Journal of Neutrosophic and Fuzzy Systems

♦ ISSN: 2771-6449 Vol: 11, No. 02, 2026 ♦ PP. 01–08

♦ Article DOI: https://doi.org/10.54216/JNFS.110201

FULL LENGTH ARTICLE

Dual-Scale Fuzzy Boundary Abstention for Selective Prototype Classification

Ajoy Kanti[1,*] **Das Suman Das**[2]

[1] Associate Professor, Department of Mathematics, Tripura University, Agartala-799022, Tripura, India

[2] Assistant Professor Grade II (Mathematics), Department of Education, National Institute of Technology Calicut, Kozhikode-673601, Kerala, India

Emails: ajoykantidas@gmail.com · dr.sumandas1995@gmail.com

Received: November 20, 2025 Revised: January 22, 2026 Accepted: March 01, 2026 ⋆ Corresponding author

ABSTRACT

A classifier can be accurate on average and still be unreliable near regions in which competing classes overlap. We study this problem as selective fuzzy classification: for $x \in \mathbb{R}^d$, the decision is either a label $\hat{y}(x) \in \{1,\dots,K\}$ or abstention $\perp$. The proposed *dual-scale fuzzy boundary index* (DFBI) decomposes local ambiguity into a distributed term $B_{\text{mass}}(x)$ and an extremal term $B_{\text{press}}(x)$. The former quantifies similarity-weighted contradictory neighborhood mass, whereas the latter compares the strongest opposing and supporting fuzzy relations. Their geometric fusion $q_{\text{DFBI}}(x) = \{B_{\text{mass}}(x)B_{\text{press}}(x)\}^{1/2} \in [0,1]$ induces the selective map $g_{\theta_c}(x) = \mathbf{1}\{q_{\text{DFBI}}(x) \le \theta_c\}$, where the empirical validation quantile θ_c targets coverage c. The analysis is entirely numerical rather than graphical and uses repeated stratified splits, selective accuracy, macro-F_1, error capture, relative risk reduction, AURC, AUGRC, paired bootstrap intervals, ablation, neighborhood sensitivity, and a nonlinear stress test. At nominal $c = 0.90$, the mean selective-accuracy vector is $\bar{\mathbf{a}} = (0.9354, 0.9882, 0.9847, 0.9744)$ for Iris, Wine, Breast Cancer, and Digits, while the corresponding error-capture vector is $\bar{\mathbf{e}} = (0.5726, 0.7528, 0.8093, 0.7937)$. Relative to membership-margin uncertainty, $\Delta\mathbf{a} = (0.0211, 0.0022, 0.0194, 0.0446)$. Thus the gain is not obtained by changing the base classifier f; it follows from a local fuzzy acceptance policy (f, g_{θ_c}) whose uncertainty ordering is informed by boundary geometry.

Keywords: Fuzzy rough sets ▪ Selective classification ▪ Reject option ▪ Uncertainty ▪ Fuzzy neighborhood ▪ Prototype classification ▪ Risk–coverage analysis

1. INTRODUCTION

Fuzzy set theory replaces a binary membership statement $x \in A$ with a grade $\mu_A(x) \in [0,1]$ [1]. Rough sets express a different form of uncertainty through lower and upper approximations, customarily satisfying $\underline{A} \subseteq A \subseteq \overline{A}$ under a crisp indiscernibility relation [2]. Fuzzy–rough models combine these ideas by allowing the relation and/or the approximation grades to be valued in $[0,1]$ [3, 4]. For classification, this boundary perspective motivates a decision problem beyond ordinary error $\Pr[f(X) \neq Y]$: when the evidence around a query x is locally contradictory, should the system expose $f(x) = \hat{y}(x)$ at all?

The reject-option formulation enlarges the action space from $\mathcal{Y}$ to $\mathcal{Y} \cup \{\perp\}$, where $\perp$ denotes abstention. Classical analysis characterizes recognition error against rejection under posterior and cost assumptions [5], whereas modern selective classification couples a classifier f with a selector $g : \mathbb{R}^d \to \{0,1\}$ [6, 7]. Writing $\ell(f(X), Y) = \mathbf{1}\{f(X) \neq Y\}$, coverage and selective risk may be summarized as $\mathscr{C}(g) = \mathbb{E}[g(X)]$

and $\mathscr{R}(f,g) = \mathbb{E}[\ell(f(X),Y)g(X)]/\mathscr{C}(g)$ for $\mathscr{C}(g) > 0$. Consequently, confidence quality is not identical to predictive accuracy: two classifiers can have similar $\Pr[f(X) = Y]$ but sharply different orderings of the error events by an uncertainty score. This distinction is especially relevant when predictive confidence is miscalibrated [8] or when the classifier's global geometry fails to resolve a local curved boundary. Selective deep models address related objectives [9, 10], while recent evaluation work emphasizes that risk–coverage summaries themselves must be interpreted carefully [11].

The proposed contribution is local and fuzzy. Let $N_k(x)$ be the k nearest training objects and let $R_\sigma(x,x_j) \in (0,1]$ denote their Gaussian fuzzy relation to x. We partition the neighborhood evidence into supporting mass $S_+(x) = \sum_{j\in N_k(x)} R_\sigma(x,x_j)\mathbf{1}\{y_j = \widehat{y}\}$ and contradictory mass $S_-(x) = \sum_{j\in N_k(x)} R_\sigma(x,x_j)\mathbf{1}\{y_j \neq \widehat{y}\}$. The distributed component B_{mass} is controlled by the ratio $S_-/(S_+ + S_-)$, while the extremal component B_{press} depends on the nearest effective competition through $r_-(x)/\{r_+(x) + r_-(x)\}$. Their geometric fusion $q_{\text{DFBI}} = \sqrt{B_{\text{mass}}B_{\text{press}}}$ satisfies $\min\{B_{\text{mass}}, B_{\text{press}}\} \leq q_{\text{DFBI}} \leq \max\{B_{\text{mass}}, B_{\text{press}}\}$ and is high only when both local mechanisms indicate boundary ambiguity. The method therefore does not retrain f, relabel observations, or aggregate experts; it constructs an explicit abstention layer over a fuzzy prototype predictor.

