Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Saturation of the Global Radius in Neighbourhood Rough-Set Attribute Reduction for High-Dimensional Small-Sample Data

Thi Thuy Tram Nguyen
Graduate University of Science and Technology, Vietnam Academy of Science and Technology
Long Giang Nguyen
Institute of Information Technology, Vietnam Academy of Science and Technology
Manh Hung Nguyen
Le Quy Don Technical University
Thanh Dai Tran
ttdaiuneti@gmail.com

University of Economics - Technology for Industries

Research Article

Keywords:

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11397419/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Saturation of the Global Radius in Neighbourhood Rough-Set Attribute Reduction for High-Dimensional Small-Sample Data

Thi Thuy Tram Nguyen[1,2], Long Giang Nguyen[3], Manh Hung Nguyen[4], Thanh Dai Tran[5*]

[1]Graduate University of Science and Technology, Vietnam Academy of Science and Technology (GUST-VAST), 18 Hoang Quoc Viet, Cau Giay District, Hanoi, 100000, Vietnam.

[2]Center for Information Technology Development, University of Information Technology, Vietnam National University Ho Chi Minh City (VNU-HCM), Quarter 6, Linh Trung Ward, Thu Duc City, Ho Chi Minh City, 815507, Vietnam.

[3]Institute of Information Technology, Vietnam Academy of Science and Technology (VAST), 18 Hoang Quoc Viet, Cau Giay District, Hanoi, 100000, Vietnam.

[4]Le Quy Don Technical University, 236 Hoang Quoc Viet, Cau Giay District, Hanoi, 100000, Vietnam.

[5*]University of Economics - Technology for Industries (UNETI), 456 Minh Khai, Hai Ba Trung District, Hanoi, 100000, Vietnam.

*Corresponding author(s). E-mail(s): ttdaiuneti@gmail.com;
Contributing authors: tramntt@uit.edu.vn; nlgiang@ioti.vn;
manhhungk12@mta.edu.vn;

Abstract

Neighbourhood rough-set (NRS) attribute reduction counts the samples whose neighbourhood of a fixed radius contains no sample of another class. We analyse why this count fails as a search criterion when attributes far outnumber samples. We show that it depends on the data only through the distance of each sample to its nearest sample of another class. Once all these distances exceed the radius, the dependency saturates and the search stops. With a radius large enough to avoid this, every small subset has an empty positive region, so greedy search

cannot start. As a reference without a distance radius, we use the published maximal homogeneous metric granules (MHMG), whose energy compares distances only with each other and bounds the training leave-one-out nearest-neighbour accuracy. On seventeen gene-expression datasets, even random attribute subsets saturated NRS after a few attributes. With 3-NN, MHMG was more accurate than NRS at every radius where NRS returned a reduct, also at equal size, and than random subsets and two nearest-neighbour variants of NRS. Against the best of eight radii chosen on the test folds, its margin was small. This evaluator favours MHMG, whose energy is tied to nearest-neighbour accuracy. With an SVM, fuzzy NRS with 50 attributes was more accurate than MHMG, and the full attribute set beat every subset of at most ten attributes. Tuned L1-regularised logistic regression (L1) was as accurate as MHMG, more stable and about 158 times faster. The rounding in its stopping rule changes the reduct size of MHMG.

Keywords: Attribute reduction, Neighbourhood rough sets, High-dimensional data, Small-sample data, Gene expression, Feature selection

1 Introduction

1.1 Motivation

Data with many more attributes than samples are now common. Gene-expression profiles, for example, measure thousands of genes on tens to a few hundred patients [1, 2]. In this regime most attributes are irrelevant or redundant, and selecting a small subset lowers the cost of measurement, reduces overfitting and makes the model easier to interpret. Attribute reduction in rough set theory addresses this task. It looks for a small subset of conditional attributes that keeps the ability of the full set to determine the decision [3]. Neighbourhood rough sets (NRS) extend it to numerical data by treating samples within a radius ε as indistinguishable [4]. A greedy search then adds attributes until the dependency of the full attribute set is reached. The radius is a parameter of the model, and the result depends on it.

The dependence on ε becomes a real obstacle when the number of attributes m far exceeds the number of samples n. Distances computed on one attribute and on ten attributes live on different scales, so a single radius cannot suit all the subsets that the search visits. Each class has few samples, so tuning ε by cross-validation is itself unreliable. Recent models therefore adapt the neighbourhood to the data, through granular balls, k nearest neighbours or data-driven radii, but they still carry a size, a purity or a radius rule.

Maximal homogeneous metric granules (MHMG) [5] have no distance radius. The granule of a sample is the open ball that reaches its nearest sample of another class, and an energy of the granule sizes scores an attribute subset. MHMG was proposed and evaluated on twelve benchmarks of the UCI Machine Learning Repository with at most a few hundred attributes [5]. It is not known why a fixed radius fails when $m \gg n$, or whether a criterion without one does better than the neighbourhood models that adapt their radius. This paper analyses fixed-radius reduction in this regime. It

does not propose a new selector: MHMG is used unchanged from [5] as the reference criterion without a distance radius.

1.2 Related work

Neighbourhood rough sets and their radius. NRS defines the positive region by ε-neighbourhoods and selects attributes greedily [4]; fuzzy neighbourhood rough sets (FNRS) replace the crisp neighbourhood by a fuzzy relation of the same radius [6]. Several models reduce the dependence on a global radius. Granular-ball rough sets cover the data with balls that are split until a purity threshold is met [7]. Sun et al. combined granular balls with fuzzy neighbourhoods for high-dimensional data [8] and for multilabel data [9]. k-nearest-neighbourhood rough sets use the k nearest samples as the neighbourhood [10], and the weighted variant of Wang and Zhao weights these neighbours by their similarity [11]. Zhang and Wang weighted the nearest neighbours by sample quality to resist label noise [12]. Adaptive neighbourhood rough sets derive a radius for each sample from the local distribution of distances [13], and personalised granularity uses nearest mutual neighbours [14]. Each of these models replaces ε by another parameter: a purity threshold, a number of neighbours, or a rule with its own constants.

Faster neighbourhood reduction. Another line of work makes NRS reduction faster. Yang et al. showed that the NRS approximations depend only on the pairs of samples of different classes that lie within the radius, and used this to skip most neighbourhood computations [15]. Thuy and Wongthanavasu defined a reduct that preserves a stripped neighbourhood cover (FSSNC). Their algorithm selects attributes by an error measure on this cover and runs fast on high-dimensional mixed tables [16]. Other recent criteria weight the attributes [17], enforce monotonicity [18], or build fuzzy neighbourhood measures for gene-expression data [19, 20]. Neighbourhood entropy criteria [20, 21] score a subset by the uncertainty of its neighbourhoods rather than by the number of pure ones, and the neighbourhood decision error rate (NDER) scores it by the leave-one-out error of a neighbourhood classifier [22]. These criteria also use a global radius.

