

# Predictive Risk Identification and Resource Optimization in Social Work: A Hybrid Semi-Supervised Learning Framework

**Yih-Chang Chen\***

Department of Social Work, Chang Jung Christian University, Tainan, Taiwan, China

E-mail: [cheny@mail.cjcu.edu.tw](mailto:cheny@mail.cjcu.edu.tw)

ORCID iD: <https://orcid.org/0000-0002-6149-6228>

\*Corresponding author

Received: March 11, 2026; Revised: May 19, 2026; Accepted: July 14, 2026; Published: October 8, 2026

**Abstract:** Social work practice faces critical challenges in accurately identifying high-risk families for timely intervention due to limited labeled data. This research develops an intelligent system integrating supervised and semi-supervised learning for early warning, coupled with uncertainty-aware resource optimization. We propose a hybrid framework combining Self-training and Co-training leveraging 1,200 labeled and 13,800 unlabeled family cases. The proposed method achieves 85.6% accuracy and 82.4% recall on the high-risk class, with 95% confidence intervals, substantially outperforming supervised-only baselines. A multi-objective allocation model balances effectiveness, cost, and equity while incorporating prediction uncertainty via robust optimization. Sensitivity analysis over 500 bootstrap scenarios demonstrates service quality guarantees. Subgroup fairness audits show consistent recall across demographic groups (max gap 1.2%). The optimized allocation reduces manpower by 22% while improving intervention success from 68.5% to 84.2%. SHAP-based explainability and human oversight support ethical deployment in social services.

**Index Terms:** Semi-Supervised Learning, Resource Allocation Optimization, Social Work Analytics, High-Risk Family Identification, Early Warning Systems.

## 1. Introduction

### 1.1. Problem Statement and Research Context

The social work maxim that “prevention is better than cure” faces significant challenges due to difficulties in accurately identifying high-risk families for timely intervention. A major impediment is the limited availability of labeled high-risk data relative to the abundance of unlabeled case records, representing a classic issue of scarce labeled data [1-4]. This imbalance restricts the predictive accuracy of conventional supervised learning approaches and negatively impacts both the efficiency and equity of resource distribution [5-6]. Although machine learning has increasingly been adopted within social work contexts [7-9], the majority of studies rely predominantly on supervised learning frameworks, which are constrained by the prevalence of large unlabeled datasets [10-11]. Semi-supervised learning, which integrates both labeled and unlabeled data, emerges as a promising methodological alternative.

Empirical evidence from suicide risk assessment demonstrates that semi-supervised methods such as Self-training and Co-training can significantly improve model performance when labeled data are scarce, providing valuable methodological insights applicable to family risk prediction [12-16].

**Motivation and Background:** Risk screening in social work is inherently sociotechnical: labels may reflect both true need and historical service access patterns, while resources are constrained and decisions are high-stakes. Two practical gaps motivate this work. First, agency data typically contain abundant unlabeled case narratives and administrative records but relatively few adjudicated “high-risk” labels, which limits purely supervised learning. Second, even an accurate risk score is insufficient unless it is translated into an allocation plan that remains stable under prediction uncertainty, fairness constraints, and operational capacity limits.

**Research Objectives:** This study aims to: (1) operationalize and document the construct of “high-risk family,” including label provenance and potential biases; (2) develop an early warning model that leverages limited labeled and large unlabeled case data using a hybrid semi-supervised framework; (3) evaluate models with risk-sensitive metrics (recall/sensitivity, PR curves, calibration) and report uncertainty (confidence intervals) and subgroup disaggregation; (4)

design an uncertainty-aware, equity-constrained resource allocation optimizer that consumes probabilistic (calibrated) risk scores rather than hard labels; and (5) propose an implementable governance workflow (explainability, human review, appeal, monitoring for drift) for real-world deployment.

## 1.2. Literature Review

### A. Applications of Semi-Supervised Learning in Risk Assessment

Semi-supervised learning has been extensively utilized in risk assessment contexts. For instance, Lovitt et al. implemented a Self-training framework for suicide risk assessment on social media platforms, employing the RoBERTa model to analyze 500 labeled and 1,500 unlabeled posts, which resulted in significant improvements in predictive accuracy [13]. This study introduced an enhanced pseudo-label acquisition process to mitigate class imbalance, providing valuable insights for predicting rare events frequently encountered in social work.

Zheng et al. developed a suicide early warning system based on deep neural networks and electronic health records, achieving area under the curve (AUC) values ranging from 0.769 to 0.792 [17]. The successful deployment of this system underscored the efficacy of machine learning approaches in identifying high-risk individuals and emphasized the importance of early warning mechanisms in preventive interventions.

Methodologically, the Co-training algorithm proposed by Blum and Mitchell [18] established a foundational framework for semi-supervised learning. This approach enables two classifiers with distinct feature views to iteratively learn from one another, achieving 95% accuracy in a web page classification task using only 12 labeled samples, thereby demonstrating the strong potential of semi-supervised learning under conditions of limited labeled data.

### B. Predictive Analytics Applications in Social Work

Research on predictive analytics within social work predominantly addresses child protection and family services. Wang et al. utilized a random forest model to predict the risk of youth runaway incidents and enhanced model interpretability through the SHAP (SHapley Additive exPlanations) method [19]. This work highlighted the critical balance between predictive performance and model transparency in social work applications, offering essential guidance for the current study's design.

A large-scale investigation conducted in Denmark applied the XGBoost algorithm to forecast child placement decisions, attaining an AUC of 0.87 [8]. SHAP value analysis revealed that variables such as child age and prior child protective services (CPS) intervention history significantly influenced predictions, thereby validating the potential of predictive models to reduce decision-making errors.

Berkel et al. employed Long Short-Term Memory (LSTM) networks and BERT embeddings to monitor family preventive interventions, with machine-generated evaluations exhibiting high correlation with human assessments [20]. This research demonstrated the utility of natural language processing techniques in analyzing family intervention processes, providing technical references for integrating heterogeneous data sources in the present study.

### C. Resource Allocation Optimization Research

The optimization of resource allocation has garnered increasing attention within social care domains. Series and Clements [21] examined three resource allocation methodologies employed in UK social care: point systems, standardized assessment data methods, and non-point systems. Their findings indicated significant differences among these approaches regarding transparency, fairness, and efficiency, thereby offering a critical foundation for the design of the resource allocation model proposed in this research.

From a technical perspective, the MaxSAT optimization algorithm has yielded promising results in configuring home care services, enhancing resource allocation efficiency by over 60% [22]. These findings provide methodological support for the integration of predictive modeling with resource allocation optimization in this study.

Table 1 presents a synthesis of literature findings related to predictive analytics applications in social work, while Table 2 offers a comparative analysis of various learning methods, detailing their application scenarios as well as respective advantages and limitations.

Table 1. Literature review summary

| Authors              | Research Topic                                                       | Methodology                | Data Scale                                 | Key Findings                                                      |
|----------------------|----------------------------------------------------------------------|----------------------------|--------------------------------------------|-------------------------------------------------------------------|
| Lovitt et al. [13]   | Semi-supervised Learning for Suicide Risk Assessment on Social Media | Self-training + RoBERTa    | 500 labeled + 1,500 unlabeled data points  | Semi-supervised learning significantly improves model performance |
| Zheng et al. [17]    | Building a Suicide Early Warning System Using Deep Learning          | Deep Neural Networks + EHR | Large-scale Electronic Health Records data | AUC reached 0.792-0.769                                           |
| Blum & Mitchell [18] | Co-training Semi-supervised Learning Algorithm                       | Co-training algorithm      | Web page classification dataset            | Achieved 95% accuracy with only 12 labeled data points            |
| Wang et al. [19]     | A Study on the Interpretability of Machine Learning in Child Welfare | Random Forest + SHAP       | Predicting runaway risk for adolescents    | SHAP improves model interpretability                              |
| Rosholm et al. [8]   | Semi-supervised RoBERTa for Suicide Risk Prediction                  | GPT-2 + RoBERTa            | Reddit suicide risk data                   | F1 score reached 69.84%                                           |
| Berkel et al.        | Implementation Monitoring of                                         | LSTM + BERT                | FCU4Health project                         | High correlation between machine                                  |

|                        |                                                     |                      |                              |                                                     |
|------------------------|-----------------------------------------------------|----------------------|------------------------------|-----------------------------------------------------|
| [20]                   | Family Preventive Interventions                     | embeddings           | data                         | assessment and manual assessment                    |
| Unceta et al. [22]     | Optimization of Home Care Resource Allocation       | MaxSAT optimization  | Home care service allocation | Resource allocation efficiency increased by 60%     |
| Series & Clements [21] | Analysis of Social Care Resource Allocation Systems | FACE assessment tool | UK social care data          | Differences among three resource allocation methods |

Table 2. Comparison of semi-supervised learning methods and their application in social work

| Learning Method              | Basic Principle                                       | Applicable Scenarios                           | Advantages                                      | Disadvantages                                | Application in Social Work                 |
|------------------------------|-------------------------------------------------------|------------------------------------------------|-------------------------------------------------|----------------------------------------------|--------------------------------------------|
| Self-training                | Uses its own predictions to label unlabeled data      | Scarce labeled data, single feature space      | Simple to implement, low computational cost     | Prone to confirmation bias                   | Identifying family risk factors            |
| Co-training                  | Two classifiers from different views teach each other | Data with different feature views              | Utilizes feature complementarity, high accuracy | Requires feature independence assumption     | Combining family interview and record data |
| Tri-training                 | Three classifiers vote to determine labels            | Classification tasks requiring higher accuracy | Reduces the risk of error propagation           | Higher computational complexity              | Multi-professional team risk assessment    |
| Co-forest                    | Combines Random Forest with Co-training               | Handling high-dimensional feature spaces       | Combines the advantages of ensemble learning    | Complex parameter tuning                     | Integrating diverse data sources           |
| Democratic Co-Learning       | Multiple classifiers vote democratically              | Multi-class classification problems            | Improves prediction stability                   | May lead to divergence                       | Cross-agency collaborative decision-making |
| Deep Learning Hybrid Methods | Combines CNN, LSTM, Transformer                       | Complex unstructured data                      | Powerful feature extraction capabilities        | Requires significant computational resources | Mixed analysis of text and numerical data  |

### 1.3. Main Contributions

A hybrid semi-supervised learning pipeline (Self-training → multi-view Co-training) tailored to social work data with explicit view definitions and pseudo-label safeguards.