The contribution is fourfold. First, a dual-scale fuzzy boundary statistic $q_{\text{DFBI}} : \mathbb{R}^d \to [0,1]$ is derived and analyzed through bounds, partial derivatives, limiting behavior, and a log-elasticity identity. Second, validation quantiles yield $g_{\theta_c}(x) = \mathbf{1}\{q_{\text{DFBI}}(x) \leq \theta_c\}$ without requiring posterior probabilities. Third, the empirical evidence is organized around the pair $(\mathscr{C},\mathscr{R})$ rather than a single accuracy number; AURC, AUGRC, error capture, paired differences, and sensitivity quantify how failures are concentrated by the uncertainty ranking [11]. Fourth, an ablation separates distributed fuzzy mass from extremal pressure. This role differs from recent fuzzy–rough methods that adapt rough learning for classification or perform metric-learning-based multi-label feature selection [12, 13]: here the local graded structure determines whether a prediction is accepted.

The remainder retains traditional scientific section titles while using a different internal organization from the preceding numerical studies. Section 2 relates fuzzy–rough uncertainty to selective prediction. Section 3 defines prototype uncertainty, coverage, and empirical risk ordering. Section 4 develops DFBI and the abstention algorithm. Section 5 specifies the repeated numerical protocol. Section 6 reports fixed-coverage, integrated-risk, paired, ablation, sensitivity, and nonlinear-boundary analyses. Section 7 discusses the supported scope of the claims, and Section 8 concludes.

2. RELATED WORK

2.1 Fuzzy, rough, and fuzzy–rough uncertainty

Distance-based fuzzy modeling replaces a hard class indicator by a membership vector $u(x) = (u_1,\ldots,u_K) \in \Delta^{K-1}$, where $\Delta^{K-1} = \{u \in [0,1]^K : \sum_k u_k = 1\}$. The fuzzy c-means family is a canonical example in which prototype distances induce such grades [14]. Their ambiguity can be summarized by entropy-like functionals; for normalized Shannon fuzzy entropy, $H(u) = -(\log K)^{-1}\sum_k u_k \log u_k \in [0,1]$, with the maximum attained at $u_k = 1/K$ [15]. Thus $\arg\max_k u_k$ and the dispersion of u carry distinct information.

Fuzzy–rough theory adds approximation structure to a graded relation $R(x,z) \in [0,1]$. Different implication/t-norm constructions lead to distinct lower and upper grades $\underline{R}A(x)$ and $\overline{R}A(x)$, so the mathematical treatment of the approximation operator matters in addition to fuzzifying the relation itself [4, 16]. Machine-learning applications include attribute selection [17], Gaussian-kernel uncertainty measures [18], fuzzy–rough nearest-neighbor classifiers [19, 20], and prototype selection [21]; the survey by [22] provides a broader account. In all these cases, local similarity is meaningful because two points with equal label disagreement need not contribute equally when $R(x,z_1) \gg R(x,z_2)$.

Recent work extends the same principle to robust learning and data representation. Adaptive relative fuzzy–rough learning has been developed for classification under varying uncertainty [12], while fuzzy rough sets combined with metric learning and label enhancement have been used for multi-label feature selection [13]. These studies confirm that graded structure can encode reliability, but their target variables differ from the present one. DFBI leaves both y_i and $\widehat{y}(x)$ unchanged and maps local contradiction to a scalar $q(x)$ used only by the selector g_θ.

2.2 Selective classification and uncertainty ranking

Nearest-neighbor analysis links local geometry to classification error [23]; reject-option theory adds the possibility of withholding a prediction [5]. For a selective classifier (f,g), the accepted set is $\mathscr{A}_g = \{x : g(x) = 1\}$ and the objective is not simply to minimize $\Pr[f(X) \neq Y]$, but to obtain a favorable risk–coverage trajectory as $\mathscr{A}_g$ changes [6]. Reject-option surrogate losses provide one formal learning route [7].

Deep selective models can learn a confidence or selection head jointly with the representation [9, 10]. The present setting is deliberately different: f is a simple interpretable prototype rule, and the question is whether a local fuzzy score $q_{\text{loc}}(x)$ ranks its failures better than a global membership score $q_{\text{glob}}(x) = h(u(x))$. Probability calibration is related but not equivalent because a calibrated confidence mapping concerns agreement between predicted probabilities and frequencies, whereas selective performance depends on the ordering of errors across thresholds [8]. Accordingly, we report both an operating point $\mathscr{C} \approx c$ and integrated criteria. AURC accumulates conditional selective risk, while AUGRC integrates accepted-error mass over coverage and addresses several comparison pathologies identified for selective systems [11].

3. PROBLEM FORMULATION

3.1 Fuzzy prototype prediction

Let the labeled training sample be $\mathscr{D} = \{(x_i,y_i)\}_{i=1}^n$, with $x_i \in \mathbb{R}^d$ and $y_i \in \{1,\ldots,K\}$. After standardization using training-set means and scales, class k is represented by

$$v_k = \frac{1}{n_k}\sum_{i:y_i=k} x_i, \qquad n_k = \sum_{i=1}^n \mathbf{1}\{y_i = k\}. \tag{1}$$

For $d_k^2(x) = \|x - v_k\|_2^2 + \varepsilon$, $\varepsilon > 0$, define

$$u_k(x) = \frac{[d_k^2(x)]^{-1/(m-1)}}{\sum_{\ell=1}^{K} [d_\ell^2(x)]^{-1/(m-1)}}, \qquad m > 1. \tag{2}$$

Then $u(x) \in \Delta^{K-1}$ and $\sum_k u_k(x) = 1$. With $m = 2$, $u_k(x) \propto d_k^{-2}(x)$, so $\widehat{y}(x) = \arg\max_k u_k(x) = \arg\min_k d_k^2(x)$. Equation (2) is therefore a supervised prototype analogue of distance-induced fuzzy membership and preserves the graded-distance logic of fuzzy prototype methods [14].