Nearest enemies and local sets. The MHMG granule is the local set of instance selection: the samples closer to x than its nearest enemy [23, 24]. Margin-based feature selection uses the distance to the nearest enemy as a hypothesis margin [25], and Relief-type estimators score each attribute by its nearest misses [26, 27]. Classification-complexity measures, such as the leave-one-out 1-NN error and the local-set average cardinality, describe the same geometry for a whole dataset [28, 29]. MHMG differs in that it scores attribute subsets by the sizes of the local sets and stops when they no longer grow.

Attribute selection when $m \gg n$. Outside rough sets, filters such as the ANOVA F statistic, minimum-redundancy maximum-relevance [30, 31], ReliefF, and embedded methods such as L1-regularised logistic regression (L1) [32] and recursive feature elimination with a support vector machine (SVM-RFE) [33] are the usual choices; their output size must be chosen or tuned. With few samples, the selected subsets also vary across resamples [34, 35]. We use these selectors as external references.

Table 1 Neighbourhood rough-set models compared in this study and the parameter that fixes their neighbourhood.

Model	Neighbourhood of a sample	Parameter
NRS [4]	ball of radius ε	ε
FNRS [6]	fuzzy ball of radius ε	ε, dependency target
GBNRS [7]	granular ball containing it	purity threshold
KNN-NRS [10]	k nearest samples	k
WKNRS [11]	k nearest samples, weighted	k, threshold α
Adaptive NRS [13]	ball of a data-driven radius	clipping constants
MHMG [5]	ball up to the nearest enemy	rounding precision of γ

To sum up, existing neighbourhood models either keep a global radius or replace it with another parameter. Their evaluations on high-dimensional data report accuracy. They do not analyse how the radius itself drives the search when $m \gg n$, and no criterion without a distance radius has been compared with them in this regime. Table 1 summarises the neighbourhood models compared in this paper by the parameter that fixes their output.

1.3 Contributions and organisation

This paper makes three contributions.

1. We show that NRS and a simplified FNRS depend on the data only through the safety radii, which extends a recent result for NRS [15] (Proposition 2). From this form we derive two failures of the NRS dependency when $m \gg n$. It saturates once every radius passes the threshold (Proposition 3), and with a large radius a greedy search cannot take its first step (Proposition 4).
2. We compare MHMG with the other neighbourhood rough-set models on seventeen gene-expression datasets. Even random attribute subsets saturated NRS after a few attributes, whereas the MHMG energy compares distances only with each other and bounds the training leave-one-out accuracy (Proposition 5, Theorem 1). With a three-nearest-neighbour (3-NN) classifier, MHMG was more accurate than NRS at every radius at which NRS returned a reduct, and more accurate than KNN-NRS and WKNRS.
3. We place the criterion among standard selectors and examine its variability. MHMG matched the best selector of its size, while tuned L1 was slightly more accurate, faster and more stable. A safety radius changes on held-out data exactly when all its nearest enemies are removed (Proposition 6), which fits the variability of the reducts across folds and under label noise.

The remainder of the paper is organised as follows. Section 2 recalls NRS, FNRS and MHMG. Section 3 explains why a fixed-radius dependency fails when $m \gg n$ and what the MHMG criterion optimises. Section 4 compares MHMG with the neighbourhood rough-set family and with standard selectors on seventeen gene-expression datasets, and reports the costs of the criterion. Section 5 discusses the results and Section 6 concludes.

Table 2 Main notation.

Symbol	Meaning
U, n; C, m	samples and their number; attributes and their number
B; R, $q = \lvert R\rvert$	an attribute subset; the reduct returned by a method and its size
d, $[x]_d$, $\overline{[x]}_d$, $n_{\max}$	class label, class of x, enemies of x, largest class size
dist_B	Euclidean distance on the attributes in B
ε; $N_B^\varepsilon(x)$, $\mathrm{POS}^\varepsilon(B)$	global radius; ε-neighbourhood of x and positive region
μ_B; $\gamma_{\mathrm{N}}^\varepsilon$, $\gamma_{\mathrm{F}}^\varepsilon$	fuzzy relation; NRS and FNRS dependency
$\delta_B(x)$, $\mathrm{NE}_B(x)$	safety radius of x and its nearest enemies
$G_B(x)$; $\phi(t)$	MHMG granule; $\log_2(t+1)$
$E(B)$, K_d, $\gamma(B)$	energy, its maximum, normalised dependency
k; $A_k(B)$; $\mathrm{acc}_k(B)$	number of neighbours; samples with $\lvert G_B(x)\rvert \geq k+1$; leave-one-out k-NN accuracy
S, s; $\delta'_B(x)$	held-out samples and their number; radius after removing S

2 Preliminaries

This section recalls decision systems, the two neighbourhood rough-set criteria with a global radius, NRS and FNRS, and the MHMG criterion of [5], which has no distance radius. The models that adapt the neighbourhood (KNN-NRS, WKNRS, GBNRS and adaptive NRS) are summarised in Table 1 and their settings are given in Section 4.1. Table 2 lists the main symbols.

2.1 Decision systems and neighbourhood rough sets

A decision system is $DS = \langle U, C \cup \{d\}\rangle$, where $U = \{x_1, \ldots, x_n\}$ is a finite set of samples, $C = \{a_1, \ldots, a_m\}$ a set of conditional attributes and d the decision attribute. The attribute values $a(x)$ lie in $[0,1]$ after min-max scaling on U; in the experiments, U is the training part of a fold. For $B \subseteq C$ and $x, y \in U$,

$$\mathrm{dist}_B(x,y) = \Big(\sum_{a\in B}\big(a(x)-a(y)\big)^2\Big)^{1/2}, \tag{1}$$

with $\mathrm{dist}_\emptyset \equiv 0$. After min-max scaling this is the range-normalised Euclidean distance used in [5], in which each attribute is weighted by the inverse square of its range. The decision class of x is $[x]_d = \{y \in U : d(y) = d(x)\}$, $n_{\max} = \max_x |[x]_d|$, and the enemy set of x is $\overline{[x]}_d = U \setminus [x]_d$.

Assumption 1 The decision d takes at least two values on U. Hence $\overline{[x]}_d \neq \emptyset$ for every $x \in U$.