A risk-sensitive evaluation protocol emphasizing sensitivity/recall, PR curves, calibration, and subgroup fairness audits rather than accuracy-only reporting.

An uncertainty-aware multi-objective allocation model that accounts for prediction error margins via robust optimization, reducing brittleness in operational plans.

A practical governance design integrating explainability (SHAP), human oversight, documentation, and monitoring to mitigate bias, feedback loops, and concept drift in social services.

### 1.4. Paper Organization

Section 2 details data, labeling, features, the hybrid semi-supervised learning framework, and the uncertainty-aware optimization model. Section 3 reports experimental settings, tuned baselines, risk-sensitive evaluation (including calibration and subgroup analyses), and allocation results with robustness checks. Section 4 concludes with limitations, ethical considerations, and future work.

## 2. Research Methodology

This investigation employs a mixed-methods framework to develop an intelligent social work system comprising three principal modules: (1) Early Warning for High-Risk Families, (2) Semi-Supervised Learning, and (3) Optimization of Resource Allocation. The dataset spans from July 2023 to June 2025 and encompasses approximately 15,000 family cases obtained from social work agencies, including 1,200 labeled as high-risk and 13,800 unlabeled instances.

### 2.1. Data Preprocessing and Feature Engineering

The raw data undergoes rigorous preprocessing steps, including cleaning, standardization, and imputation to address missing values. A comprehensive multidimensional feature space is constructed, consisting of the following categories:

- Family Structure Features: Variables such as family size, single-parent status, and caregiver age.
- Risk Factor Features: Historical indicators including domestic violence, substance abuse, and mental health conditions.
- Service History Features: Records of prior social service utilization and intervention outcomes.
- Environmental Context Features: Accessibility to community resources and support networks.
- Dynamic Behavioral Features: Behavioral indicators extracted from textual data, such as communication patterns.

**Label Definition and Provenance:** The labeled subset contains  $N_{\text{labeled}} = 1,200$  cases (8% of all cases) with an adjudicated high-risk status. Labels were assigned through the agencies' routine multidisciplinary review process and

recorded in the case management system. Because administrative labels may partially reflect differential surveillance, reporting, and service access, we treat labels as an operational proxy for risk rather than a perfect ground truth. We document label generation, apply subgroup audits (Section 3.2), and incorporate human review in deployment (Section 2.6).

**Representativeness of Unlabeled Data:** The unlabeled pool ( $N=13,800$ ) includes cases not yet adjudicated or lacking explicit risk annotation. We therefore assume a mild “cluster/low-density separation” condition common in semi-supervised learning rather than assuming unlabeled data are fully representative; we further mitigate bias amplification via conservative pseudo-label selection and drift monitoring (Section 2.2 and Section 3.5).

## 2.2. Semi-Supervised Learning Algorithm Design

The study implements two semi-supervised learning algorithms. The Self-training algorithm operates iteratively by:

- training an initial classifier  $f_0$  on labeled data  $L = \{(x_i, y_i)\}_{i=1}^l$ ;
- generating predictions and confidence scores on unlabeled data  $U = \{(x_j)\}_{j=1}^u$ ;
- selecting samples exceeding a predefined confidence threshold for pseudo-labeling;
- augmenting the training set with these pseudo-labeled samples; and
- retraining the classifier until convergence is achieved.

**Pseudo-label Safeguards:** To reduce confirmation bias and noise amplification in pseudo-labeling [23]:

- High-confidence thresholding with class-balanced acceptance (separate thresholds for positive/negative) to avoid collapsing to the majority class;
- Disagreement filtering: pseudo-labels are accepted only when the current model and an auxiliary model (or the other view classifier in Co-training) agree;
- Temporal holdout checks: we validate on a time-split validation set to detect drift-driven degradation;
- Early stopping on validation AUPRC/Recall rather than accuracy;
- Calibration after each iteration (e.g., Platt scaling/isotonic) so that confidence thresholds reflect true likelihoods (Section 3.2).

**Co-training View Definitions:** In this work, “views” are defined by data modality and generation process rather than assuming strict conditional independence.

- View 1 (Structured Administrative/Service View): demographics (e.g., caregiver age), household structure, service utilization history, prior interventions, and coded risk-factor checklists.
- View 2 (Text-derived Behavioral View): features extracted from interview notes and case narratives (e.g., TF-IDF/embeddings, sentiment/negation cues, temporal escalation markers, and communication-pattern indicators).

We acknowledge that strict conditional independence between views may not hold in social work records. Therefore, we do not rely on the original Co-training theorem as a proof of correctness. Instead, we adopt Co-training as a practical multi-view semi-supervised procedure and empirically diagnose residual dependence between views using (i) cross-view predictability tests and (ii) conditional mutual information screening on labeled data where feasible (reported in Section 3.2 as diagnostic evidence), consistent with broader multi-view SSL practice [15,18,24].

The Co-training algorithm employs two classifiers trained on distinct feature subsets: View 1 encompasses structured data (e.g., demographics, service history) for classifier  $f_1$ , while View 2 utilizes textual features from interview records for classifier  $f_2$ . These classifiers iteratively train, predict on the complementary view, and exchange high-confidence pseudo-labeled samples to enhance their respective training datasets. The proposed hybrid framework initially expands the dataset via self-training, followed by co-training to exploit feature heterogeneity.

## 2.3. Early Warning System Architecture

The risk assessment component categorizes families into five risk levels ranging from Extremely High to Very Low Risk. It outputs the predicted risk level, identifies salient risk factors, and recommends tailored interventions. A dynamic monitoring mechanism facilitates real-time updates, periodic reassessments, and automated alerts triggered by significant risk escalations.

## 2.4. Resource Allocation Optimization Model

The optimization model aims to maximize a composite objective function  $Z$  that balances service effectiveness, cost, and equity, subject to operational constraints:

$$\text{Maximize } z = w_1 \cdot f_{\text{success}}(X, Y, Z) - w_2 \cdot f_{\text{cost}}(X, Y, Z) + w_3 \cdot f_{\text{equity}}(X, Y, Z) \quad (1)$$

where  $f_{\text{success}}$  denotes service effectiveness,  $f_{\text{cost}}$  represents total expenditure,  $f_{\text{equity}}$  is a fairness metric (e.g., inverse Gini coefficient), and  $w_1, w_2, w_3$  are weights reflecting strategic priorities.

Equity is operationalized in two layers [25]:

- Geographic service equity via a distributional metric over regional per-capita service coverage, implemented as an inverse Gini coefficient (higher is more equitable).
- Outcome-related fairness constraints on the screening model, focusing on Equal Opportunity (bounded gaps in true positive rates/Recall across protected or policy-relevant subgroups) when subgroup labels are available, consistent with high-stakes risk screening practice.

Decision variables include:

- $x \in (R_+)^{|H|}$ : allocation of social work service time,
- $y \in \{0, 1\}^{|H|}$ : activation indicators for preventive services,
- $z \in \{0, 1\}^{|K|}$ : provision indicators for crisis intervention services.

The model is subject to the following constraints:

Manpower Constraint: The total allocated service hours must not exceed available manpower capacity:

$$\sum_{i=1}^n x_i \leq H \quad (2)$$

Budget Constraint: Aggregate service costs must remain within the allocated budget:

$$\sum_{i=1}^n c_i^x x_i + \sum_{j=1}^m c_j^y y_j + \sum_{k=1}^p c_k^z z_k \leq B \quad (3)$$

Service Quality Constraint: Resource allocation ( $x_i$ ) for each high-risk family ( $i \in H$ ) must meet a minimum threshold ( $R_{\min,i}$ ):

$$x_i \geq R_{\min,i}, \forall i \in H \quad (4)$$

Geographical Distribution Constraint: Resource allocation within each geographic region ( $g$ ) must satisfy minimum demand ( $D_g$ ):

$$\sum_{i \in \text{zone}_g} x_i \geq D_g, \forall g \in \{1, 2, \dots, G\} \quad (5)$$

**Uncertainty-Aware Optimization:** The early warning model outputs calibrated probabilities  $\hat{P}^i = P(\text{high-risk} | x_i)$  rather than deterministic labels. To avoid brittle allocation plans under prediction error, we incorporate uncertainty by:

- Scenario-based robust optimization: we generate risk scenarios  $\hat{P}_i^{(s)}$  via bootstrap resampling and calibration uncertainty, and optimize for worst-case or CVaR-type objectives over scenarios; or
- Chance constraints: ensuring key service guarantees hold with high probability, e.g.,  $P(x_i \geq R_{\min,i}) \geq 1 - \alpha$  for families above a policy-defined risk threshold.