Three global uncertainty baselines are functions of $u(x)$ alone. Let $u_{(1)} \geq u_{(2)}$ be the two largest components. Maximum-membership uncertainty is $q_{\max} = 1 - u_{(1)}$, margin uncertainty is $q_{\mathrm{mar}} = 1 - (u_{(1)} - u_{(2)})$, and normalized entropy is

$$q_H(x) = -\frac{1}{\log K} \sum_{k=1}^{K} u_k(x) \log u_k(x), \qquad 0 \log 0 := 0. \tag{3}$$

Because $0 \leq -\sum_k u_k \log u_k \leq \log K$, it follows that $q_H \in [0,1]$. These scores distinguish concentration from dispersion in u, but none sees the labels or distances inside $N_k(x)$ once the prototypes are fixed.

3.2 Selective risk, coverage, and empirical ordering

Let $q(x)$ be an uncertainty score for which larger values indicate less reliable predictions. The threshold rule

$$g_\theta(x) = \mathbf{1}\{q(x) \leq \theta\} \tag{4}$$

defines acceptance when $g_\theta = 1$ and abstention otherwise. Population coverage and selective 0–1 risk are

$$\mathscr{C}(\theta) = \Pr[g_\theta(X) = 1], \tag{5}$$

$$\mathscr{R}(\theta) = \frac{\mathbb{E}[\mathbf{1}\{\widehat{y}(X) \neq Y\} g_\theta(X)]}{\mathscr{C}(\theta)}, \qquad \mathscr{C}(\theta) > 0. \tag{6}$$

Equivalently, $\mathscr{R}(\theta) = \Pr[\widehat{y}(X) \neq Y \mid g_\theta(X) = 1]$. The objectives conflict: decreasing $\mathscr{R}$ is trivial if $\mathscr{C} \to 0$, while $\mathscr{C} = 1$ recovers the full classifier. Hence we evaluate both a prescribed operating region $\mathscr{C} \approx c$ and the complete ordering induced by q.

For a test sample $\mathscr{D}_{\mathrm{test}} = \{(x_i, y_i)\}_{i=1}^{N}$, sort the cases so $q_{(1)} \leq \cdots \leq q_{(N)}$ and define $e_{(j)} = \mathbf{1}\{\widehat{y}_{(j)} \neq y_{(j)}\}$. At prefix j, empirical coverage, selective risk, and generalized risk are

$$\widehat{C}_j = \frac{j}{N}, \qquad \widehat{R}_j = \frac{1}{j} \sum_{t=1}^{j} e_{(t)},$$

$$\widehat{G}_j = \frac{1}{N} \sum_{t=1}^{j} e_{(t)} = \widehat{C}_j \widehat{R}_j. \tag{7}$$

AURC and AUGRC are computed by numerical integration of $\{(\widehat{C}_j, \widehat{R}_j)\}$ and $\{(\widehat{C}_j, \widehat{G}_j)\}$, respectively. Following [11], AUGRC is included because it directly tracks accepted-error mass rather than only conditional error. Lower values are preferable, but the two integrals need not rank methods identically; this is why fixed-coverage and integrated analyses are both retained.

4. PROPOSED METHOD

4.1 Fuzzy neighborhood relation

Prototype memberships in (2) are global summaries. To expose local geometry, let $N_k(x)$ denote the k nearest training points to x after standardization. If $d_{i,(k)}$ is the distance from training point i to its kth non-self neighbor, the training-only scale is

$$\sigma = \mathrm{median}_{1 \leq i \leq n} d_{i,(k)}. \tag{8}$$

The Gaussian fuzzy relation is

$$R_\sigma(x, x_j) = \exp\left(-\frac{\|x - x_j\|_2^2}{2\sigma^2}\right). \tag{9}$$

Thus $0 < R_\sigma \leq 1$ and, for $r = \|x - x_j\|_2 > 0$, $\partial R_\sigma / \partial r = -(r/\sigma^2) R_\sigma < 0$: similarity decays smoothly with distance. Gaussian fuzzy relations are established in fuzzy–rough approximation [18]; here R_σ constructs a selective boundary statistic rather than a rough lower approximation.

4.2 Distributed fuzzy boundary mass

Fix x and abbreviate $\widehat{y} = \widehat{y}(x)$. Define supporting and contradictory fuzzy masses

$$S_+(x) = \sum_{j \in N_k(x)} R_\sigma(x, x_j) \mathbf{1}\{y_j = \widehat{y}\},$$

$$S_-(x) = \sum_{j \in N_k(x)} R_\sigma(x, x_j) \mathbf{1}\{y_j \neq \widehat{y}\}. \tag{10}$$

The first boundary component is

$$B_{\mathrm{mass}}(x) = 1 - \frac{S_+(x)}{S_+(x) + S_-(x) + \varepsilon} = \frac{S_-(x) + \varepsilon}{S_+(x) + S_-(x) + \varepsilon}. \tag{11}$$

Ignoring the numerical safeguard ε, (11) is exactly the contradictory fraction $S_-/(S_+ + S_-)$. Hence $\partial B_{\mathrm{mass}} / \partial S_- > 0$ and $\partial B_{\mathrm{mass}} / \partial S_+ < 0$ for positive masses. A crisp-neighborhood baseline instead uses $q_{\mathrm{crisp}}(x) = k^{-1} \sum_{j \in N_k(x)} \mathbf{1}\{y_j \neq \widehat{y}\}$, assigning equal weight $1/k$ regardless of whether $R_\sigma(x, x_j)$ is near one or near zero.

4.3 Extremal boundary pressure

Distributed averaging can dilute a narrow competing boundary. We therefore define

$$r_+(x) = \max_{j \in N_k(x) : y_j = \widehat{y}} R_\sigma(x, x_j), \tag{12}$$

$$r_-(x) = \max_{j \in N_k(x) : y_j \neq \widehat{y}} R_\sigma(x, x_j), \tag{13}$$

with an empty-set maximum implemented as zero, and set

$$B_{\mathrm{press}}(x) = \frac{r_-(x)}{r_+(x) + r_-(x) + \varepsilon}. \tag{14}$$

For $\varepsilon \approx 0$, the odds satisfy $B_{\mathrm{press}} / (1 - B_{\mathrm{press}}) \approx r_- / r_+$. Thus $B_{\mathrm{press}} < 1/2$ corresponds approximately to stronger nearest support, $B_{\mathrm{press}} \approx 1/2$ to balanced boundary contact, and $B_{\mathrm{press}} > 1/2$ to a stronger nearest opponent. Unlike B_{mass}, this statistic depends on an extremum rather than an average and can react to a localized competing class even when S_- is modest.