Definition 1 (Neighbourhood dependency [4]) For $\varepsilon > 0$ and $B \subseteq C$, let $N_B^\varepsilon(x) = \{y \in U : \mathrm{dist}_B(x,y) \leq \varepsilon\}$, $\mathrm{POS}^\varepsilon(B) = \{x \in U : N_B^\varepsilon(x) \subseteq [x]_d\}$ and $\gamma_{\mathrm{N}}^\varepsilon(B) = |\mathrm{POS}^\varepsilon(B)|/n$. A subset $B \subseteq C$ is an NRS reduct if $\gamma_{\mathrm{N}}^\varepsilon(B) = \gamma_{\mathrm{N}}^\varepsilon(C)$ and $\gamma_{\mathrm{N}}^\varepsilon(B \setminus \{a\}) < \gamma_{\mathrm{N}}^\varepsilon(B)$ for every $a \in B$.

Definition 2 (Fuzzy neighbourhood dependency) For $\varepsilon > 0$ and $B \subseteq C$, let $\mu_B(x,y) = \max\big(0, 1 - \mathrm{dist}_B(x,y)/\varepsilon\big)$ and

$$\gamma_{\mathrm{F}}^{\varepsilon}(B) = \frac{1}{n} \sum_{x \in U} \min_{y \in [x]_d} \big(1 - \mu_B(x,y)\big), \tag{2}$$

which is well defined by Assumption 1. Since $1 - \mu_B(x,y) = \min\big(1, \mathrm{dist}_B(x,y)/\varepsilon\big)$, $\gamma_{\mathrm{F}}^{\varepsilon}(B)$ is the mean degree to which each sample is separated from its closest enemy. Each term is the membership of x in the lower approximation of its own class when the fuzzy implication $I(u,v) = \max(1-u, v)$ is used; this is a simplified form of the model of [6].

Both criteria compare distances with the global radius ε. MHMG replaces it by one radius per sample (Fig. 1).

2.2 Maximal homogeneous metric granules

Definition 3 (Safety radius, granule and nearest enemies [5]) For $B \subseteq C$ and $x \in U$, the safety radius of x is

$$\delta_B(x) = \min_{y \in [x]_d} \mathrm{dist}_B(x,y), \tag{3}$$

which exists by Assumption 1 because $[x]_d$ is finite and non-empty. The maximal homogeneous metric granule of x is the open ball

$$G_B(x) = \{y \in U : \mathrm{dist}_B(x,y) < \delta_B(x)\}, \tag{4}$$

and the nearest-enemy set of x is $\mathrm{NE}_B(x) = \{y \in [x]_d : \mathrm{dist}_B(x,y) = \delta_B(x)\}$, which is non-empty because the minimum in (3) is attained.

Definition 4 (Energy and normalised dependency [5]) With $\phi(t) = \log_2(t+1)$ for $t \in \mathbb{Z}_{\geq 0}$, the energy of B and its normalised form are

$$\begin{aligned} E(B) &= \sum_{x \in U} \phi\big(|G_B(x)|\big), \qquad \gamma(B) = \frac{E(B)}{K_d}, \\ K_d &= \sum_{x \in U} \phi\big(|[x]_d|\big). \end{aligned} \tag{5}$$

Since $|[x]_d| \geq 1$ and $\phi(1) = 1$, $K_d \geq n > 0$.

The MHMG reduct is found by a greedy forward search that adds the attribute with the largest γ as long as γ, rounded to two decimals, strictly increases (Algorithm 1 in Section 3.6). No radius or size is set; the rounding precision of the stopping test is the only constant.

Definition 5 (Separating subset) A subset $B \subseteq C$ is *separating* if $\delta_B(x) > 0$ for every $x \in U$.

Lemma 1 *A subset $B \subseteq C$ is separating if and only if*

$$\forall x, y \in U : \quad d(x) \neq d(y) \implies \exists a \in B : \; a(x) \neq a(y). \tag{6}$$

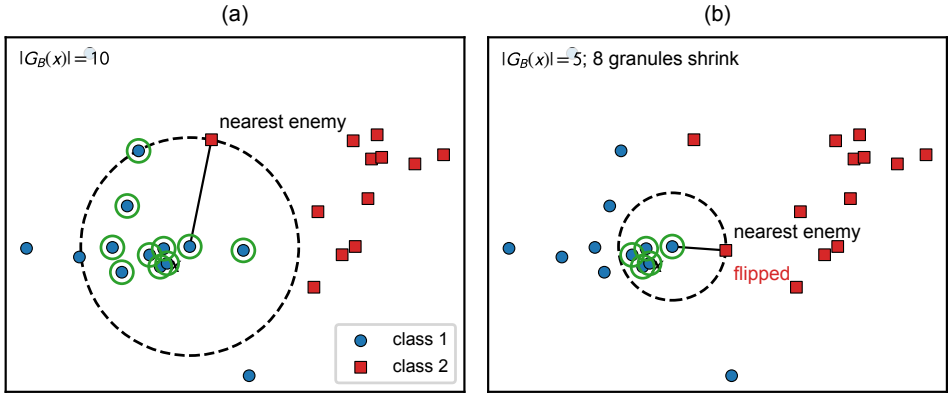


Fig. 1 Two-dimensional granules. (a) The safety radius of x is its distance to the nearest other-class sample (dashed circle); samples inside the open ball form $G_B(x)$. (b) A flipped label near x shrinks granules

Proof By (1), $\mathrm{dist}_B(x, y) = 0$ if and only if $a(x) = a(y)$ for all $a \in B$. By (3) and since the minimum is attained, $\delta_B(x) = 0$ if and only if there is $y \in \overline{[x]}_d$ with $\mathrm{dist}_B(x, y) = 0$. Hence B is not separating if and only if there are $x \in U$ and $y \in \overline{[x]}_d$, that is, $d(y) \neq d(x)$, with $a(x) = a(y)$ for all $a \in B$, which is the negation of (6). □

Proposition 1 (Homogeneity and bounds) *Let $B \subseteq C$ be separating. Then for every $x \in U$,*

$$x \in G_B(x) \subseteq [x]_d, \qquad 1 \leq |G_B(x)| \leq |[x]_d|, \tag{7}$$

and

$$n \leq E(B) \leq K_d, \qquad 0 < \gamma(B) \leq 1. \tag{8}$$

Proof Fix $x \in U$. Since $\mathrm{dist}_B(x, x) = 0 < \delta_B(x)$, (4) gives $x \in G_B(x)$. For $y \in \overline{[x]}_d$, (3) gives $\mathrm{dist}_B(x, y) \geq \delta_B(x)$, so $y \notin G_B(x)$ by (4). Hence

$$G_B(x) \subseteq U \setminus \overline{[x]}_d = [x]_d,$$

and $\{x\} \subseteq G_B(x) \subseteq [x]_d$ gives $1 \leq |G_B(x)| \leq |[x]_d|$. As ϕ is strictly increasing, $1 = \phi(1) \leq \phi(|G_B(x)|) \leq \phi(|[x]_d|)$; summing over $x \in U$ gives $n \leq E(B) \leq K_d$, and dividing by $K_d \geq n > 0$ gives $0 < \gamma(B) \leq 1$. □

