支持向量机等传统方法<sup>[11]</sup>。发表于《Journal of Quantitative Criminology》的研究显示, 在暴力再犯预测任务中, XGBoost 的 AUC 值达 0.677, 与随机森林相当, 优于逻辑回归<sup>[12]</sup>。家庭暴力再犯预测具有样本特征复杂、数据非线性关联显著、高风险个案识别要求高的特点, 而 XGBoost 算法的技术优势与再犯预测领域的实践表现, 与家暴再犯预测的需求高度契合。本项目正是在这一技术背景与研究基础上, 选择 XGBoost 作为家暴再犯预测模型的核心算法, 结合家暴治理的本土数据特征与实战需求进行算法优化与模型构建, 以期破解家庭暴力再犯防控中的“预测难、识别难、施策难”问题贡献科学、可行的方案。

## 2 该模型构建与干预研究的核心前提

通过调研, 我们发现在风险评估层面, 我国尚未建立科学的量化评估体系, 基层工作中仍高度依赖民警的主观经验开展家暴再犯风险判断, 缺乏统一、客观的评估标准, 导致出现“高风险个案漏判、低风险过度干预”的问题, 资源配置失衡, 难以实现精准防控。

同时在数据层面存在壁垒, 家暴相关数据来源仍以警方接处警记录为主, 数据维度单一, 对精神控制、经济虐待等非显性暴力行为缺乏有效的数据支撑与识别能力, 直接影响家暴再犯风险预测的准确性。

这也导致在基层民警实践中, 家暴再犯防控的执行难点突出, 一方面在面对非显性暴力时, 民警因缺乏明确统一的判定标准, 难以全面、真实掌握家暴案件的实际情况, 存在明显的识别难题; 另一方面因无专业量化工具支撑, 经验式的风险判断易受主观因素影响, 评估的科学性与准确性不足。

此外在防控模式适配性层面, 现有防控模型与我国“家庭本位”的文化背景契合度不足, 未能充分将婚姻关系修复、社区调解这类符合本土社会需求的干预手段与技术化的风险预测结果有机融合, 使得防控措施本土化适配性与实际落地效果大打折扣。同时民警在开展干预工作时, 存在“重事后处理、轻事前预防”问题, 预测模型尚未有效嵌入公安基层工作流程, 技术层面的风险研判与一线的警务实战应用严重脱节, 使得防控工作的落地效果大打折扣。与司法、民政、心理服务等部门的协同机制仍需优化, 仅靠民警单打独斗必然无法有效解决问题。

由此可见, 基于大数据的家庭暴力再犯干预模型的搭建和干预研究刻不容缓, 势在必行。

## 3 家暴再犯干预的综合实践与应用路径

### 3.1 本家暴再犯预测模型的构建过程

基于上述问题, 结合理论基础与前人经验, 我们以“数据收集-干预适配-机制优化”为核心逻辑, 构建闭环实现路径:

首先, 我们通过两次制作并发放调查问卷, 敲定家暴者核心特征, 弄清再犯成因, 搭建模型框架。从第一次的调研数据来看, 家暴施暴者年龄集中在家庭责任与社会压力叠加阶段, 文化与道德素养的不足易引发其对暴力行为的认知偏差。经济层面则有超五成施暴者存在长期失业、债务缠身或收入不稳定的状况, 可见经济压力是一个重要的行为诱因。同时数据显示, 施暴者的暴力行为呈现出显著的反复性与升级性特征, 首次家暴造成轻微及中度身体损伤的占比达 61.34%, 且 65.13% 的施暴者存在再次施暴行为, 2-5 次暴力发生频次占比 71.42%。根据上述关键词, 我们制作了初期模型, 主要选取了“暴力频次、经济压力、文化水平”等显性特征作为模型变量, 进行小样本数据 XGBoost 算法初训。

通过第二份调查问卷, 我们深入了解家暴者的再犯成因, 根据数据我们意识到再犯成因既包括施暴者自身的性格缺陷、对暴力行为存在认知偏差、酗酒等不良生活习惯等个体层面问题, 也受原生家庭背景、心理状态等内在因素影响。在外部层面, 家庭经济状况不佳、婚姻家庭矛盾突出、缺乏稳定的社会支持等问题加剧了暴力行为的发生, 再结合当前对家暴行为的干预存在措施同质化、针对性不足, 法律惩戒与心理矫治的约束效果有限等问题, 无法有效切断暴力循环, 最终导致家暴再犯率居高不下, 多次施暴且无有效干预的情况下, 形成暴力行为的恶性循环。