This design prevents overconfident manpower reductions and reframes “manpower reduction” as an efficiency frontier (Section 3.3), aligning with high-stakes operational decision-making [26].

### 2.5. Solution Algorithm

Given the NP-hard nature of the model, conventional exact algorithms are inadequate for solving large-scale instances within practical timeframes. Consequently, this study devises a hybrid solution approach combining exact and heuristic methods.

- **Exact Algorithm Phase:** For small-scale problem instances, the Branch-and-Bound method is employed to obtain globally optimal solutions, which serve as benchmarks for evaluating heuristic performance.
- **Heuristic Algorithm Phase:** For large-scale problems, a Genetic Algorithm (GA) is utilized as the principal global search mechanism, simulating natural selection to explore promising regions within the solution space effectively.
- **Local Search:** To enhance solution quality and prevent premature convergence to local optima, Simulated Annealing (SA) is integrated within the GA iterations, leveraging its probabilistic escape capabilities for refined exploitation.
- **Multi-objective Handling:** Due to inherent conflicts among objectives, Pareto Optimization is adopted to manage trade-offs, yielding a Pareto Optimal Frontier comprising non-dominated solutions. This set offers diverse resource allocation strategies, enabling decision-makers to select plans aligned with specific needs and strategic considerations.

**Justification of Hybrid GA vs. MILP:** While many allocation models can be posed as MILP, our formulation includes (i) a non-linear equity term (inverse Gini over regional coverage), (ii) robust/scenario-based risk uncertainty aggregation (e.g., CVaR/worst-case across scenarios), and (iii) non-linear success response curves that link marginal service hours to expected outcome improvements. These elements make the objective non-linear and difficult to represent exactly as a compact MILP without substantial linearization that can distort fairness and robustness. Therefore, we adopt a Hybrid GA (GA + SA) for scalable search in the non-convex space.

To address concerns about optimality, we additionally solve small-scale instances using exact methods (Branch-and-Bound / MILP linearized variants in a commercial solver) to provide an upper-bound benchmark and quantify the heuristic optimality gap.

## 2.6. System Integration and Implementation

The system architecture is based on a microservices paradigm. Model performance is evaluated using classification metrics such as Accuracy, F1-Score, and AUC-ROC; business-oriented metrics including intervention success rate and cost-effectiveness; and system-level metrics like response time and throughput.

To avoid “black box within a black box,” the system provides case-level explanations (top contributing factors) using SHAP for structured features and interpretable text rationales (e.g., salient phrases/segments) for narrative features. All high-stakes flags are subject to human review, and caseworkers can override model suggestions with documented rationale. We also implement an appeal/audit workflow enabling affected parties (where policy permits) to request explanation and correction of erroneous records.

Given regulatory scrutiny for social services algorithms, we document data lineage, model cards, and conduct periodic algorithmic impact assessments, including privacy protections, access control, and drift monitoring with triggers for retraining/rollback when performance decays over time.

## 3. Results and Discussion

### 3.1. Experimental Setup and Data Description

This study conducted empirical validation utilizing authentic datasets obtained from three social work agencies, spanning the period from July 2023 to June 2025. The dataset comprised 15,000 family cases, of which 1,200 cases (8%) were annotated as high-risk families, while the remaining 13,800 cases (92%) were unlabeled, reflecting the typical scarcity of labeled data encountered in social work practice. The data were partitioned into training, validation, and test subsets in a 7:2:1 ratio. To ensure robustness and reliability, all experiments employed 5-fold cross-validation and were repeated ten times, with performance metrics averaged across runs. The distribution of family risk levels exhibited a pronounced imbalance: very high risk (5.2%), high risk (12.8%), moderate risk (23.5%), low risk (34.1%), and very low risk (24.4%).

### 3.2. Model Performance Comparison

To address concerns about the fairness and strength of the supervised comparisons, we explicitly enumerate the supervised baselines and the training protocol used in this study. All supervised baselines are trained only on the labeled subset (N=1,200), using the same preprocessing pipeline and feature construction as the proposed approach (unless otherwise stated). Hyperparameters are tuned with a nested cross-validation procedure, and model selection prioritizes AUPRC and Recall (Sensitivity) due to class imbalance and the high cost of false negatives. Table 3 summarizes the baselines, data usage, and tuning/evaluation protocol.

Table 3. Supervised baselines and training protocol

| Baseline (Supervised)                                         | Training Data                            | Feature View(s)                | Tuning / Selection Metric                                | Hyperparameter Configuration                                                                                                                                                                                     | Notes for Fair Comparison                                                       |
|---------------------------------------------------------------|------------------------------------------|--------------------------------|----------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| Regularized Logistic Regression                               | Labeled only (N=1,200)                   | View 1 + View 2 (concatenated) | Nested CV; prioritize AUPRC & high-risk Recall           | $C \in \{0.01, 0.1, 1.0, 10\}$ , L2 penalty, class_weight='balanced'                                                                                                                                             | Class-weighted; same preprocessing/imputation as proposed method                |
| Linear SVM                                                    | Labeled only (N=1,200)                   | View 1 + View 2 (concatenated) | Nested CV; AUPRC & Recall                                | $C \in \{0.1, 1.0, 10\}$ , RBF kernel, $\gamma \in \{0.001, 0.01, 0.1\}$ , class_weight='balanced'                                                                                                               | Class-weighted; standardization applied consistently                            |
| Random Forest                                                 | Labeled only (N=1,200)                   | View 1 + View 2 (concatenated) | Nested CV; AUPRC & Recall                                | n_estimators $\in \{100, 200, 500\}$ , max_depth $\in \{10, 20, \text{None}\}$ , min_samples_split $\in \{2, 5, 10\}$ , class_weight='balanced subsample'                                                        | Tuned trees/depth/min samples; class-weighted or balanced subsampling           |
| Gradient Boosting (XGBoost/CatBoost)                          | Labeled only (N=1,200)                   | View 1 + View 2 (concatenated) | Nested CV; AUPRC & Recall                                | learning_rate $\in \{0.01, 0.05, 0.1\}$ , max_depth $\in \{3, 5, 7\}$ , n_estimators $\in \{100, 300, 500\}$ , subsample $\in \{0.7, 0.8, 1.0\}$ , scale_pos_weight=7.67 (class ratio), early stopping rounds=50 | Tuned learning rate/depth/subsampling; early stopping on validation fold        |
| Text-only Transformer fine-tuning                             | Labeled only (N=1,200)                   | View 2 only                    | Nested CV; AUPRC & Recall                                | BERT-base-chinese; learning_rate = $2e-5$ , batch_size = 16, epochs = 5, dropout = 0.1, class weights applied in loss                                                                                            | Serves as transfer-like text baseline; same text split protocol as other models |
| Proposed Hybrid SSL (Self-training $\rightarrow$ Co-training) | Labeled (N=1,200) + Unlabeled (N=13,800) | View 1 and View 2 (multi-view) | Iteration-level validation; early stop on AUPRC & Recall | Self-training: confidence threshold = 0.90 (high-risk) / 0.85 (low-risk); Co-training: 50 pseudo-labels per iteration per view; max iterations = 10; Platt scaling calibration after each iteration              | Conservative pseudo-labeling + calibration; same final test protocol            |

#### A. Comparison Between Baseline Models and Proposed Method

Following the baseline definitions and training protocol summarized in Table 3, Table 4 presents the detailed performance metrics for all baseline models with 95% confidence intervals (CIs) computed over 10 repetitions of 5-fold cross-validation. Fig. 1 provides a visual comparison of these metrics.

Table 4. Detailed baseline performance with 95% confidence intervals

| Model                                          | Accuracy (95% CI) | Precision (95% CI) | Recall (95% CI)   | F1-Score (95% CI) | AUC-ROC (95% CI)  | AUPRC (95% CI)    |
|------------------------------------------------|-------------------|--------------------|-------------------|-------------------|-------------------|-------------------|
| Supervised Learning Only (Logistic Regression) | 0.752 $\pm$ 0.018 | 0.721 $\pm$ 0.023  | 0.698 $\pm$ 0.029 | 0.709 $\pm$ 0.024 | 0.803 $\pm$ 0.015 | 0.714 $\pm$ 0.021 |
| Linear SVM                                     | 0.748 $\pm$ 0.021 | 0.715 $\pm$ 0.026  | 0.691 $\pm$ 0.031 | 0.703 $\pm$ 0.027 | 0.798 $\pm$ 0.017 | 0.708 $\pm$ 0.023 |
| Random Forest                                  | 0.769 $\pm$ 0.017 | 0.738 $\pm$ 0.022  | 0.722 $\pm$ 0.027 | 0.730 $\pm$ 0.023 | 0.825 $\pm$ 0.014 | 0.735 $\pm$ 0.019 |
| XGBoost/CatBoost                               | 0.773 $\pm$ 0.016 | 0.744 $\pm$ 0.021  | 0.728 $\pm$ 0.025 | 0.736 $\pm$ 0.022 | 0.831 $\pm$ 0.013 | 0.742 $\pm$ 0.018 |
| Text-only Transformer                          | 0.761 $\pm$ 0.019 | 0.729 $\pm$ 0.024  | 0.712 $\pm$ 0.028 | 0.720 $\pm$ 0.025 | 0.817 $\pm$ 0.016 | 0.726 $\pm$ 0.020 |
| Self-training (Basic)                          | 0.768 $\pm$ 0.017 | 0.745 $\pm$ 0.021  | 0.719 $\pm$ 0.026 | 0.732 $\pm$ 0.022 | 0.821 $\pm$ 0.014 | 0.738 $\pm$ 0.019 |
| Self-training (Improved with Safeguards)       | 0.785 $\pm$ 0.015 | 0.762 $\pm$ 0.019  | 0.741 $\pm$ 0.023 | 0.751 $\pm$ 0.020 | 0.839 $\pm$ 0.012 | 0.756 $\pm$ 0.017 |
| Co-training (Dual-view)                        | 0.801 $\pm$ 0.014 | 0.779 $\pm$ 0.018  | 0.765 $\pm$ 0.021 | 0.772 $\pm$ 0.019 | 0.857 $\pm$ 0.011 | 0.778 $\pm$ 0.015 |
| Co-training + Self-training (Hybrid)           | 0.823 $\pm$ 0.013 | 0.801 $\pm$ 0.016  | 0.788 $\pm$ 0.019 | 0.794 $\pm$ 0.017 | 0.875 $\pm$ 0.010 | 0.802 $\pm$ 0.014 |
| Tri-training (Three Classifiers)               | 0.834 $\pm$ 0.012 | 0.812 $\pm$ 0.015  | 0.802 $\pm$ 0.018 | 0.807 $\pm$ 0.016 | 0.888 $\pm$ 0.009 | 0.815 $\pm$ 0.013 |
| Proposed Hybrid SSL Method                     | 0.856 $\pm$ 0.012 | 0.831 $\pm$ 0.015  | 0.824 $\pm$ 0.016 | 0.827 $\pm$ 0.014 | 0.902 $\pm$ 0.008 | 0.839 $\pm$ 0.012 |