Remark 1 (Non-separating subsets) If B is not separating, some x has $\delta_B(x) = 0$, and then $G_B(x) = \emptyset$ by (4) because no distance is negative. The bounds (7) therefore require Definition 5. By Lemma 1, a single attribute is separating if and only if no two samples of different classes share its value. Ties between samples of different classes on a single attribute are not rare in measured data. In 9 of the 17 datasets of Section 4, more than two thirds of

the attributes have such a tie, and on the Prostate data a median of 63 percent of the samples have an empty one-attribute granule. A few attributes together break these ties, and the MHMG subsets were separating in all 85 training folds of the first repetition (no radius equal to zero). On fully discretised or binary data the ties may persist for every subset, granules stay empty, and the search may stop at $\emptyset$; such datasets were excluded.

3 Fixed-Radius and Nearest-Enemy Criteria

This section contrasts the criteria that compare distances with a global radius and the MHMG criterion, which has no distance radius. NRS and FNRS see the data only through the safety radii (Proposition 2). Because radii can only grow as attributes are added, both criteria have an absorbing state, which is reached as soon as every radius passes the threshold (Proposition 3). MHMG compares distances only with each other (Proposition 5); its only absorbing state requires every granule to fill its class, and its energy bounds the leave-one-out nearest-neighbour accuracy (Theorem 1). Proposition 6 gives a limit that all three criteria share. All results of this section use the Euclidean distance (1) on attributes scaled to $[0, 1]$. The proofs rest on the definitions and Assumption 1 of Section 2. Section 4 measures the saturation, the failed start of the search and the change of the radii on the benchmark data.

3.1 Neighbourhood dependency through safety radii

Proposition 2 (NRS and FNRS in terms of safety radii) *For every $\varepsilon > 0$, $B \subseteq C$ and $x \in U$,*

$$x \in \mathrm{POS}^{\varepsilon}(B) \iff \delta_B(x) > \varepsilon,$$
$$\gamma_{\mathrm{F}}^{\varepsilon}(B) = \frac{1}{n} \sum_{x \in U} \min\Big(1, \frac{\delta_B(x)}{\varepsilon}\Big). \tag{9}$$

Proof Since U is the disjoint union of $[x]_d$ and $\overline{[x]}_d$, we have $N_B^{\varepsilon}(x) \subseteq [x]_d$ if and only if $N_B^{\varepsilon}(x) \cap \overline{[x]}_d = \emptyset$, that is, $\mathrm{dist}_B(x, y) > \varepsilon$ for every $y \in \overline{[x]}_d$. As $\overline{[x]}_d$ is finite and non-empty, this holds if and only if $\min_{y \in \overline{[x]}_d} \mathrm{dist}_B(x, y) > \varepsilon$, which is $\delta_B(x) > \varepsilon$ by (3). For the second identity, $t \mapsto \min(1, t/\varepsilon)$ is non-decreasing on $[0, \infty)$, so the minimum over $\overline{[x]}_d$ of $\min\big(1, \mathrm{dist}_B(x, y)/\varepsilon\big)$ is attained at a nearest enemy and equals $\min\big(1, \delta_B(x)/\varepsilon\big)$; averaging over x gives (9). □

The first part of (9) is closely related to a recent result of Yang et al. [15]. They showed that the lower and upper approximations of NRS are determined by the pairs of samples of different classes that lie within distance ε of each other, and used this to accelerate the reduction. Proposition 2 states the same fact for each sample through its safety radius and extends it to FNRS. This form leads to the saturation result below.

Thus NRS counts the safety radii above ε, FNRS averages them up to ε, and MHMG counts the samples inside them.

3.2 Saturation of NRS

Proposition 3 (Saturation) *Let $\varepsilon > 0$ and $B \subseteq B' \subseteq C$. Then $\delta_B(x) \le \delta_{B'}(x)$ for every $x \in U$. If $\min_{x \in U} \delta_B(x) > \varepsilon$, then*

$$\gamma^{\varepsilon}_{\mathrm{N}}(B') = \gamma^{\varepsilon}_{\mathrm{N}}(B) = \gamma^{\varepsilon}_{\mathrm{N}}(C) = 1, \tag{10}$$

and if $\min_{x \in U} \delta_B(x) \ge \varepsilon$, then $\gamma^{\varepsilon}_{\mathrm{F}}(B') = \gamma^{\varepsilon}_{\mathrm{F}}(B) = \gamma^{\varepsilon}_{\mathrm{F}}(C) = 1$. A forward search that has reached such a B cannot increase the dependency by adding attributes.

Proof For $y \in U$, (1) gives $\mathrm{dist}_B(x,y)^2 = \sum_{a \in B}(a(x) - a(y))^2 \le \sum_{a \in B'}(a(x) - a(y))^2 = \mathrm{dist}_{B'}(x,y)^2$, because $B \subseteq B'$ and every term is non-negative. Taking the minimum over $y \in [x]_d$ in (3) gives $\delta_B(x) \le \delta_{B'}(x)$. If $\min_x \delta_B(x) > \varepsilon$, then $\delta_{B''}(x) \ge \delta_B(x) > \varepsilon$ for every x and every $B'' \supseteq B$, so $x \in \mathrm{POS}^{\varepsilon}(B'')$ by Proposition 2, and $\gamma^{\varepsilon}_{\mathrm{N}}(B'') = 1$. Taking $B'' = B$, B' and C gives (10). In the same way, $\min_x \delta_B(x) \ge \varepsilon$ gives $\min\bigl(1, \delta_{B''}(x)/\varepsilon\bigr) = 1$ for every x and every $B'' \supseteq B$, and Proposition 2 gives $\gamma^{\varepsilon}_{\mathrm{F}}(B'') = 1$. □

The proposition does not say when saturation occurs, but it shows that it is absorbing. Once every radius passes ε, no attribute can improve the criterion. Squared distances are sums over attributes, so when many attributes are available a few of them may be enough to push every radius past a fixed ε; Section 4.2 measures how early this happened on the benchmark data. A larger ε delays saturation, but the same ε must then serve subsets of very different sizes. A large radius fails at the other end of the search (Proposition 4). Saturation is therefore a property of the dependency as much as of the radius. NRS counts the samples whose radius exceeds ε, so it becomes flat as soon as all radii pass ε. FNRS uses a global radius too, but it averages the truncated radii and stays informative until the smallest radius reaches ε. The FNRS of our experiments uses a radius of 0.3 times the median pairwise distance on all attributes, which is large for small subsets and therefore saturates late.