然后, 在摸清具体成因后我们继续完善模型的详细构成, 破解“评估粗放”与“适配不足”的问题。基于小样本数据 XGBoost 算法初训基础, 扩大样本量, 优化算法参数, 提升模型对非显性暴力特征的识别能力。在原有“暴力频次、经济压力、文化水平”等显性特征基础上, 新增隐性特征变量: 在施暴者特征中增加情绪稳定性、成瘾行为史两个角度; 在家庭特征上补充婚姻满意度、矛盾冲突两点隐性变量; 在社会特征上聚焦社区干预记录、司法惩戒力度和邻里反馈情况, 构建“显性+隐性”双维度特征体系。同时融入“家庭本位”文化需求, 在模型特征权重中增加“婚姻关系修复意愿”“社区调解成效”等本土化指标, 使预测结果更贴合我国社会实际。

最终, 模型通过 XGBoost 算法对多源数据进行特征

挖掘,识别出与再犯风险强关联的关键因子。基于238份问卷数据的分析显示,“暴力发生频次 $\geq 2$ 次”“存在经济压力”“文化水平初中及以下”等因素与再犯概率呈显著正相关,为风险分级提供了实证依据。模型设定了三级风险阈值:低风险(预测再犯概率 $< 20\%$ )、中风险( $20\% - 50\%$ )、高风险( $> 50\%$ )。在干预措施匹配层面,模型根据风险等级自动生成差异化的干预指引:低风险由社区主导季度回访;中风险由公安+妇联协同开展每月跟进与心理矫治;高风险由司法+公安+社区联动实施保护令申请、重点管控和24小时应急响应。这一设计将预测结果直接转化为可操作的干预指引,有效衔接了技术研判与警务实战。

### 3.2 该模型的综合实践与应用路径

为推动模型成果向可操作干预策略转化,构建科学化、精准化、系统化的治理体系。我们主张紧扣最高人民法院、最高人民检察院及公安部反家暴最新政策导向,从技术赋能、机制创新与社会治理三维度构建闭环干预体系。

#### 3.2.1 深化预测模型与基层警务融合

为提升预测模型的实战应用效能,必须将其有效转化为基层民警在日常勤务中可理解、可操作、可执行的标准化行动指南。针对基层警务工作强度高、处警时效性强等特点,建议开发关于家暴风险智能评估的应用程序,并将其深度集成于现有的移动警务终端系统。该应用程序将依据模型快速采集核心指标,将数据实时传输至后台服务器,调用已训练完成的再犯风险预测模型进行运算,迅速反馈案件风险等级(低、中、高)及对应的标准化处置建议清单。同时为各级公安机关指挥中心定制可视化看板,基于地理信息系统实时呈现家暴警情热力分布与再犯风险等级构成,动态追踪各派出所及网格单元对中、高风险案件的干预任务执行进度与反馈情况,实现对辖区家暴风险态势的动态监控与精准调度。

#### 3.2.2 完善跨部门数据共享与协同机制

打破数据壁垒、强化协同是提升治理效能的关键。一要强化数据共享,建设轻量化的协同信息平台。由地方政府牵头,依托政务外网构建集数据采集、风险预警、任务派发、反馈监督于一体的信息化平台。平台建设遵循数据标准化、流程闭环化、监督透明化原则,统一公安、妇联、法院等部门的数据接口,实施内容脱敏与权限管理;实现预警—派单—执行—反馈全流程线上留痕;自动生成工作量与成效报表,以支撑绩效考核。二要签订协同合作备忘录。由公安牵头,联合法院、检察院、司法、妇联、社区等部门建立联席机

制,明确各方在分级干预中的职责分工,如公安负责处置与管控、妇联开展心理矫治、社区落实巡查与响应。规定信息传递时限,如红色预警须在30分钟内推送至相关单位,以确保应急响应及时启动。建立问责机制,对因履职不力导致的恶性事件进行依法追责。