4.4 Dual-scale fuzzy boundary index

The proposed score is

$$q_{\mathrm{DFBI}}(x) = \sqrt{B_{\mathrm{mass}}(x) B_{\mathrm{press}}(x)}. \tag{15}$$

Because $q_{\mathrm{DFBI}}^2 = B_{\mathrm{mass}} B_{\mathrm{press}}$, the two components interact multiplicatively: neither can be replaced by a constant without changing the ranking. For $a, b \in [0,1]$, the geometric mean satisfies $\min\{a,b\} \le \sqrt{ab} \le \max\{a,b\}$ and, by AM–GM, $\sqrt{ab} \le (a+b)/2$. Consequently, q_{DFBI} remains between its two components and is conservative relative to both their maximum and arithmetic mean.

Proposition 1 (Bounds, monotonicity, and elasticity) *For every x, $0 \le B_{\mathrm{mass}}(x), B_{\mathrm{press}}(x), q_{\mathrm{DFBI}}(x) \le 1$. For $B_{\mathrm{mass}}, B_{\mathrm{press}} > 0$,*

$$\frac{\partial q_{\mathrm{DFBI}}}{\partial B_{\mathrm{mass}}} = \frac{1}{2}\sqrt{\frac{B_{\mathrm{press}}}{B_{\mathrm{mass}}}} > 0, \qquad \frac{\partial q_{\mathrm{DFBI}}}{\partial B_{\mathrm{press}}} = \frac{1}{2}\sqrt{\frac{B_{\mathrm{mass}}}{B_{\mathrm{press}}}} > 0,$$

and

$$\frac{\partial \log q_{\mathrm{DFBI}}}{\partial \log B_{\mathrm{mass}}} = \frac{\partial \log q_{\mathrm{DFBI}}}{\partial \log B_{\mathrm{press}}} = \frac{1}{2}.$$

Thus either form of boundary evidence can increase the score, and their local proportional influence on q_{DFBI} is symmetric on the log scale.

Proof. From (11), $0 < B_{\mathrm{mass}} \le 1$ for $\varepsilon > 0$; from (14), $0 \le B_{\mathrm{press}} < 1$ because the denominator exceeds or equals the numerator. Equation (15) then gives $q_{\mathrm{DFBI}} \in [0,1]$. Differentiation yields the stated partial derivatives, while $\log q_{\mathrm{DFBI}} = \frac{1}{2}(\log B_{\mathrm{mass}} + \log B_{\mathrm{press}})$ gives both elasticities. □

Proposition 2 (Local agreement limit) *If every $j \in N_k(x)$ has $y_j = \widehat{y}(x)$, then $r_-(x) = 0$, $B_{\mathrm{press}}(x) = 0$, and therefore $q_{\mathrm{DFBI}}(x) = 0$. Moreover, $S_-(x) = 0$ and $B_{\mathrm{mass}}(x) = \varepsilon/(S_+(x)+\varepsilon) \to 0$ as $\varepsilon \downarrow 0$.*

Proof. Under unanimous local agreement, the opposing index set in (13) is empty, so the implemented maximum is $r_- = 0$. Equation (14) gives $B_{\mathrm{press}} = 0$ and (15) gives $q_{\mathrm{DFBI}} = 0$. Substitution of $S_- = 0$ into (11) proves the final limit. □

4.5 Coverage calibration and algorithm

Let $n_v = |\mathscr{D}_{\mathrm{val}}|$. For target coverage $c \in (0,1)$ we use the empirical higher quantile

$$\theta_c = \inf\left\{ t : \frac{1}{n_v} \sum_{x_i \in \mathscr{D}_{\mathrm{val}}} \mathbf{1}\{q_{\mathrm{DFBI}}(x_i) \le t\} \ge c \right\}. \tag{16}$$

The selector is then $g_{\theta_c}(x) = \mathbf{1}\{q_{\mathrm{DFBI}}(x) \le \theta_c\}$. Ties can make realized validation or test coverage exceed c, which is a property of discrete empirical scores rather than a test-label adjustment. No label from $\mathscr{D}_{\mathrm{test}}$ enters (8)–(16).

The resulting **Dual-Scale Fuzzy Boundary Abstention (DFBA)** algorithm is the composition

$$x \xrightarrow{(2)} u(x) \to \widehat{y}(x) \xrightarrow{(9)} (B_{\mathrm{mass}}, B_{\mathrm{press}}),$$
$$(B_{\mathrm{mass}}, B_{\mathrm{press}}) \xrightarrow{(15)} q_{\mathrm{DFBI}}(x) \xrightarrow{(16)} g_{\theta_c}(x).$$

Table 1. Benchmark characteristics and repeated-split design.

Data set	n	d	K	Reps.
Iris	150	4	3	60
Wine	178	13	3	60
Breast Cancer	569	30	2	60
Digits	1797	64	10	40

Operationally: (i) standardize all features with training statistics; (ii) compute $V = \{v_k\}_{k=1}^K$ by (1); (iii) fit the k-neighbor structure and σ by (8); (iv) score validation cases using (11)–(15); (v) calibrate θ_c; and (vi) return $\widehat{y}(x)$ if $g_{\theta_c}(x) = 1$, otherwise $\perp$. Once neighbors are retrieved, local aggregation is $O(k)$ and prototype membership is $O(Kd)$. A brute-force neighbor query is $O(nd)$ per test case; faster indexing can reduce retrieval cost in favorable geometries. Because DFBA never updates V or u, its full-coverage predictions are exactly those of the base prototype classifier.

5. EXPERIMENTAL DESIGN

5.1 Data and repeated-split protocol

Four public benchmark data sets are used: Iris, Wine, Breast Cancer, and Digits from the scikit-learn distribution [24]; Iris originates from Fisher's taxonomic study [25]. Their dimensions satisfy $4 \le d \le 64$ and their class counts satisfy $2 \le K \le 10$. For each replication, the sample is partitioned as $\mathscr{D} = \mathscr{D}_{\mathrm{tr}} \dot\cup \mathscr{D}_{\mathrm{val}} \dot\cup \mathscr{D}_{\mathrm{test}}$ with proportions $(0.60, 0.20, 0.20)$ under stratification. We use $R_D = 60$ repetitions for Iris, Wine, and Breast Cancer and $R_D = 40$ for Digits, so $\sum_D R_D = 220$. Standardization, V, σ, and θ_c are re-estimated within each replication, preventing information flow from $\mathscr{D}_{\mathrm{test}}$ to training or calibration.