Proposition 4 (Search that cannot start) *Let every attribute take values in $[0,1]$ on U, let dist_B be the Euclidean distance (1), let $\varepsilon > 0$ and let $B \subseteq C$ with $|B| \le \varepsilon^2$. Then*

$$\mathrm{POS}^{\varepsilon}(B) = \emptyset, \qquad \gamma^{\varepsilon}_{\mathrm{N}}(B) = 0. \tag{11}$$

In particular, if $\varepsilon \ge 1$, then $\gamma^{\varepsilon}_{\mathrm{N}}(\{a\}) = \gamma^{\varepsilon}_{\mathrm{N}}(\emptyset) = 0$ for every $a \in C$, and a forward search that adds an attribute only when $\gamma^{\varepsilon}_{\mathrm{N}}$ strictly increases stops at $\emptyset$, whatever the value of $\gamma^{\varepsilon}_{\mathrm{N}}(C)$.

Proof For $x, y \in U$, (1) and $|a(x) - a(y)| \le 1$ give $\mathrm{dist}_B(x,y) \le |B|^{1/2} \le \varepsilon$. Fix $x \in U$. By Assumption 1 there is $y \in [x]_d$, and $y \in N^{\varepsilon}_B(x)$, so $N^{\varepsilon}_B(x) \not\subseteq [x]_d$ and $x \notin \mathrm{POS}^{\varepsilon}(B)$. Hence $\mathrm{POS}^{\varepsilon}(B) = \emptyset$. If $\varepsilon \ge 1$, both $\emptyset$ and every $\{a\}$ satisfy $|B| \le 1 \le \varepsilon^2$, so the first step of the search finds no attribute with $\gamma^{\varepsilon}_{\mathrm{N}}(\{a\}) > \gamma^{\varepsilon}_{\mathrm{N}}(\emptyset)$ and the search stops. □

Proposition 4 is the mirror image of Proposition 3. A small radius makes the dependency flat at 1 after a few attributes, and a large radius makes it flat at 0 for every small subset, so a greedy forward search cannot leave $\emptyset$ even when $\gamma^{\varepsilon}_{\mathrm{N}}(C) = 1$. The failure at a large radius is therefore a property of the greedy search on the counting dependency, not of the NRS model itself.

Table 3 Decision table of Example 1.

	x_1	x_2	x_3	x_4	x_5	x_6
a_1	0.3	0.7	0.2	0.9	1.0	0.0
a_2	0.1	0.3	0.3	0.4	0.0	1.0
d	1	1	1	2	2	2

The bound $|B| \leq \varepsilon^2$ depends on the metric. It holds for the Euclidean distance (1) on attributes scaled to $[0, 1]$. Under the Chebyshev distance, $\mathrm{dist}_B(x, y) \leq 1$ for every B, so $\varepsilon \geq 1$ empties the positive region of every subset, including C, while a smaller ε gives no bound on $|B|$. For other metrics or scalings the bound must be derived again.

Propositions 3 and 4 concern the NRS dependency, which counts the samples whose neighbourhood is pure, and the simplified FNRS of Proposition 2. They do not cover every criterion with a global radius. NDER [22] uses the class counts inside each neighbourhood, neighbourhood entropy [20, 21] uses the sizes of the neighbourhoods and their overlap with the classes, and FSSNC [16] uses an error measure on a stripped neighbourhood cover. These quantities depend on more than the safety radii, so the two propositions do not apply to them, and we did not test these criteria.

Example 1 Table 3 lists six samples with two attributes and two classes, and let $\varepsilon = 0.15$. On $\{a_1\}$ the safety radii are $0.3, 0.2, 0.2, 0.2, 0.3, 0.2$. All of them exceed ε, so $\gamma^{\varepsilon}_{\mathrm{N}}(\{a_1\}) = \gamma^{\varepsilon}_{\mathrm{N}}(C) = 1$ and NRS returns $\{a_1\}$. By Proposition 3, adding a_2 cannot change this value. The granules show what NRS misses. With $K_d = 12$, MHMG first selects a_1, because $\gamma(\{a_1\}) = 0.69$ and $\gamma(\{a_2\}) = 0.65$ after rounding. The granule sizes on $\{a_1\}$ are $2, 1, 2, 2, 2, 1$. Adding a_2 changes them to $3, 1, 3, 1, 2, 1$: the granules of x_1 and x_3 now contain their whole class, and $\gamma(\{a_1, a_2\}) = 0.72$. MHMG therefore adds a_2, an improvement that the saturated NRS criterion cannot register.

Remark 2 (Models with an adaptive neighbourhood) KNN-NRS, WKNRS, GBNRS and adaptive NRS define the neighbourhood by k nearest samples, granular balls or a data-driven radius, so they are not functions of the safety radii alone and Proposition 3 does not apply to them. They share its mechanism in one respect, since each counts the samples whose neighbourhood passes a purity or similarity test, and the searches of KNN-NRS, GBNRS and adaptive NRS stop when this count reaches its value on the full attribute set, while WKNRS stops when the count no longer increases. Whether that happens after a few attributes when $m \gg n$ is an empirical question, which Section 4.2 answers.

3.3 Ordinal invariance of MHMG

Proposition 5 (Invariance) *Let $g : [0, \infty) \to [0, \infty)$ be strictly increasing, and let $\tilde{\delta}_B$, $\tilde{G}_B$ and $\tilde{E}$ be defined as in Definitions 3 and 4 with $g \circ \mathrm{dist}_B$ in place of dist_B. Then $\tilde{\delta}_B(x) = g(\delta_B(x))$, $\tilde{G}_B(x) = G_B(x)$ for every $x \in U$, and $\tilde{E}(B) = E(B)$. The NRS dependency is not invariant in this sense: with $g \circ \mathrm{dist}_B$ and the same ε, $x \in \mathrm{POS}^{\varepsilon}$ if and only if $\delta_B(x) > g^{-1}(\varepsilon)$ whenever ε lies in the range of g.*

Proof Fix x. As $\overline{[x]}_d$ is finite and g is increasing, $\min_{y\in\overline{[x]}_d} g(\mathrm{dist}_B(x,y)) = g\big(\min_{y\in\overline{[x]}_d} \mathrm{dist}_B(x,y)\big)$, that is, $\tilde{\delta}_B(x) = g(\delta_B(x))$. As g is strictly increasing, $g(\mathrm{dist}_B(x,y)) < g(\delta_B(x))$ if and only if $\mathrm{dist}_B(x,y) < \delta_B(x)$, so $\tilde{G}_B(x) = G_B(x)$ by (4), and $\tilde{E}(B) = E(B)$ by (5). For NRS, Proposition 2 applied to $g \circ \mathrm{dist}_B$ gives $x \in \mathrm{POS}^{\varepsilon}$ if and only if $g(\delta_B(x)) > \varepsilon$, which is $\delta_B(x) > g^{-1}(\varepsilon)$ when $\varepsilon = g(t)$ for some t. □