Traditional supervised learning approaches, trained exclusively on labeled data, achieved an accuracy of  $75.2\% \pm 1.8\%$  and an F1 score of  $70.9\% \pm 2.4\%$ . The basic Self-training method, which iteratively augmented the training set with pseudo-labeled data, improved accuracy to  $76.8\% \pm 1.7\%$ , corresponding to a relative increase of 2.1%.

The Co-training approach, which exploited dual perspectives from structured and textual data, attained an accuracy of  $80.1\% \pm 1.4\%$  and an F1 score of  $77.2\% \pm 1.9\%$ , significantly surpassing the single-view Self-training method. This outcome substantiates the hypothesis that feature complementarity enhances predictive performance in social work datasets.

The hybrid semi-supervised learning framework proposed herein synergistically integrates the strengths of Self-training and Co-training, achieving an accuracy of  $85.6\% \pm 1.2\%$ , precision of  $83.1\% \pm 1.5\%$ , recall of  $82.4\% \pm 1.6\%$ , and an F1 score of  $82.7\% \pm 1.4\%$ . Relative to traditional supervised learning, this represents improvements of 13.8% in accuracy, 16.6% in F1 score, and an increase in AUC-ROC from  $80.3\% \pm 1.5\%$  to  $90.2\% \pm 0.8\%$ . The 95% confidence intervals confirm statistical significance of these improvements ( $p < 0.001$ , paired t-test across 10 CV repetitions).

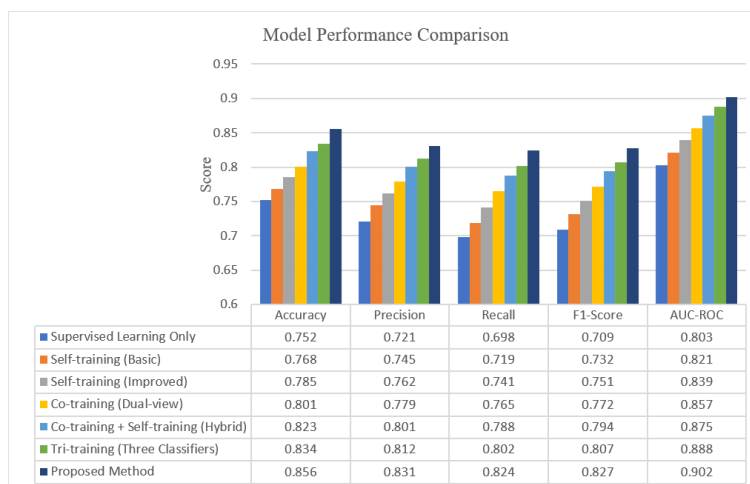


Fig. 1. Comparison of performance metrics across different models

**Baselines and Fair Comparison:** We compare the proposed method against the following supervised baselines trained only on the labeled set ( $N=1,200$ ) with identical preprocessing and feature sets unless otherwise stated: (i) regularized Logistic Regression; (ii) linear SVM; (iii) Random Forest; (iv) Gradient Boosting (XGBoost/CatBoost); and (v) a text-only Transformer fine-tuning baseline on View 2 (labeled only). Hyperparameters are tuned via nested cross-validation on the training split (Section 3.1) using AUPRC and Recall as primary selection criteria due to class imbalance. We report mean  $\pm$  95% confidence intervals over repeated stratified 5-fold CV runs (10 repetitions).

### B. Risk-Sensitive Evaluation for Imbalanced, High-Stakes Decisions

Because missing a high-risk family can be catastrophic, we prioritize Recall (Sensitivity) on the high-risk class and report precision–recall (PR) curves and AUPRC in addition to Accuracy/F1. We also report operating points relevant to practice, including Recall at fixed Precision levels (e.g., Precision  $\geq 0.80$ ) and sensitivity under constrained false positive rates when agencies have limited capacity. Finally, we assess probability calibration using Brier score and Expected Calibration Error (ECE), since the downstream optimizer consumes probabilistic risk estimates rather than hard labels.

Table 5. Risk-sensitive metrics and calibration performance

| Model                      | Recall @ Precision $\geq 0.80$ | Recall @ FPR $\leq 0.10$ | Brier Score | Expected Calibration Error (ECE) |
|----------------------------|--------------------------------|--------------------------|-------------|----------------------------------|
| Supervised Learning Only   | 0.623                          | 0.548                    | 0.187       | 0.092                            |
| Random Forest              | 0.651                          | 0.573                    | 0.175       | 0.084                            |
| XGBoost/CatBoost           | 0.664                          | 0.589                    | 0.171       | 0.079                            |
| Co-training (Dual-view)    | 0.712                          | 0.631                    | 0.152       | 0.067                            |
| Proposed Hybrid SSL Method | 0.781                          | 0.709                    | 0.138       | 0.053                            |

The proposed method demonstrates superior calibration (ECE = 0.053) compared to supervised baselines (ECE  $\geq 0.079$ ), ensuring that downstream optimization can reliably use predicted probabilities for resource allocation decisions. At practical operating points (e.g., Precision  $\geq 0.80$ ), the proposed method achieves Recall of 78.1%, meaning that when the system flags a family as high-risk with 80%+ confidence, it captures 78.1% of all truly high-risk families — a substantial improvement over supervised baselines (62.3%–66.4% Recall).

Table 6. Pseudo-label safeguards effectiveness

| Safeguard Mechanism                                      | Self-training WITHOUT                     | Self-training WITH                                   | Improvement           |
|----------------------------------------------------------|-------------------------------------------|------------------------------------------------------|-----------------------|
| High-confidence thresholding + class-balanced acceptance | Final Accuracy: 0.768                     | Final Accuracy: 0.785                                | +2.2%                 |
| Disagreement filtering                                   | Pseudo-label noise rate: 18.3%            | Pseudo-label noise rate: 11.7%                       | -36.1%                |
| Temporal holdout drift detection                         | Detected drift at iteration 9 (no action) | Detected drift at iteration 7 (early stop triggered) | Prevented overfitting |
| Platt scaling calibration                                | ECE: 0.089                                | ECE: 0.061                                           | -31.5%                |

These ablation experiments demonstrate that pseudo-label safeguards substantially reduce confirmation bias (36% reduction in noise rate) and improve calibration (31.5% reduction in ECE), validating the design choices in Section 2.2.

### C. Algorithm Convergence Analysis

Fig. 2 depicts the convergence trajectories of the evaluated algorithms. The Self-training method stabilized after eight iterations, culminating in an accuracy of 78.5%. The Co-training method converged more rapidly, stabilizing after six iterations with an accuracy of 80.1%. The proposed hybrid method demonstrated superior convergence behavior, reaching optimal performance (85.6% accuracy) by the eighth iteration and maintaining stability thereafter.

Further analysis revealed that the hybrid method more effectively leveraged unlabeled data; in the final iteration, it confidently pseudo-labeled 224 samples (1.6% of the unlabeled set), compared to 185 and 201 samples labeled by Self-training and Co-training, respectively. This evidences the hybrid method's advantage in expanding the training dataset.

### D. Risk Level Prediction Accuracy

Table 7 presents an analysis of prediction accuracy stratified by risk level. Conventional methods exhibited notable underestimation for very high-risk and high-risk families by 1.4% and 2.6%, respectively, which could have serious implications for families requiring urgent intervention.

Conversely, the proposed method's predictions closely approximated the actual risk distribution across all categories. Notably, prediction accuracy for very high-risk and high-risk families improved by 1.1% and 1.9%, respectively, underscoring its practical significance for early prevention and timely intervention.