The prespecified configuration is $(m,k,c) = (2,15,0.90)$. Comparator scores are $q_{\max}$, q_H, q_{mar}, crisp neighborhood disagreement q_{crisp}, and the direct ablation $q = B_{\mathrm{mass}}$. Because every score shares the same $\widehat{y}(x)$, differences are entirely attributable to the ordering induced by q and the resulting accepted set $\mathscr{A}_\theta = \{x : g_\theta(x) = 1\}$.

5.2 Numerical evaluation criteria

Let $\mathscr{E} = \{i : \widehat{y}_i \ne y_i\}$ be the set of full-coverage test errors and let $\mathscr{A}_\theta = \{i : g_\theta(x_i) = 1\}$. Error capture is

$$\mathrm{EC}(\theta) = \frac{|\mathscr{E} \cap \mathscr{A}_\theta^c|}{|\mathscr{E}|}. \tag{17}$$

If rejection were independent of error, one would expect roughly $\mathrm{EC} \approx 1 - \widehat{C}$; hence $\mathrm{EC} \gg 1 - \widehat{C}$ indicates that rejected cases are enriched for mistakes. Relative risk reduction is $\mathrm{RRR} = [\mathscr{R}(1) - \mathscr{R}(\theta)]/\mathscr{R}(1)$ when $\mathscr{R}(1) > 0$. We additionally report selective macro-F_1, AURC, and AUGRC from (7).

All score comparisons are paired because they use the same train/validation/test split in replication r. For a metric M, define $\Delta_r = M_r(\mathrm{DFBI}) - M_r(b)$ for "larger is better" metrics; for AUGRC we use $\Delta_r = M_r(b) - M_r(\mathrm{DFBI})$ so that $\Delta_r > 0$ always favors DFBI. We bootstrap $\{\Delta_r\}_{r=1}^{R_D}$ 5000 times and report percentile 95% intervals for $\mathbb{E}[\Delta_r]$. This pairing removes between-split variation from the comparison and

is more informative than contrasting independent standard errors.

5.3 Sensitivity and nonlinear stress test

Neighborhood sensitivity is examined at $k \in \{7,11,15,21,31\}$ on independent repeated splits; no k^* is selected retrospectively from test performance. A nonlinear two-moons problem with $n = 2400$ is generated at noise $\nu \in \{0.08,0.15,0.25\}$ with 80 repetitions per level. One centroid per class intentionally misspecifies the curved boundary, creating a setting in which $q_{\text{glob}}(x) = h(u(x))$ may be weak even when local label geometry is informative.

All pseudo-random generators use fixed seeds and the supplied Python implementation. Nearest-neighbor routines and benchmark loaders are from scikit-learn [24]. No result figure is used: every inferential statement below is traceable to a table, an algebraic contrast, or a paired confidence interval.

6. RESULTS AND NUMERICAL ANALYSIS

6.1 Fixed-coverage performance

Table 2 gives the principal operating-point results. Let $\mathbf{a}_{\text{full}} = (0.8683,0.9648,0.9268,0.8885)$ denote mean full-coverage accuracy for Iris, Wine, Breast Cancer, and Digits. Since every uncertainty mechanism is post-predictive, $\mathbf{a}_{\text{full}}$ is invariant across selectors. At $c = 0.90$, DFBI yields $\mathbf{a}_{\text{sel}} = (0.9354,0.9882,0.9847,0.9744)$, so $\Delta\mathbf{a} = \mathbf{a}_{\text{sel}} - \mathbf{a}_{\text{full}} = (0.0671,0.0234,0.0579,0.0859)$. All four coordinates are positive, indicating that the rejected set $\mathscr{A}_{\theta_c}^c$ contains a disproportionate share of base-classifier errors.

The error-capture column is more diagnostic than accuracy alone. For DFBI, the rejection rates $1-\widehat{C}$ are approximately $(0.0794,0.0986,0.1035,0.1016)$, whereas EC $= (0.5726,0.7528,0.8093,0.7937)$. The descriptive enrichment ratios $\text{EC}/(1-\widehat{C})$ are therefore about $(7.21,7.64,7.82,7.81)$: errors are several times more prevalent in the rejected set than random rejection at the same size would imply. This concentration produces RRR $= (0.5486,0.5197,0.7914,0.7721)$. Relative to the crisp score, DFBI changes selective accuracy by $(0.0058,0.0056,0.0048,0.0078)$ and error capture by $(0.0501,0.1500,0.0546,0.0672)$, confirming that fuzzy distance weighting and extremal pressure contribute information beyond an unweighted disagreement count.

The direct ablation $q = B_{\text{mass}}$ is intentionally competitive. Defining $\Delta a_{\text{mass}} = a_{\text{DFBI}} - a_{\text{mass}}$, the four differences are approximately $(-1.5\times10^{-4}, 2.70\times10^{-3}, 6.6\times10^{-4}, 3.26\times10^{-3})$. Thus Iris shows a negligible reversal, whereas Wine, Breast Cancer, and Digits favor the full fusion. The corresponding error-capture change on Wine is especially visible: $0.7528 - 0.6778 = 0.0750$. Hence the contribution of B_{press} in (14) is measurable but data dependent; the results do not support the stronger statement $M(\text{DFBI}) > M(B_{\text{mass}})$ for every metric M and every data set.

6.2 Threshold-integrated risk

A fixed coverage can conceal poor score ordering elsewhere, so Table 3 reports AURC and AUGRC. With G_b and G_D denoting comparator and DFBI AUGRC, define $\rho_G = (G_b - G_D)/G_b$. Against membership margin, $\rho_G = (0.389,0.173,0.426,0.470)$ for Iris, Wine, Breast Cancer, and Digits. Numerically, AUGRC changes from $0.02008 \to 0.01227$, $0.00320 \to 0.00265$, $0.00982 \to 0.00563$, and $0.01576 \to 0.00835$. Thus the fixed-coverage improvements are accompanied by materially lower accepted-error mass over the full threshold ordering.