#### 3.2.3 构建社会支持与长效治理路径

以技术与机制为基础,社会支持与文化引导为延伸,共筑长效治理体系。一是要建立受暴者“庇护—赋能”一体化支持系统。民政、司法、人社协同,建立临时庇护快速响应机制,社区专员可凭电子凭证安排高风险受害者入住协议机构。司法部门开通法律援助绿色通道,提供诉讼代理服务。人社部门提供职业技能培训与就业推荐,增强受害者经济独立性,降低依附风险。二是要探索施暴者强制矫治与司法裁量的联动机制。对中高风险施暴者实施强制心理矫治(如8次/疗程),由妇联或专业机构执行,并建立《干预效果跟踪档案》。档案内容作为检察逮捕、起诉及法院审理保护令、离婚案时评估人身危险性的重要依据。对拒不配合或矫治无效者,依法提请治安处罚或强化管控,形成司法震慑。三是要营造社区“零容忍”的文化氛围。将反家暴纳入社区治理,强化网格员识别矛盾、初期调解与引导求助的能力。培育“反家暴大使”队伍,发挥社区领袖、退休人员等群体的积极作用,参与普法宣传与邻里调解。通过持续动员,构建邻里互助、共同抵制家暴的社区防线,实现源头预防。

综上所述,我们认为该路径设计层层递进,既立足大数据技术的精准赋能,又贴合我国家暴治理的本土实践与基层执法需求,实现了预测模型从技术构建到实践应用的转化,也推动家暴再犯防控从单一部门处置向多主体协同、从事后被动应对向事前精准预警+事中科学干预的现代化转型,为构建科学化、精准化、系统化的家暴治理体系提供了可落地、可操作的实践方案。

## 4 结语

由于缺乏足够的社会实践经验与实地考察,本研究多搜寻相关研究与政府报告作为分析基础,存在一定的理论化。在大数据与社会治理深度融合的大环境下,我国反家暴工作中风险识别和干预难题亟待解决。基于大数据的家庭暴力再犯预测模型的构建,为实现精准化、科学化的社会治理提供了新的技术路径,提升了对家庭暴力再犯风险的识别能力,也为早期干预和资源精准投放奠定了科学基础。模型的应用有助于司法、民政、公安等多部门协同联动,实现从“事后处置”向“事前预警、事中干预”的转

变,推动家庭暴力防治体系的现代化转型。然而,技术的应用必须始终以伦理为底线。在数据采集与使用过程中,必须严格保护个人隐私,确保预测结果服务于帮扶而非惩罚。未来,应进一步优化模型,增强公众信任,并结合多元手段形成综合干预机制。

#### 参考文献:

- [1]《第四期中国妇女社会地位调查主要数据情况》,2022.1.15.
- [2]中国人大网《我国反家庭暴力法实施成效显著-7年签发人身安全保护令1.5万余份》,2023.8.30.
- [3]《新时代中国推进妇女全面发展的实践与成就》白皮书,2025.9.19.
- [4]高雅.公安机关应对家庭暴力之分级干预模式研究[J].公安研究,2023(11):47-55.
- [5]赵子雄.数据导侦:大数据时代侦查工作的实然困境、必然选择与应然指向[J].河北公安警察职业学院学报,2022(3).
- [6]李国军.论大数据驱动下的预测警务创新[J].中国人民公安大学学报(社会科学版),2015(6):3-8.
- [7]孙菲菲,曹卓,肖晓雷.基于随机森林的分类器在犯罪预测中的应用研究[J].情报杂志,2014,33(10):148-152.
- [8]徐梓童.以基层治理共同体干预家庭暴力研究——以公安机关为视角[J].铁道警察学院学报,2022(5):55-60.
- [9]王燕,张晋,潘飞等.社会性别主流化视野下警察干预家庭暴力的思考[J].河南司法警官职业学院学报,2019(1):47-50.
- [10]Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System[C]. ACM SIGKDD, 2016: 785-794.
- [11]Won Ha-ram, et al. A Recidivism Prediction Model Based on XGBoost Considering Asymmetric Error Costs[J]. 国民大学校, 2019.
- [12]In Pursuit of Interpretable, Fair and Accurate Machine Learning for Criminal Recidivism Prediction[J]. Journal of Quantitative Criminology, 2022.

基金项目:江苏警官学院2025年省级大学生创新创业训练计划项目“基于大数据的家庭暴力再犯预测模型构建和干预研究”(项目编号:ZA202510329110)资助。