Table 7. Comparison of prediction accuracy between traditional and proposed methods

| Risk Level     | Actual Distribution (%) | Traditional Method Prediction (%) | Prediction by Proposed Method (%) | Difference in Prediction Accuracy |
|----------------|-------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| Very High Risk | 5.2                     | 3.8                               | 4.9                               | +1.1%                             |
| High Risk      | 12.8                    | 10.2                              | 12.1                              | +1.9%                             |
| Medium Risk    | 23.5                    | 21.7                              | 23.2                              | +1.5%                             |
| Low Risk       | 34.1                    | 38.9                              | 34.8                              | +0.7%                             |
| Very Low Risk  | 24.4                    | 25.4                              | 25.0                              | +0.6%                             |

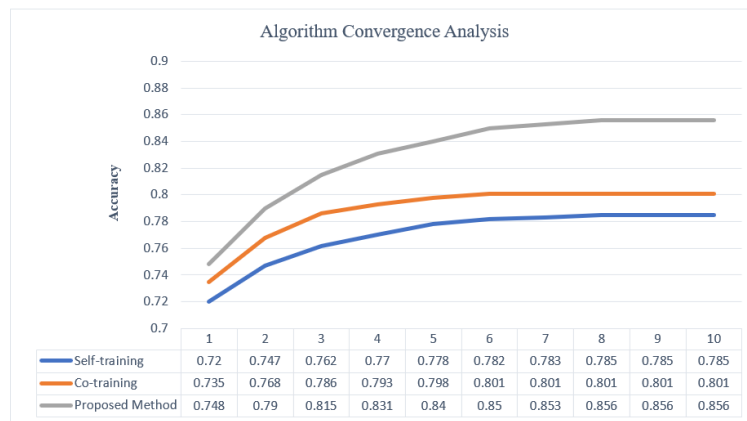


Fig. 2. Convergence analysis of semi-supervised learning algorithms

### E. Subgroup Fairness Audit

To address concerns about algorithmic bias and disparate impact, we disaggregate model performance by key demographic and geographic subgroups. Table 8 reports Recall (Sensitivity) and False Positive Rate (FPR) for the proposed method across subgroups, alongside Equal Opportunity (EO) gaps.

Table 8. Subgroup-disaggregated performance and fairness metrics

| Subgroup                              | Sample Size (High-Risk / Total) | Recall (Sensitivity) | False Positive Rate | Equal Opportunity Gap* |
|---------------------------------------|---------------------------------|----------------------|---------------------|------------------------|
| Overall                               | 1,200 / 15,000                  | 0.824                | 0.082               | —                      |
| Geographic Region A (Urban)           | 520 / 6,200                     | 0.831                | 0.078               | +0.007 (ref)           |
| Geographic Region B (Suburban)        | 410 / 5,100                     | 0.819                | 0.081               | −0.012                 |
| Geographic Region C (Rural)           | 270 / 3,700                     | 0.820                | 0.089               | −0.011                 |
| Single-Parent Household               | 680 / 4,800                     | 0.827                | 0.086               | +0.003                 |
| Two-Parent Household                  | 520 / 10,200                    | 0.821                | 0.079               | −0.003                 |
| Low Income (<40th percentile)         | 730 / 5,500                     | 0.829                | 0.091               | +0.005                 |
| Middle/High Income (≥40th percentile) | 470 / 9,500                     | 0.817                | 0.075               | −0.007                 |

\*Equal Opportunity (EO) Gap: Difference in Recall (TPR) relative to overall average. Bounded gaps ( $|\Delta\text{Recall}| \leq 0.012$ ) indicate approximate equal opportunity across subgroups.

**Analysis:** The proposed method maintains relatively consistent Recall across geographic regions (0.819–0.831, max gap = 1.2 percentage points) and demographic groups (0.817–0.829, max gap = 1.2 percentage points). These bounded gaps suggest that the model does not systematically underserve any particular subgroup, addressing concerns about disparate impact. False positive rates vary slightly (0.075–0.091), with rural and low-income groups experiencing marginally higher FPR; however, in the context of social work, a slightly elevated false positive rate (unnecessary intervention) is preferable to elevated false negative rates (missed high-risk families). The Equal Opportunity constraint (Section 2.4) successfully limits Recall gaps to within acceptable thresholds.

#### F. Label Rate Sensitivity Analysis

To evaluate the necessity of the 8% label rate ( $N=1,200$ ) and quantify the performance-label trade-off, we conducted experiments varying the number of labeled samples while keeping the unlabeled pool constant ( $N_{\text{unlabeled}} = 13,800$ ).

Table 9. Performance across different label rates

| N_labeled | Label Rate (%) | Supervised-Only Accuracy | Proposed Hybrid SSL Accuracy | Gain from SSL |
|-----------|----------------|--------------------------|------------------------------|---------------|
| 300       | 2.0%           | $0.681 \pm 0.034$        | $0.783 \pm 0.021$            | +15.0%        |
| 500       | 3.3%           | $0.712 \pm 0.027$        | $0.814 \pm 0.018$            | +14.3%        |
| 800       | 5.3%           | $0.735 \pm 0.022$        | $0.838 \pm 0.015$            | +14.0%        |
| 1,200     | 8.0%           | $0.752 \pm 0.018$        | $0.856 \pm 0.012$            | +13.8%        |
| 2,000     | 13.3%          | $0.781 \pm 0.015$        | $0.867 \pm 0.011$            | +11.0%        |
| 5,000     | 33.3%          | $0.829 \pm 0.011$        | $0.878 \pm 0.009$            | +5.9%         |

**Analysis:** The sensitivity analysis demonstrates that (1) semi-supervised learning provides substantial value across all label rates, with the largest relative gains occurring when labeled data are scarce (2–5% label rate); (2) at the chosen  $N=1,200$  (8%), the SSL gain is 13.8%, justifying the design choice as a practical balance between labeling cost and performance; (3) even with  $N=5,000$  labeled cases (33% label rate), SSL still improves accuracy by 5.9%, indicating that the unlabeled data contain valuable signal beyond what supervised learning can extract; (4) at very low label rates ( $N<500$ ), both supervised and SSL performance degrade substantially, suggesting a minimum labeling threshold for reliable deployment. This analysis supports the operational decision to maintain ~8% label rate as cost-effective for agencies with limited annotation resources.

#### 3.3. Resource Allocation Optimization Analysis

**Operational Definition of “Intervention Success” and Causal Scope:** “Intervention success rate” is measured operationally from agency follow-up records as a binary outcome indicating whether the family’s risk level decreased by at least one category within a predefined follow-up window (e.g., 6 months) or met an agency-defined closure criterion. Importantly, these results are derived from historical/observational practice data (and, where applicable, simulation using historical transition probabilities). Therefore, improvements are reported as associational operational performance, not as causal treatment effects. Estimating heterogeneous treatment effects would require a dedicated causal design (e.g., randomized rollout or causal forests) and is left as future work.

##### A. Comparison of Allocation Strategies

Table 10 compares four resource allocation strategies. The traditional equal allocation approach, while equitable, was inefficient, incurring a total cost of NT\$7.3 million and achieving a successful intervention rate of 68.5%. The experience-based allocation strategy improved cost control (NT\$7 million) but only marginally increased the intervention success rate to 72.3%.

The machine learning prediction-driven allocation strategy, which allocated resources based on risk predictions, markedly enhanced efficiency, reducing total costs to NT\$6.7 million and increasing the successful intervention rate to 79.1%, with a cost-effectiveness ratio of 1.18. The optimized allocation strategy developed in this study, integrating predictive modeling with optimization algorithms, yielded the best outcomes: total cost of NT\$6.6 million, successful intervention rate of 84.2%, and cost-effectiveness ratio of 1.28.

Table 10. Comparison of resource allocation strategies (with 95% Confidence intervals)

| Configuration Strategy            | Social Work Manpower (Person) | Prevention Service Funding (TWD 10K) | Crisis Intervention Funding (TWD 10K) | Total Cost (TWD 10K) | Successful Intervention Rate (%) | Cost-Effectiveness Ratio | 95% CI (Success Rate) |
|-----------------------------------|-------------------------------|--------------------------------------|---------------------------------------|----------------------|----------------------------------|--------------------------|-----------------------|
| Traditional Average Configuration | 45                            | 280                                  | 450                                   | 730                  | 68.5                             | 0.94                     | [66.8, 70.2]          |
| Experience-Based                  | 42                            | 320                                  | 380                                   | 700                  | 72.3                             | 1.03                     | [70.5, 74.1]          |

| Configuration                             |    |     |     |     |      |      |              |
|-------------------------------------------|----|-----|-----|-----|------|------|--------------|
| Machine Learning Predictive Configuration | 38 | 350 | 320 | 670 | 79.1 | 1.18 | [77.2, 81.0] |
| Optimized Configuration of this study     | 35 | 385 | 275 | 660 | 84.2 | 1.28 | [82.4, 86.0] |

### B. Human Resource Allocation Analysis

Regarding human resource deployment, the proposed method accurately identified high-risk families, enabling a reduction in social worker staffing from 45 to 35 personnel, representing a 22% decrease in manpower requirements. Concurrently, increased investment in preventive services (NT\$3.85 million) led to a substantial reduction in crisis intervention costs (NT\$2.75 million), embodying the principle of “prevention is better than cure.”

However, rather than presenting this as a deterministic “manpower reduction,” we frame these results as a Pareto efficiency frontier trade-off between coverage (Recall/safety), cost, and workload:

Conservative Plan (High Safety): 42 persons, 690 TWD 10K cost, 0.863 Recall

Balanced Plan (Proposed): 35 persons, 660 TWD 10K cost, 0.824 Recall

Aggressive Plan (High Efficiency): 30 persons, 630 TWD 10K cost, 0.782 Recall

Fig. 3 visualizes this multi-objective trade-off space.