The local ablations reveal a more nuanced result. DFBI has the lowest AUGRC on Wine and Breast Cancer, whereas on Iris $G_{\text{mass}} - G_{\text{DFBI}} = 0.01212 - 0.01227 = -1.5\times10^{-4}$. On Digits, $G_{\text{crisp}} - G_{\text{DFBI}} \approx -1.45\times10^{-4}$ and $G_{\text{mass}} - G_{\text{DFBI}} \approx -1.3\times10^{-4}$, so the two local ablations are marginally better in integrated generalized risk. Therefore $a_{\text{sel}}^{\text{DFBI}} > a_{\text{sel}}^{\text{crisp}}$ at the target operating point does not imply $\text{AUGRC}_{\text{DFBI}} < \text{AUGRC}_{\text{crisp}}$ globally. Keeping both summaries prevents an unsupported dominance claim.

6.3 Paired uncertainty contrasts

Table 4 summarizes the empirical distribution of paired contrasts Δ_r. Against margin uncertainty, $\widehat{\mathbb{E}}[\Delta_r]$ for selective accuracy is $(0.02109,0.00217,0.01943,0.04458)$. The 95% intervals exclude zero for Iris, Breast Cancer, and Digits, while Wine gives $[-0.00214,0.00681]$. For AUGRC, the sign convention $\Delta_r = G_r(\text{margin}) - G_r(\text{DFBI})$ yields intervals entirely above zero on all four data sets, with mean advantages from 5.53×10^{-4} to 7.81×10^{-3}. The paired formulation therefore distinguishes a stable ranking improvement from a fixed-coverage difference that may be small relative to split-to-split variability.

Against the crisp neighborhood, DFBI's error-capture advantage is positive with intervals excluding zero on all four data sets. The selective-accuracy interval excludes zero on Wine, Breast Cancer, and Digits but not Iris. This paired evidence supports a modest claim: fuzzy weighting plus extremal pressure usually improves the selected 90%-coverage operating point over an unweighted local disagreement count. It does not establish universal superiority over every fuzzy ablation or every threshold.

6.4 Neighborhood sensitivity

Table 5 quantifies sensitivity to k. Let $\omega_D = \max_k a_D(k) - \min_k a_D(k)$ over $k \in \{7,11,15,21,31\}$. From the reported means, $\omega_D = (0.0219,0.0075,0.0062,0.0046)$ for Iris, Wine, Breast Cancer, and Digits. Thus $k = 15$ is not a singular optimum: Breast Cancer is highest at $k = 11$, while Digits is essentially tied at $k = 11$ and 15. The small ω_D values outside Iris indicate that the main fixed-coverage conclusion is not produced by a sharply tuned neighborhood size.

AURC/AUGRC sensitivity tells the same general story. On Wine, AUGRC is $(0.00272,0.00260,0.00314)$ for $k = (11,15,21)$; on Breast Cancer it is $(0.00551,0.00534,0.00552)$. On Digits, increasing k from 7 to 31 changes AUGRC from 0.00732 to 0.00830, a relative increase of $(0.00830/0.00732 - 1)\times100\% \approx 13.4\%$. This direction is consistent with a locality tradeoff: very large k broadens $N_k(x)$ and can mix boundary evidence from farther regions.

6.5 Nonlinear boundary stress test

The two-moons experiment tests a different failure mode. Because a single prototype per class cannot represent the

Table 2. Fixed-coverage numerical comparison. Coverage is validation-calibrated toward $c = 0.90$. EC is error capture. RRR is relative risk reduction.

Data set	Score	Full acc.	Coverage	Sel. acc.	Sel. F_1	EC	RRR
Iris	Margin	.8683	.9072	.9144	.9096	.4108	.3609
	Crisp neighborhood	.8683	.9289	.9296	.9266	.5225	.4998
	Fuzzy mass	.8683	.9189	**.9356**	**.9324**	.5654	.5419
	DFBI	.8683	.9206	.9354	.9322	**.5726**	**.5486**
Wine	Margin	.9648	.8981	.9860	.9859	.7139	.4771
	Crisp neighborhood	.9648	.9181	.9826	.9831	.6028	.3557
	Fuzzy mass	.9648	.9032	.9855	.9858	.6778	.4334
	DFBI	.9648	.9014	**.9882**	**.9885**	**.7528**	**.5197**
Breast Cancer	Margin	.9268	.9003	.9652	.9609	.5662	.5203
	Crisp neighborhood	.9268	.9096	.9799	.9776	.7548	.7344
	Fuzzy mass	.9268	.8962	.9840	.9821	.8036	.7849
	DFBI	.9268	.8965	**.9847**	**.9829**	**.8093**	**.7914**
Digits	Margin	.8885	.8997	.9298	.9272	.4338	.3725
	Crisp neighborhood	.8885	.9078	.9666	.9652	.7265	.7005
	Fuzzy mass	.8885	.9006	.9712	.9699	.7668	.7428
	DFBI	.8885	.8984	**.9744**	**.9733**	**.7937**	**.7721**

Table 3. Threshold-integrated comparison (lower is better).

Data	Score	AURC	AUGRC
Iris	Margin	.02368	.02008
	Crisp	.01395	.01268
	Fuzzy mass	**.01320**	**.01212**
	DFBI	.01339	.01227
Wine	Margin	.00349	.00320
	Crisp	.00357	.00329
	Fuzzy mass	.00303	.00283
	DFBI	**.00283**	**.00265**
Breast Cancer	Margin	.01229	.00982
	Crisp	.00797	.00599
	Fuzzy mass	**.00726**	.00578
	DFBI	.00755	**.00563**
Digits	Margin	.01795	.01576
	Crisp	.00919	**.00821**
	Fuzzy mass	**.00891**	.00822
	DFBI	.00944	.00835

curved geometry, the base risk remains high: at $v = 0.08$, full accuracy 0.8602 corresponds to $\mathscr{R}(1) = 0.1398$. Crisp and fuzzy-mass scores have enough ties that the higher validation quantile accepts all test points, $\widehat{C} = 1$, leaving selective risk unchanged. DFBI breaks this degeneracy: $\widehat{C} = 0.8990$, selective accuracy 0.9570 gives $\widehat{R} \approx 0.0430$, and EC $= 0.7217$. Thus roughly 72% of the errors are concentrated into about 10% of rejected cases.