MHMG therefore uses only the order of the distances from each sample and has no threshold on their scale. Its energy can still grow after every radius has passed any fixed value, as long as the granules gain members. For a separating B, Proposition 1 and the strict increase of ϕ show that its maximum $\gamma(B) = 1$ is reached only when $G_B(x) = [x]_d$ for every x. This requires every sample of a class to be closer to x than the nearest enemy of x, which is much stronger than $\delta_B(x) > \varepsilon$ for a fixed ε. This does not mean that the MHMG search never stops early. Algorithm 1 stops when γ, rounded to two decimals, no longer increases, and this plateau plays the role of a threshold on the gain. Section 5 reports how the rounding precision changes the result.

3.4 Granules and nearest-neighbour classification

Definition 6 (Leave-one-out k-NN rule) Let $B \subseteq C$, $x \in U$ and $1 \le k \le n-1$. A set $V \subseteq U \setminus \{x\}$ is a *k-NN set* of x if $|V| = k$ and

$$\mathrm{dist}_B(x,y) \le \mathrm{dist}_B(x,z) \qquad \forall y \in V,\ \forall z \in U \setminus (\{x\} \cup V). \tag{12}$$

Given a k-NN set V, the rule assigns to x a class c that maximises $|\{y \in V : d(y) = c\}|$, ties between classes being broken in any way. We say that the rule *classifies x correctly* if, for every k-NN set of x and every tie-breaking, it assigns $d(x)$, and we write

$$\mathrm{acc}_k(B) = \frac{|\mathrm{Cor}_k(B)|}{n}, \tag{13}$$

$$A_k(B) = \{x \in U : |G_B(x)| \ge k+1\}, \tag{14}$$

where $\mathrm{Cor}_k(B)$ is the set of samples that the rule classifies correctly.

Theorem 1 (Granule size and k-NN accuracy) *Let $B \subseteq C$ be separating and $1 \le k \le n-1$. Then:*

(i) for every $x \in A_k(B)$ and every k-NN set V of x,

$$V \subseteq G_B(x) \setminus \{x\} \subseteq [x]_d; \tag{15}$$

(ii)

$$\mathrm{acc}_k(B) \ge \frac{|A_k(B)|}{n}; \tag{16}$$

(iii) if $n_{\max} > k$, then

$$\frac{|A_k(B)|}{n} \ge \frac{E(B)/n - \phi(k)}{\phi(n_{\max}) - \phi(k)}; \tag{17}$$

(iv) if $\mathrm{dist}_B(x,y) \ne \delta_B(x)$ for all $x \in U$ and $y \in [x]_d \setminus \{x\}$, then

$$\mathrm{acc}_1(B) = \frac{|A_1(B)|}{n}. \tag{18}$$

Proof (i) Let $x \in A_k(B)$, let V be a k-NN set of x, and put $W = G_B(x) \setminus \{x\}$. By Proposition 1, $x \in G_B(x)$, so $|W| = |G_B(x)| - 1 \geq k$, and every $w \in W$ satisfies $\mathrm{dist}_B(x, w) < \delta_B(x)$ by (4). Suppose that some $y \in V$ has $\mathrm{dist}_B(x, y) \geq \delta_B(x)$. Then $y \notin W$, so $W \cap V \subseteq V \setminus \{y\}$ and

$$|W \setminus V| = |W| - |W \cap V| \geq k - (k-1) = 1.$$

Take $w \in W \setminus V$. Then $w \in U \setminus (\{x\} \cup V)$ and $\mathrm{dist}_B(x, w) < \delta_B(x) \leq \mathrm{dist}_B(x, y)$, which contradicts (12). Hence every $y \in V$ satisfies $\mathrm{dist}_B(x, y) < \delta_B(x)$, that is, $V \subseteq G_B(x) \setminus \{x\}$, and $G_B(x) \subseteq [x]_d$ by (7).

(ii) Let $x \in A_k(B)$ and let V be any k-NN set of x. By (i), $|\{y \in V : d(y) = d(x)\}| = k$ and $|\{y \in V : d(y) = c\}| = 0$ for every class $c \neq d(x)$, so $d(x)$ is the unique maximiser and the rule assigns $d(x)$ whatever the tie-breaking. Hence $A_k(B) \subseteq \mathrm{Cor}_k(B)$, and dividing cardinalities by n gives (16).

(iii) Write $A = A_k(B)$. Since ϕ is increasing, $x \in U \setminus A$ gives $|G_B(x)| \leq k$ and $\phi(|G_B(x)|) \leq \phi(k)$, and $x \in A$ gives $|G_B(x)| \leq |[x]_d| \leq n_{\max}$ by (7) and $\phi(|G_B(x)|) \leq \phi(n_{\max})$. Splitting the sum in (5),

$$\begin{aligned} E(B) &= \sum_{x \in U \setminus A} \phi(|G_B(x)|) + \sum_{x \in A} \phi(|G_B(x)|) \\ &\leq (n - |A|)\, \phi(k) + |A|\, \phi(n_{\max}), \end{aligned}$$

so $|A|\big(\phi(n_{\max}) - \phi(k)\big) \geq E(B) - n\,\phi(k)$. Since $n_{\max} > k$ implies $\phi(n_{\max}) > \phi(k)$, dividing by $n\big(\phi(n_{\max}) - \phi(k)\big) > 0$ yields (17).

(iv) By (ii) with $k = 1$, $A_1(B) \subseteq \mathrm{Cor}_1(B)$. Conversely, let $x \in U \setminus A_1(B)$. Then $|G_B(x)| \leq 1$, and $x \in G_B(x)$ gives $G_B(x) = \{x\}$. For $y \in [x]_d \setminus \{x\}$ we have $y \notin G_B(x)$, so $\mathrm{dist}_B(x, y) \geq \delta_B(x)$ by (4), and the hypothesis gives

$$\mathrm{dist}_B(x, y) > \delta_B(x) \qquad \forall y \in [x]_d \setminus \{x\}. \tag{19}$$

Take $z \in \mathrm{NE}_B(x)$, which exists by Definition 3, and let $V = \{v\}$ be a 1-NN set of x. If $v \in [x]_d$, then $v \neq z$, so $z \in U \setminus (\{x\} \cup V)$ and (12) gives $\mathrm{dist}_B(x, v) \leq \mathrm{dist}_B(x, z) = \delta_B(x)$, contradicting (19). Hence $v \in \overline{[x]}_d$, the rule assigns $d(v) \neq d(x)$, and $x \notin \mathrm{Cor}_1(B)$. Hence $\mathrm{Cor}_1(B) = A_1(B)$, which gives (18). □

The bound (17) is trivial when $E(B)/n < \phi(k)$. Raising the energy thus raises a lower bound on the leave-one-out k-NN accuracy on the training samples. For $k = 1$ and without ties at the radius, part (iv) says that $|A_1(B)|/n$ is one minus the leave-one-out 1-NN error, the complexity measure N3 of Ho and Basu [28]. The guarantee concerns the training samples and does not extend to test samples.