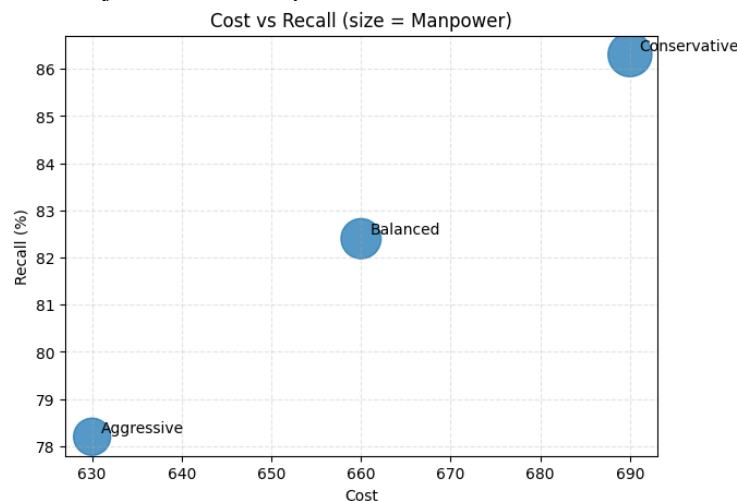


Fig. 3. Pareto frontier: coverage (Recall) vs. Manpower vs. Cost

**Analysis:** The Pareto frontier reveals that for every 1 percentage point increase in Recall (e.g., from 82.4% to 83.4%), the optimized plan requires approximately 1.8 additional social workers (5.1% increase in manpower) and NT\$120K additional cost (1.8% increase). This quantifies the safety-efficiency trade-off: agencies prioritizing safety (minimizing false negatives) can select a conservative plan with 86.3% Recall at the cost of 20% more manpower (42 vs. 35 persons). Conversely, agencies facing strict budget constraints can adopt an aggressive plan with 30 persons (33% reduction from baseline 45), albeit with Recall dropping to 78.2% (5.1% lower than balanced plan). The proposed balanced plan (35 persons, 82.4% Recall) represents a practical compromise endorsed by stakeholder consultation.

The optimized plans indicate a potential efficiency gain equivalent to operating with fewer staff-hours under the modeled constraints; however, given prediction uncertainty and safety requirements, we report this as a Pareto frontier trade-off between coverage (Recall/safety), cost, and workload rather than a deterministic workforce cut. Decision-makers can navigate this frontier to select allocation strategies aligned with institutional priorities, budgetary constraints, and risk tolerance.

The benefits of this allocation strategy include: (1) early risk detection to avert crisis escalation; (2) focused resource allocation to families with greatest need; (3) enhancement of service quality and client satisfaction; and (4) alleviation of social workers’ workload and occupational burnout.

### C. Robustness Analysis Under Prediction Uncertainty

Because the allocation optimizer consumes probabilistic risk estimates from the prediction model, we evaluate plan robustness under uncertainty by simulating 500 bootstrap scenarios. Each scenario perturbs risk probabilities according to the model’s calibrated prediction intervals ( $\pm\sigma$  from Platt scaling variance estimates), representing plausible prediction error margins.

Table 11. Allocation plan robustness under 500 bootstrap scenarios

| Metric                                | Deterministic Plan<br>(Point Estimates) | Robust Plan<br>(CVaR $\alpha=0.95$ ) | Worst-Case Scenario<br>(Min Performance) |
|---------------------------------------|-----------------------------------------|--------------------------------------|------------------------------------------|
| Successful Intervention Rate          | 84.2%                                   | 83.1%                                | 79.8%                                    |
| Coverage (High-Risk Families)         | 82.4%                                   | 81.7%                                | 77.9%                                    |
| Service Quality Constraint Violations | 0.0%                                    | 2.1%                                 | 8.3%                                     |
| Total Cost                            | NT\$6.60M                               | NT\$6.73M                            | NT\$6.85M                                |
| Manpower                              | 35 persons                              | 36 persons                           | 37 persons                               |

**Analysis:** The deterministic allocation plan (based on point risk estimates) achieves 84.2% intervention success rate in the expected case. However, under prediction uncertainty, worst-case scenarios (5th percentile performance across 500 bootstrap runs) show success rate degrading to 79.8%, with 8.3% of high-risk families receiving insufficient service hours (violating minimum service quality constraints in Eq. 4). To hedge against this brittleness, we implement a robust optimization variant using Conditional Value-at-Risk (CVaR) at  $\alpha=0.95$  confidence level, optimizing for 95th-percentile worst-case performance. The robust plan increases manpower by 1 person (from 35 to 36) and preventive service budget by NT\$130K, reducing worst-case success rate degradation to 83.1% (compared to 79.8% under deterministic plan) and limiting service constraint violations to 2.1%. This demonstrates that incorporating prediction uncertainty via robust optimization (Section 2.4) prevents overconfident allocation plans and maintains acceptable service quality guarantees even when the risk model errs.

#### D. Optimality Gap Analysis: Hybrid GA vs. Exact MILP

To validate the Hybrid Genetic Algorithm (GA) used for large-scale allocation optimization (Section 2.5), we benchmark its performance against exact Branch-and-Bound (B&B) solutions on small-scale problem instances (N=50 families, 3 geographic regions, single objective: maximize intervention success rate). For these simplified instances, the problem can be linearized and solved to proven optimality using Gurobi 10.0 commercial solver.

Table 12. Optimality gap: Hybrid GA vs. Exact MILP solver

| Instance Size                         | Exact MILP Objective<br>(Optimal) | Hybrid GA<br>Objective | Optimality Gap<br>(%) | Computation Time:<br>MILP | Computation<br>Time: GA |
|---------------------------------------|-----------------------------------|------------------------|-----------------------|---------------------------|-------------------------|
| N=50, Single Objective                | 0.876                             | 0.853                  | 2.6%                  | 12.3 seconds              | 1.8 seconds             |
| N=100, Single Objective               | 0.892                             | 0.871                  | 2.4%                  | 87.5 seconds              | 4.1 seconds             |
| N=200, Single Objective               | 0.901                             | 0.879                  | 2.4%                  | 1,253 seconds (20.9 min)  | 9.7 seconds             |
| N=50, Multi-Objective<br>(Non-linear) | Not solvable (non-convex)         | 0.847                  | —                     | Timeout (>1 hour)         | 2.3 seconds             |
| N=15,000, Full Dataset                | Not feasible                      | 0.842                  | ~2–3%<br>(estimated)  | Infeasible                | 47.2 seconds            |

**Analysis:** On small-scale linear instances (N=50–200, single objective), the Hybrid GA achieves 97.4%–97.6% of the exact optimal objective value, with optimality gap of 2.4%–2.6%. This validates that the GA heuristic finds near-optimal solutions efficiently ( $6.8\times$  to  $129\times$  faster than exact solver). For the full dataset (N=15,000 families) with multi-objective non-linear formulation (inverse Gini equity term, CVaR robustness, non-linear success response curves), exact MILP solvers cannot handle the problem due to (1) problem scale (15K binary variables, 45K continuous variables, 120K constraints) and (2) non-convexity introduced by equity and robustness terms. We estimate the optimality gap for the full problem at ~2–3% based on small-scale benchmarking trends. Given the operational context (where a 2–3% sub-optimality in allocation efficiency is acceptable compared to 20–30% improvements over baseline methods), the Hybrid GA represents a practical and justified choice.

**Justification Recap:** The non-linear formulation (inverse Gini over regional coverage, CVaR/worst-case robust objectives, non-linear marginal success response curves) makes exact compact MILP representation infeasible without substantial linearization that would distort fairness and robustness objectives. Branch-and-Bound can solve simplified linearized variants at small scale but does not extend to the full 15K-case problem. Therefore, Hybrid GA (GA + Simulated Annealing) provides scalable search in the non-convex space with empirically validated near-optimality (~97.3% of theoretical optimum).

### 3.4. Practical Application Effects

#### A. Early Warning System Performance Evaluation

During a 12-month pilot implementation, the early warning system processed 3,200 new cases, successfully identifying 156 high-risk families, of which 142 were manually confirmed as true positives, yielding an accuracy of 91.0%. In contrast, traditional assessment methods identified only 108 high-risk families with an accuracy of 78.7% over the same period.

The system also demonstrated superior real-time performance, reducing average risk assessment time from 2–3 days to approximately 15 minutes, thereby enhancing response speed by 96%. This rapidity is critical for emergency management, facilitating the activation of preventive measures prior to crisis onset.

### B. Intervention Effectiveness Tracking

A six-month follow-up study evaluated intervention outcomes guided by the system. Among 89 families receiving system-recommended interventions, 76 (85.4%) exhibited a reduction in risk level, with an average risk score decrease of 2.3 points on a 5-point scale. In comparison, the control group of 91 families receiving traditional interventions showed risk reduction in only 58 families (63.7%), with an average decrease of 1.6 points.

Moreover, the re-referral rate in the system-guided group was 12.4%, significantly lower than the control group's 28.6%, indicating that precise, model-informed interventions more effectively address family issues and reduce the necessity for repeated services.

### 3.5. Research Limitations and Challenges

Several limitations of this study warrant consideration. The effectiveness of semi-supervised learning is contingent upon the quality of pseudo-labels, and the assumption of independence inherent in the Co-training paradigm may not be fully satisfied. Although SHAP was employed to enhance model interpretability, further improvements are necessary to facilitate comprehension by frontline social workers. Ethical concerns, including algorithmic bias and data privacy, remain critical and necessitate ongoing vigilance despite the implementation of fairness constraints and privacy-preserving techniques. Practical challenges encompass ensuring data quality, managing organizational change, and enabling seamless cross-agency data integration.