At $v = 0.15$, the fixed-coverage difference over margin is $0.9539 - 0.8838 = 0.0701$. At $v = 0.25$, however, $a_{\text{mass}} - a_{\text{DFBI}} = 9 \times 10^{-4}$ and the crisp AUGRC is lower by about 7.7×10^{-5}. This sign reversal is informative rather than problematic: as local disagreement becomes diffuse, S_- already captures much of the ambiguity and the marginal information supplied by the extremum r_- diminishes. The evidence therefore supports conditional utility of the pressure term, not universal dominance.

7. DISCUSSION

The numerical evidence separates three uncertainty mechanisms. Global prototype scores have the form $q_{\text{glob}}(x) = h(u(x))$ and therefore inherit the geometry of $V = \{v_k\}$. Crisp neighborhood disagreement is local but replaces all similarities by $1/k$. Fuzzy mass retains R_σ, whereas DFBI adds the extremal ratio implicit in $B_{\text{press}} \approx r_-/(r_+ + r_-)$. In symbols, the information path changes from $x \mapsto u(x) \mapsto q_{\text{glob}}$ to $x \mapsto N_k(x) \mapsto (S_+, S_-, r_+, r_-) \mapsto q_{\text{DFBI}}$. The largest gains on Digits and on the misspecified two-moons classifier are consistent with the hypothesis that local boundary variables contain information absent from global prototype confidence.

The multiplicative fusion in (15) should not be read as a theorem that two components always dominate one. A natural generalization is

$$q_\lambda(x) = B_{\text{mass}}(x)^\lambda B_{\text{press}}(x)^{1-\lambda}, \qquad 0 \le \lambda \le 1,$$

for which $\log q_\lambda = \lambda \log B_{\text{mass}} + (1-\lambda)\log B_{\text{press}}$. The present method fixes $\lambda = 1/2$ to avoid an additional tuned parameter. Tables 3 and 6 show why this restraint matters: B_{mass} alone has slightly lower integrated risk in some cases, and at high moons noise its fixed-coverage accuracy is marginally higher. The supported claim is therefore narrower: B_{press} is useful when strong local competition is hidden by averaging or when score ties obscure a discrete disagreement count.

The formulation also clarifies how DFBA differs from fuzzy–rough prediction itself. Existing fuzzy–rough nearest-neighbor and prototype methods use graded approximations to determine or simplify the class decision [19, 20, 21]; recent methods adapt fuzzy–rough learning for classification or inte-

Table 4. Paired bootstrap contrasts for DFBI. Positive values favor DFBI. Intervals are 95% percentile intervals from 5000 paired bootstrap resamples.

Comparator	Data set	Sel.-accuracy advantage	Error-capture advantage	AUGRC advantage
Margin	Iris	.02109 [.01078,.03184]	.16188 [.07376,.25443]	.00781 [.00596,.00972]
	Wine	.00217 [-.00214,.00681]	.03889 [-.06389,.14444]	.00055 [.00009,.00102]
	Breast Cancer	.01943 [.01595,.02298]	.24311 [.19987,.28671]	.00419 [.00352,.00487]
	Digits	.04458 [.04013,.04902]	.35989 [.32456,.39560]	.00740 [.00675,.00805]
Crisp neighborhood	Iris	.00585 [-.00063,.01223]	.05012 [.00944,.09183]	–
	Wine	.00560 [.00245,.00911]	.15000 [.07222,.24167]	–
	Breast Cancer	.00475 [.00273,.00703]	.05458 [.03336,.07851]	–
	Digits	.00779 [.00579,.00969]	.06719 [.04977,.08430]	–

Table 5. DFBI neighborhood sensitivity. Entries are selective accuracy / error capture.

Data	$k=7$	11	15	21	31
Iris	.9456/.520	.9562/.650	.9488/.587	.9550/.635	.9343/.481
Wine	.9926/.825	.9911/.808	.9896/.758	.9865/.683	.9851/.658
Breast	.9843/.803	.9872/.839	.9833/.791	.9844/.808	.9810/.753
Digits	.9743/.772	.9745/.775	.9745/.775	.9720/.751	.9699/.735

Table 6. Nonlinear two-moons stress test (80 repetitions per noise level).

Noise / score	Cov.	Sel. acc.	EC	AUGRC
.08 Margin	.9013	.8859	.2642	.0542
.08 Crisp	1.0000	.8602	.0000	.0099
.08 Fuzzy mass	1.0000	.8602	.0000	.0099
.08 DFBI	.8990	**.9570**	**.7217**	.0099
.15 Margin	.8991	.8838	.2617	.0581
.15 Crisp	.9756	.8820	.1676	.0107
.15 Fuzzy mass	.9750	.8826	.1717	**.0106**
.15 DFBI	.8998	**.9539**	**.7057**	.0107
.25 Margin	.8975	.8725	.2282	.0661
.25 Crisp	.9074	.9305	.5722	**.0170**
.25 Fuzzy mass	.9023	**.9346**	**.5995**	.0177
.25 DFBI	.9030	.9337	.5939	.0171

grate it with metric learning and label enhancement for feature selection [12, 13]. DFBA preserves $f(x) = \widehat{y}(x)$ and acts only through g_{θ_c}. Hence the deployable object is the pair (f, g_{θ_c}), with empirical behavior summarized by $(\widehat{C}, \widehat{R}, \widehat{G})$ rather than by accuracy alone. This separation is useful when rejected cases can be routed to a human, a second-stage model, or a costlier measurement process.

Several limitations remain. First, the experiments use compact public data and a deliberately simple representation $\phi(x) = x \in \mathbb{R}^d$; learned embeddings $\phi(x) \in \mathbb{R}^p$ may alter both prototype and neighborhood geometry. Second, (16) guarantees a validation coverage of at least c under the empirical higher quantile, not exact test coverage c; ties can produce $\widehat{C} > c$, as observed in the moons experiment. Third, k and σ are geometric parameters. Although the observed ranges ω_D are modest for three of four benchmarks, nonuniform densities may favor an adaptive bandwidth $\sigma(x)$. Fourth, the paired bootstrap intervals quantify variability over the specified repeated-split mechanism and seeds; they are not distribution-free guarantees for unrelated populations. Finally, rejection cost is not modeled explicitly, so an application-specific loss $L = \ell g + c_\perp(1-g)$ would be needed when abstention itself has a known economic or clinical cost.