3.5 A limit shared by the family

Cross-validation trains on $U' = U \setminus S$ for a held-out set $S \subset U$. In this subsection the attribute values of the samples in U' are kept fixed, that is, the scaling is not refitted on U'. Then dist_B is the same function on U' as on U. For $x \in U'$, the enemy set of x in U' is $\overline{[x]}_d \setminus S$, and its safety radius in U' is

$$\delta'_B(x) = \begin{cases} \min_{y \in \overline{[x]}_d \setminus S} \mathrm{dist}_B(x, y), & \overline{[x]}_d \setminus S \neq \emptyset, \\ +\infty, & \overline{[x]}_d \setminus S = \emptyset. \end{cases} \tag{20}$$

Proposition 6 (Radius change under removal) *Let $B \subseteq C$, $S \subset U$ and $x \in U \setminus S$. Then*

$$\delta'_B(x) \geq \delta_B(x), \qquad \delta'_B(x) \neq \delta_B(x) \iff \mathrm{NE}_B(x) \subseteq S. \tag{21}$$

Proof If $\overline{[x]}_d \setminus S = \emptyset$, then $\overline{[x]}_d \subseteq S$, hence $\mathrm{NE}_B(x) \subseteq \overline{[x]}_d \subseteq S$, and $\delta'_B(x) = +\infty > \delta_B(x)$ by (20); both parts of (21) hold. Let now $\overline{[x]}_d \setminus S \neq \emptyset$. Since $\overline{[x]}_d \setminus S \subseteq \overline{[x]}_d$,

$$\delta'_B(x) = \min_{y \in \overline{[x]}_d \setminus S} \mathrm{dist}_B(x, y) \geq \min_{y \in \overline{[x]}_d} \mathrm{dist}_B(x, y) = \delta_B(x).$$

Case (a): $\mathrm{NE}_B(x) \not\subseteq S$. Take $z \in \mathrm{NE}_B(x) \setminus S \subseteq \overline{[x]}_d \setminus S$. Then $\delta'_B(x) \leq \mathrm{dist}_B(x, z) = \delta_B(x)$, so $\delta'_B(x) = \delta_B(x)$. *Case (b):* $\mathrm{NE}_B(x) \subseteq S$. Then $\overline{[x]}_d \setminus S \subseteq \overline{[x]}_d \setminus \mathrm{NE}_B(x)$, and every y in this set satisfies $\mathrm{dist}_B(x, y) \geq \delta_B(x)$ and $\mathrm{dist}_B(x, y) \neq \delta_B(x)$. As $\overline{[x]}_d \setminus S$ is finite and non-empty, $\delta'_B(x) = \min_{y \in \overline{[x]}_d \setminus S} \mathrm{dist}_B(x, y) > \delta_B(x)$. Cases (a) and (b) are exhaustive and mutually exclusive, which proves the equivalence in (21). □

Corollary 1 (Expected number of changed radii) *Let $B \subseteq C$ be fixed, let S be drawn uniformly from the subsets of U of size s, $1 \leq s \leq n-1$, and let $t_x = |\mathrm{NE}_B(x)|$ for $x \in U$. Then, for every $x \in U$,*

$$\Pr\big[\delta'_B(x) \neq \delta_B(x) \,\big|\, x \notin S\big] = \prod_{i=0}^{t_x - 1} \frac{s-i}{n-1-i} \leq \frac{s}{n-1}, \tag{22}$$

with equality if and only if $t_x = 1$ or $s = n-1$ (the product is 0 when $t_x > s$). Consequently, the expected number of samples of U' whose radius changes satisfies

$$\mathbb{E}\Big[\big|\{x \in U' : \delta'_B(x) \neq \delta_B(x)\}\big|\Big] \leq \frac{s(n-s)}{n-1}, \tag{23}$$

with equality if and only if $s = n-1$ or every sample has a unique nearest enemy.

Proof Fix $x \in U$ and write $t = t_x$. By Definition 3, $t \geq 1$ and $\mathrm{NE}_B(x) \subseteq \overline{[x]}_d \subseteq U \setminus \{x\}$. Conditional on $x \notin S$, S is uniformly distributed over the s-subsets of $U \setminus \{x\}$, which has $n-1$ elements. If $t > s$, no such subset contains $\mathrm{NE}_B(x)$; otherwise the subsets that contain it are obtained by adding $s-t$ of the remaining $n-1-t$ elements, so

$$\Pr\big[\mathrm{NE}_B(x) \subseteq S \,\big|\, x \notin S\big] = \frac{\binom{n-1-t}{s-t}}{\binom{n-1}{s}} = \prod_{i=0}^{t-1} \frac{s-i}{n-1-i},$$

where the product vanishes when $t > s$. By Proposition 6, $\delta'_B(x) \neq \delta_B(x)$ if and only if $\mathrm{NE}_B(x) \subseteq S$, which gives the equality in (22). If $t > s$, then $t \geq 2$, $s < n-1$ and the probability is $0 < s/(n-1)$. If $t \leq s$, the factor for $i = 0$ is $s/(n-1)$, and every factor for $1 \leq i \leq t-1$ satisfies $0 < (s-i)/(n-1-i) \leq 1$ because $s \leq n-1$, with equality if and only if $s = n-1$. In both cases the probability is at most $s/(n-1)$, with equality if and only if $t = 1$ or $s = n-1$. By linearity of expectation and $\Pr[x \notin S] = (n-s)/n$,

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{x \in U} \mathbf{1}[x \notin S]\, \mathbf{1}[\delta'_B(x) \neq \delta_B(x)]\Big] \\
&\quad = \sum_{x \in U} \Pr[x \notin S]\, \Pr\big[\delta'_B(x) \neq \delta_B(x) \,\big|\, x \notin S\big] \\
&\quad \leq n \cdot \frac{n-s}{n} \cdot \frac{s}{n-1} = \frac{s(n-s)}{n-1},
\end{aligned}$$

with equality if and only if $s = n-1$ or $t_x = 1$ for every $x \in U$. □

Algorithm 1 MHMG forward selection [5]