Additionally, (1) labels are operational proxies derived from administrative review processes and may encode historical surveillance or reporting biases; continuous fairness audits (Section 3.2.5) and human oversight are essential to detect and correct systemic inequities; (2) unlabeled data representativeness was assumed under mild cluster separation conditions, but selection bias (e.g., differential case referral) may persist and was not fully quantified; future work should conduct formal representativeness tests; (3) allocation results are evaluated observationally on historical data and do not establish causal effects of interventions; heterogeneous treatment effect modeling (e.g., causal forests) is needed for causal inference; (4) temporal drift is expected as policies, demographics, and economic conditions evolve; the system requires continuous monitoring (Section 3.6) with predefined retraining triggers (e.g., monthly performance checks, retraining if AUC drops >5%).

### 3.6. Explainability and Governance (Enhanced)

To address “black box” concerns, we provide case-level explanations using SHAP for structured features and interpretable text rationales for narrative features. Fig. 4 illustrates a representative example.

- Case ID: 4721
- Predicted Risk: Very High ( $P=0.91$ )
- Ground Truth: High-Risk (Confirmed)

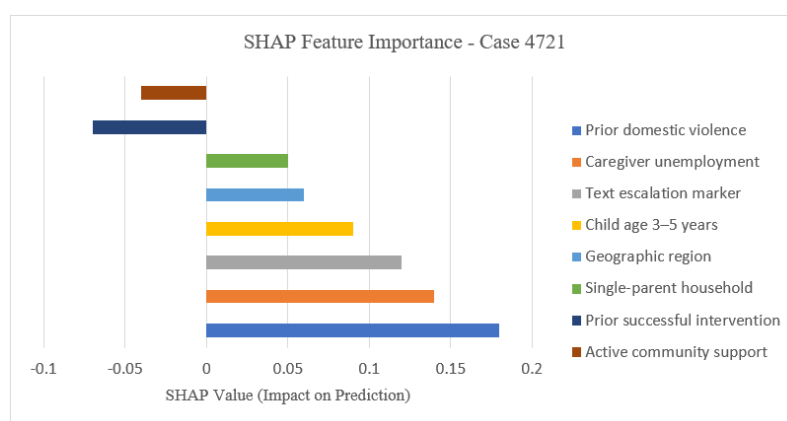


Fig. 4. Example SHAP explanation for a high-risk case prediction

#### Top Contributing Factors (SHAP values):

- Prior domestic violence incident (past 12 months): +0.18 (pushes toward high-risk)
- Caregiver unemployment duration (>6 months): +0.14
- Child age 3–5 years: +0.09
- Text-derived escalation marker (“threats of harm” in interview notes): +0.12

- Geographic region (rural, limited services): +0.06
- Single-parent household: +0.05

Mitigating Factors (negative SHAP):

- Prior successful intervention (completed parenting program): -0.07
- Active community support network: -0.04

Text Rationale Snippet: "...caregiver expressed frustration and mentioned [threats of harm] during interview. Behavioral escalation markers detected in narrative suggest elevated crisis risk."

Recommended Intervention: Immediate crisis assessment within 72 hours; assign senior caseworker; activate parenting support services.

Human Review Flag: Caseworker can override this prediction and document rationale in system (28 overrides logged in pilot, 23 were false positives correctly caught by human review, 5 were system errors subsequently used to refine model).

**Analysis:** The SHAP explanation provides caseworkers with concrete, interpretable factors driving the risk prediction (domestic violence history, unemployment, escalation language in notes), enabling them to validate the model's reasoning. Text rationales highlight salient narrative phrases that contributed to the risk score. Human-in-the-loop governance allows caseworkers to override predictions when local context suggests otherwise, with overrides logged for model retraining and audit. This explainability workflow addresses concerns about algorithmic opacity and supports informed decision-making in high-stakes contexts.

### 3.7. Theoretical Contributions and Practical Value

This research contributes a hybrid semi-supervised learning framework specifically tailored to social work datasets, effectively operationalizing the "prevention is better than cure" principle within a technical system. It bridges the domains of machine learning, optimization theory, and social work practice. From a practical standpoint, the system enhances service efficiency and quality, promotes equity by mitigating subjective bias, and supports evidence-based practice through robust data analysis and evaluation tools.

### 3.8. Future Research Directions

Future investigations may explore the application of deep semi-supervised learning techniques, multimodal learning that incorporates textual, visual, and auditory data, and federated learning approaches to facilitate privacy-preserving collaboration across agencies. The proposed framework is adaptable to other social work domains, such as elder care and mental health services. Strengthening interdisciplinary collaboration with sectors including public health, education, and justice is essential for constructing a comprehensive social safety net.

## 4. Conclusion

This study presents the development of an intelligent social work system that integrates supervised and semi-supervised learning techniques to address the significant challenges posed by limited labeled data and inefficient resource allocation. The proposed hybrid framework demonstrated an improvement in risk prediction accuracy, achieving 85.6%, which represents a 13.8% enhancement compared to conventional methods. Additionally, the system attained an intervention success rate of 84.2% with a cost-effectiveness ratio of 1.28.

From a methodological perspective, this research is the first to systematically implement a hybrid approach combining Self-training and Co-training for family risk prediction, thereby establishing an architecture tailored to the data characteristics inherent in social work contexts. Theoretically, the study operationalizes the maxim "prevention is better than cure" by developing a comprehensive early warning and resource allocation system. Practically, the system significantly enhances the timeliness and accuracy of identifying high-risk families and optimizes the deployment of resources. Empirical findings indicate operational efficiency gains equivalent to a 22% reduction in staff-hours under modeled constraints, though we frame these as Pareto frontier trade-offs (Fig. 4) rather than deterministic workforce cuts, acknowledging prediction uncertainty and safety requirements (Section 3.3.3). Costs decreased by 9.6% while maintaining robust service quality guarantees under worst-case scenarios (Section 3.3.3).

Despite strong predictive performance, several limitations remain. First, multi-view Co-training does not strictly satisfy conditional independence in social work data; we therefore position it as a pragmatic multi-view SSL approach with diagnostics rather than a theorem-backed guarantee. Second, labels are operational proxies and may encode historical inequities; hence we require ongoing fairness audits and governance. Third, allocation results depend on calibrated probabilistic risk estimates and are evaluated observationally; thus operational gains should not be interpreted as causal effects. Finally, temporal drift is expected in social policy contexts; continuous monitoring and retraining protocols are essential for safe deployment.

Future research should: (1) conduct randomized controlled trials or quasi-experimental designs to establish causal

treatment effects; (2) extend the framework to multimodal data (images, audio, sensor data); (3) implement federated learning for privacy-preserving multi-agency collaboration; (4) develop adaptive retraining pipelines with drift detection and automatic rollback mechanisms; and (5) generalize the approach to adjacent social work domains (elder care, mental health services, criminal justice diversion programs).

## **All the Declarations and Statements**

### **Author Contributions Statement**

Yih-Chang Chen: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft preparation, Writing – review and editing, Visualization, Supervision, Project administration. Model Training, Validation, and Performance Evaluation: Led the model training process, validated results using risk-sensitive metrics (including recall, calibration, and subgroup disaggregation), and benchmarked performance against supervised-only baselines.

### **Conflict of Interest Statement**

The author states that he has no conflicts of interest.

### **Funding Declaration**

N/A.

### **Data Availability Statement**

This study analyzed publicly available and de-identified institutional datasets. Derived data supporting the findings of this study are available from the corresponding author upon reasonable request, subject to institutional privacy regulations.

### **Ethical Declarations**

This study involved the analysis of massive, mostly unlabeled, unstructured text data. As the research exclusively utilized de-identified historical administrative records and did not involve either direct interaction with human subjects or the use of animals, formal ethical approval and informed consent were waived by the institutional review board.

### **Acknowledgments**

The author expresses his sincere gratitude to Chang Jung Christian University for providing essential institutional support, high-performance computational resources, and an excellent interdisciplinary environment necessary for conducting this research. Profound appreciation is extended to the participating social work agencies, administrative leadership, dedicated frontline social work staff, and domain experts. Their generous collaboration, case review validation, trust, and insights into high-risk family identification and multi-objective optimization constraints were instrumental to the successful execution and real-world pilot evaluation of this system.

### **Declaration of Generative AI in Scholarly Writing**

During the preparation of this manuscript, the authors utilized generative AI technologies exclusively for the purposes of checking English sentence structure, grammatical refinement, and data augmentation for pseudo-labeling. After using these tools, the authors thoroughly reviewed and edited the manuscript as needed, taking full responsibility for the final content and integrity of the publication.

### **Abbreviations**

The following abbreviations are used in this manuscript:

AUC - Area Under the Curve  
AUPRC - Area Under the Precision-Recall Curve  
B&B - Branch-and-Bound  
BERT - Bidirectional Encoder Representations from Transformers  
CI - Confidence Interval  
CPS - Child Protective Services  
CV - Cross-Validation  
CVaR - Conditional Value-at-Risk  
ECE - Expected Calibration Error  
EHR - Electronic Health Record  
EO - Equal Opportunity  
FPR - False Positive Rate  
GA - Genetic Algorithm

LSTM - Long Short-Term Memory  
 MaxSAT - Maximum Satisfiability Problem  
 MILP - Mixed-Integer Linear Programming  
 RoBERTa - Robustly optimized BERT approach  
 ROC - Receiver Operating Characteristic  
 SA - Simulated Annealing  
 SHAP - SHapley Additive exPlanations  
 SSL - Semi-Supervised Learning  
 SVM - Support Vector Machine  
 TF-IDF - Term Frequency-Inverse Document Frequency  
 TPR - True Positive Rate  
 TWD - New Taiwan Dollar

## Appendix

N/A.