8. CONCLUSION

This paper formulated fuzzy selective classification as a composition of prediction, local boundary measurement, and acceptance:

$$x \mapsto \big(f(x), q_{\mathrm{DFBI}}(x)\big),$$
$$(f, g_{\theta_c})(x) \in \mathscr{Y} \cup \{\perp\}.$$

The Gaussian relation R_σ generates distributed contradictory mass $B_{\mathrm{mass}} \in [0,1]$ and extremal boundary pressure $B_{\mathrm{press}} \in [0,1]$; their product relation $q^2_{\mathrm{DFBI}} = B_{\mathrm{mass}} B_{\mathrm{press}}$ yields an interpretable uncertainty ordering. The higher validation quantile in (16) then determines the operating coverage without changing V, $u(x)$, or the full-coverage label $\widehat{y}(x)$.

Across $\sum_D R_D = 220$ real-data partitions, mean selective accuracy lies in $[0.9354, 0.9882]$ at approximately 90% coverage, while error capture lies in $[0.5726, 0.8093]$. Against membership margin, the fixed-coverage accuracy differences are $\Delta a \in [0.0022, 0.0446]$ and the AUGRC reductions are approximately 17.3%–47.0%. Paired intervals support the strongest accuracy improvements on Iris, Breast Cancer, and Digits, while Wine's fixed-coverage interval includes zero. The nonlinear stress test further shows that local fuzzy pressure can reduce selected risk from about 0.1398 to 0.0430 at $\nu = 0.08$ when global prototype geometry is misspecified. At the same time, ablation reversals show that B_{mass} alone can be equally good or slightly better when disagreement is already diffuse. The resulting conclusion is therefore specific: DFBA is an interpretable, numerically auditable abstention layer whose dual-scale term is most useful when local competition is concentrated, with adaptive fusion, local bandwidths, and explicit reject costs remaining open extensions.

REFERENCES

[1] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[2] Z. Pawlak, "Rough sets," *International Journal of Computer and Information Sciences*, vol. 11, pp. 341–356, 1982.

[3] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy

rough sets," *International Journal of General Systems*, vol. 17, no. 2–3, pp. 191–209, 1990.

[4] A. M. Radzikowska and E. E. Kerre, "A comparative study of fuzzy rough sets," *Fuzzy Sets and Systems*, vol. 126, no. 2, pp. 137–155, 2002.

[5] C.-K. Chow, "On optimum recognition error and reject tradeoff," *IEEE Transactions on Information Theory*, vol. 16, no. 1, pp. 41–46, 1970.

[6] R. El-Yaniv and Y. Wiener, "On the foundations of noise-free selective classification," *Journal of Machine Learning Research*, vol. 11, pp. 1605–1641, 2010.

[7] P. L. Bartlett and M. H. Wegkamp, "Classification with a reject option using a hinge loss," *Journal of Machine Learning Research*, vol. 9, pp. 1823–1840, 2008.

[8] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," *Proceedings of the 34th International Conference on Machine Learning*, vol. 70, pp. 1321–1330, 2017.

[9] Y. Geifman and R. El-Yaniv, "Selective classification for deep neural networks," *Advances in Neural Information Processing Systems*, vol. 30, pp. 4878–4887, 2017.

[10] Y. Geifman and R. El-Yaniv, "Selectivenet: A deep neural network with an integrated reject option," *Proceedings of the 36th International Conference on Machine Learning*, vol. 97, pp. 2151–2159, 2019.

[11] J. Traub, T. J. Bungert, C. T. Lüth, M. Baumgartner, K. H. Maier-Hein, L. Maier-Hein, and P. F. Jäger, "Overcoming common flaws in the evaluation of selective classification systems," *Advances in Neural Information Processing Systems*, vol. 37, 2024.

[12] Y. Zhang, C. Wang, Y. Huang, W. Ding, and Y. Qian, "Adaptive relative fuzzy rough learning for classification," *IEEE Transactions on Fuzzy Systems*, vol. 32, no. 11, pp. 6267–6276, 2024.

[13] M. Cai, M. Yan, P. Wang, and F. Xu, "Multi-label feature selection based on fuzzy rough sets with metric learning and label enhancement," *International Journal of Approximate Reasoning*, vol. 168, p. 109149, 2024.

[14] J. C. Bezdek, R. Ehrlich, and W. Full, "Fcm: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2–3, pp. 191–203, 1984.

[15] A. De Luca and S. Termini, "A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory," *Information and Control*, vol. 20, no. 4, pp. 301–312, 1972.

[16] M. De Cock, C. Cornelis, and E. E. Kerre, "Fuzzy rough sets: The forgotten step," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 1, pp. 121–130, 2007.

[17] R. Jensen and Q. Shen, "Fuzzy-rough sets assisted attribute selection," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 1, pp. 73–89, 2007.

[18] Q. Hu, L. Zhang, D. Chen, W. Pedrycz, and D. Yu, "Gaussian kernel based fuzzy rough sets: Model, uncertainty measures and applications," *International Journal of Approximate Reasoning*, vol. 51, no. 4, pp. 453–471, 2010.

[19] M. Sarkar, "Fuzzy-rough nearest neighbor algorithms in classification," *Fuzzy Sets and Systems*, vol. 158, no. 19, pp. 2134–2152, 2007.

[20] R. Jensen and C. Cornelis, "Fuzzy-rough nearest neighbour classification and prediction," *Theoretical Computer Science*, vol. 412, no. 42, pp. 5871–5884, 2011.

[21] N. Verbiest, C. Cornelis, and F. Herrera, "Frps: A fuzzy rough prototype selection method," *Pattern Recognition*, vol. 46, no. 10, pp. 2770–2782, 2013.

[22] S. Vluymans, L. D'eer, Y. Saeys, and C. Cornelis, "Applications of fuzzy rough set theory in machine learning: A survey," *Fundamenta Informaticae*, vol. 142, no. 1–4, pp. 53–86, 2015.

[23] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[25] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.