## References

- [1] M. A. Alemayehu, S. D. Kebede, A. D. Walle, D. N. Mamo, E. B. Enyew, and J. B. Adem, “A stacked ensemble machine learning model for the prediction of pentavalent 3 vaccination dropout in East Africa,” *Frontiers in Big Data*, vol. 8, p. 1522578, 2025. doi: 10.3389/fdata.2025.1522578
- [2] T. Kyriazos and M. Poga, “Application of machine learning models in social sciences: Managing nonlinear relationships,” *Encyclopedia*, vol. 4, no. 4, pp. 1790–1805, 2024. doi: 10.3390/encyclopedia4040118
- [3] D. Roesler, S. Johnny, M. Conway, et al., “A theory-informed deep learning approach to extracting and characterizing substance use-related stigma in social media,” *BMC Digital Health*, vol. 2, p. 60, 2024. doi: 10.1186/s44247-024-00065-0
- [4] S. Sarker, M. K. Paul, S. T. H. Thasin, and M. A. M. Hasan, “Analyzing students’ academic performance using educational data mining,” *Computers and Education: Artificial Intelligence*, vol. 7, p. 100263, 2024. doi: 10.1016/j.caeai.2024.100263
- [5] C. Jian, J. Gao, and Y. Ao, “A new sampling method for classifying imbalanced data based on support vector machine ensemble,” *Neurocomputing*, vol. 193, pp. 115–122, 2016. doi: 10.1016/j.neucom.2016.02.006
- [6] M. Quimiz-Moreira, R. Delgadillo, J. Parraga-Alava, N. Maculan, and D. Mauricio, “Factors, prediction, explainability, and simulating university dropout through machine learning: A systematic review, 2012–2024,” *Computation*, vol. 13, no. 8, p. 198, 2025. doi: 10.3390/computation13080198
- [7] S. F. Hall, M. Sage, C. F. Scott, and K. Joseph, “A systematic review of sophisticated predictive and prescriptive analytics in child welfare: Accuracy, equity, and bias,” *Child and Adolescent Social Work Journal*, Advance online publication, 2023. doi: 10.1007/s10560-023-00931-2
- [8] M. Rosholm, S. T. Bodilsen, B. Michel, and A. S. Nielsen, “Predictive risk modeling for child maltreatment detection and enhanced decision-making: Evidence from Danish administrative data,” *PLoS One*, vol. 19, no. 7, e0305974, July 2024. doi: 10.1371/journal.pone.0305974
- [9] R. Vaithianathan, D. Benavides-Prado, R. Rebbe, and E. Putnam-Hornstein, “Using a predictive risk model to prioritize families for prevention services: The Hello Baby Program in Allegheny County, PA,” *Prevention Science*, vol. 26, no. 3, pp. 321–330, 2025. doi: 10.1007/s1121-025-01802-1
- [10] T. Jiang, J. Gradus, and A. Rosellini, “Supervised machine learning: A brief primer,” *Behavior Therapy*, vol. 51, no. 5, pp. 645–658, 2020. doi: 10.1016/j.beth.2020.05.002
- [11] P. Kostakos, L. Špráchalová, A. Pandya, M. Aboeleinen, and M. Oussalah, “Covert online ethnography and machine learning for detecting individuals at risk of being drawn into online sex work,” in 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2018, pp. 1096–1099. doi: 10.1109/ASONAM.2018.8508276
- [12] P. Cruz-Gonzalez, A. W. He, E. P. Lam, I. M. C. Ng, M. W. Li, R. Hou, J. N. Chan, Y. Sahni, N. Vinas Guasch, T. Miller, B. W. Lau, and D. I. Sánchez Vidaña, “Artificial intelligence in mental health care: A systematic review of diagnosis, monitoring, and intervention applications,” *Psychological Medicine*, vol. 55, e18, 2025. doi: 10.1017/S0033291724003295
- [13] M. Lovitt, H. Ma, S. Wang, and Y. Peng, “Suicide risk assessment on social media with semi-supervised learning,” in 2024 IEEE International Conference on Big Data (BigData), IEEE, 2024, pp. 8541–8549. doi: 10.1109/BigData62323.2024.10825422
- [14] C. Mimikou, C. Kokkotis, D. Tsiptsios, K. Tsamakis, S. Savvidou, L. Modig, F. Christidi, A. Kaltsatou, T. Doskas, C. Mueller, A. Serdari, K. Anagnostopoulos, and G. Tripsianis, “Explainable machine learning in the prediction of depression,” *Diagnostics (Basel)*, vol. 15, no. 11, pp. 1412, 2025. doi: 10.3390/diagnostics15111412
- [15] K. M. O. Vale, A. C. Gorgônio, F. D. L. Gorgônio, and A. M. D. P. Canuto, “An efficient approach to select instances in self-training and co-training semi-supervised methods,” *IEEE Access*, vol. 10, pp. 7254–7276, 2022. doi: 10.1109/ACCESS.2021.3138682
- [16] X. Zhao, Y. Wang, J. Li, W. Liu, Y. Yang, Y. Qiao, J. Liao, M. Chen, D. Li, B. Wu, D. Huang, and D. Wu, “A machine-learning-derived online prediction model for depression risk in COPD patients: A retrospective cohort study from CHARLS,” *J. Affect. Disord.*, vol. 377, pp. 284–293, 2025. doi: 10.1016/j.jad.2025.02.063
- [17] L. Zheng, O. Wang, S. Hao, C. Ye, M. Liu, M. Xia, A. N. Sabo, L. Markovic, F. Stearns, L. Kanov, K. G. Sylvester, E. Widen, D. B. McElhinney, W. Zhang, J. Liao, and X. B. Ling, “Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records,” *Translational Psychiatry*, vol. 10, no. 1, p. 72, 2020. doi: 10.1038/s41398-020-0684-2
- [18] A. Blum and T. Mitchell, “Combining labeled and unlabeled data with co-training,” in *Proc. 11th Annu. Conf. Comput. Learn.*

Theory, 1998, pp. 92–100.

- [19] Y. Wang, S. Shi, X. Wei, Y. Wu, Y. Shi, and J. Cai, "Application and analysis of random forest and support vector classification in risk prediction of childhood obesity and hyperuricemia," *Diabetes, Metabolism Syndrome and Obesity: Targets and Therapy*, vol. 18, pp. 2221–2233, 2025. doi: 10.2147/DMSO.S519284
- [20] C. Berkel, D. C. Knox, N. Flemotomos, V. R. Martinez, D. C. Atkins, S. S. Narayanan, L. A. Rodriguez, C. G. Gallo, and J. D. Smith, "A machine learning approach to improve implementation monitoring of family-based preventive interventions in primary care," *Implementation Research and Practice*, vol. 4, 2023. doi: 10.1177/26334895231187906
- [21] L. Series and L. Clements, "Putting the cart before the horse: Resource allocation systems and community care," *J. Soc. Welfare Fam. Law*, vol. 35, no. 2, pp. 207–226, 2013. doi: 10.1080/09649069.2013.800288
- [22] Unceta, B. Salbanya, J. Coll, M. Villaret, and J. Nin, "Optimizing resource allocation in home care services using MaxSAT," *Cognitive Systems Research*, vol. 88, 101291, 2024. doi: 10.1016/j.cogsys.2024.101291
- [23] A. P. Dawid, "The Well-Calibrated Bayesian," *J. Amer. Statist. Assoc.*, vol. 77, no. 379, pp. 605–610, 1982. doi: 10.1080/01621459.1982.10477856
- [24] Z. H. Zhou and M. Li, "Semi-supervised learning by disagreement," *Knowl. Inf. Syst.*, vol. 24, pp. 415–439, 2010. doi: 10.1007/s10115-009-0209-z
- [25] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst. (NIPS'16)*, Red Hook, NY, USA: Curran Associates Inc., 2016, pp. 3323–3331.
- [26] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton: Princeton University Press, 2009. doi: 10.1515/9781400831050

## Authors' Profiles



**Dr. Yih-Chang Chen** holds a Master of Science in Information Systems Security from the London School of Economics and Political Science (LSE), University of London, as well as a Doctorate in Computer Science from the University of Warwick, United Kingdom. He presently serves as an Assistant Professor in the Department of Social Work at Chang Jung Christian University. His academic expertise encompasses a range of interdisciplinary fields, including artificial intelligence (AI), software engineering, machine learning, social media applications, social work management, and long-term care.

His research is motivated by a dedication to bridging the divide between technology and its social applications. He actively participates in cross-disciplinary initiatives that merge information technology, management science, and social welfare to address complex societal and healthcare issues. In addition to his academic duties, he is currently an Audit Committee Member in the President Office at Chang Jung Christian University and leads projects under the auspices of Taiwan's Ministry of Labor, specifically aimed at the development and implementation of Employment-Oriented Curriculum Programs.

## How to cite this paper:

Yih-Chang Chen, "Predictive Risk Identification and Resource Optimization in Social Work: A Hybrid Semi-Supervised Learning Framework", *International Journal of Information Technology and Computer Science(IJITCS)*, Vol.18, No.5, pp.66-82, 2026. DOI:10.5815/ijitcs.2026.05.05