Input: decision system $\langle U, C \cup \{d\} \rangle$ with C scaled to $[0,1]$; rounding precision of two decimals

Output: subset $R \subseteq C$ at which no single added attribute increases the rounded γ of (5)

$R \leftarrow \emptyset$; $\gamma \leftarrow 0$; $D \leftarrow \mathbf{0}_{n \times n}$ $\{D_{ij} = \mathrm{dist}_R(x_i, x_j)^2\}$
compute K_d by (5)
loop
  $(a^*, \gamma^*) \leftarrow (\mathrm{null}, -1)$
  for all $a \in C \setminus R$ in index order **do**
    $D^a \leftarrow D + \big(a(x_i) - a(x_j)\big)^2_{i,j}$
    $\delta_i^2 \leftarrow \min_{j : d(x_j) \neq d(x_i)} D^a_{ij}$ for all i $\{(3)\}$
    $G_i \leftarrow \{j : D^a_{ij} < \delta_i^2\}$ for all i $\{(4)\}$
    $\tilde{\gamma} \leftarrow \mathrm{round}\big(\sum_i \phi(|G_i|)/K_d,\ 2\big)$ $\{(5)\}$
    if $\tilde{\gamma} > \gamma^*$ **then**
      $(a^*, \gamma^*) \leftarrow (a, \tilde{\gamma})$
    end if
  end for
  if $a^* = \mathrm{null}$ **or** $\gamma^* \leq \gamma$ **then**
    break
  end if
  $R \leftarrow R \cup \{a^*\}$; $\gamma \leftarrow \gamma^*$; $D \leftarrow D^{a^*}$
end loop
return R

With five-fold cross-validation, $s/(n-1) \approx 0.2$: for a subset chosen independently of the folds, at most about a fifth of the safety radii change between the full data and a training fold. A subset selected on the training fold can exceed this bound. Because NRS, FNRS and MHMG all depend on the radii, all three inherit this sensitivity when n is small; Section 4.6 shows its effect.

3.6 Selection algorithm and complexity

Algorithm 1 is the published search [5] written on squared distances, which is how it was run here; Proposition 7 gives its cost. The initial value $\gamma = 0$ equals $\gamma(\emptyset)$, because $\mathrm{dist}_\emptyset \equiv 0$ makes every granule empty (Remark 1).

Proposition 7 (Cost of Algorithm 1) *If Algorithm 1 returns R with $|R| = q$, its running time and working memory (in addition to the $O(nm)$ input) satisfy*

$$T_{\mathrm{MHMG}}(n, m, q) = O\big((q+1)\, m\, n^2\big), \qquad M_{\mathrm{MHMG}}(n) = O(n^2). \tag{24}$$

Table 4 Time complexity of the selectors as implemented (n samples, m attributes, q selected attributes, $h = 10$ hits and misses for ReliefF, $|U/d|$ classes; for adaptive NRS and GBNRS, $m' = 80$ prefiltered attributes and $c(n)$ the cost of one evaluation of the criterion).

Selector	Time
MHMG (Alg. 1)	$O\big((q+1)\,m\,n^2\big)$
NRS, greedy + backward pass	$O\big((q+1)\,m\,n^2 + q^2 n^2\big)$
FNRS, capped at 50 attributes	$O\big((q+1)\,m\,n^2\big)$, $q \le 50$
F	$O(nm)$
FCQ	$O(q\,n\,m)$
ReliefF	$O(n^2 m + n\,\lvert U/d\rvert\,h\,m)$
L1 (liblinear)	$O(I\,n\,m)$ for I solver passes
RFE (linear SVM, 10 % per round)	$O\big(\log_{10/9}(m/q)\big)$ SVM fits
KNN-NRS, WKNRS	$O\big((q+1)\,m\,n^2\big)$
Adaptive NRS, GBNRS	$O\big((q+1)\,m'\,c(n)\big)$

Proof Computing K_d once costs $O(n)$. Let pass $t = 0, 1, \ldots$ of the loop start with $|R| = t$. Pass t visits the $m - t$ attributes of $C \setminus R$, and for each attribute line 6 costs $O(n^2)$, lines 7 and 8 compute n row minima over enemies and n counts over rows of D^a in $O(n^2)$, and line 9 costs $O(n)$; the pass therefore costs at most $c_0\,(m-t)\,n^2$ for a constant c_0. Passes $0, \ldots, q-1$ accept an attribute and pass q, if it is reached, stops the loop, so at most $q+1$ passes run and

$$\begin{aligned} T_{\mathrm{MHMG}}(n, m, q) &\le O(n) + \sum_{t=0}^{q} c_0\,(m-t)\,n^2 \\ &\le O(n) + c_0\,(q+1)\,m\,n^2, \end{aligned}$$

which is $O((q+1)mn^2)$. The algorithm stores D, one matrix D^a at a time and $O(n)$ further values, hence $M_{\mathrm{MHMG}}(n) = O(n^2)$. □

Table 4 compares this cost with the other methods as implemented here. MHMG, NRS and FNRS share the $O((q+1)mn^2)$ form, which is why their times in Table 9 lie within one order of magnitude of each other, whereas the filters are linear in n and the embedded methods depend on the solver. In floating point, a candidate whose rounded $\tilde{\gamma}$ ties with the best is not taken, because the comparison is strict and candidates are visited in index order; this reproduces the published implementation [5], with which our search was checked to coincide.

4 Experiments

This section compares MHMG with the neighbourhood rough-set models and with standard selectors on seventeen gene-expression datasets, and measures what the criterion costs in stability, robustness and time.

4.1 Setup

Datasets

Gene-expression data are the standard benchmark for $m \gg n$. Table 5 lists the seventeen datasets. Ten come from OpenML [36], including the leukemia [37], colon [38],

Table 5 Gene-expression datasets (n: samples, m: attributes, K: classes).

Dataset	n	m	K	Source
Lymphoma	45	4026	2	OpenML
Leukemia	72	7129	2	OpenML
Burkitt	220	22283	3	OpenML
GCM	190	16063	14	OpenML
Tumors-C	60	7129	2	OpenML
RSCTC-1	105	22283	3	OpenML
RSCTC-3	95	22277	5	OpenML
AP-OmU	201	10935	2	OpenML
Colon	62	2000	2	OpenML
Duke	86	7129	2	OpenML
CLL-SUB	111	11340	3	scikit-feature
Glioma	50	4434	4	scikit-feature
GLI-85	85	22283	2	scikit-feature
Prostate	102	5966	2	scikit-feature
SMK-CAN	187	19993	2	scikit-feature
TOX-171	171	5748	4	scikit-feature
Lung	203	3312	5	scikit-feature

lymphoma [39], global cancer map (GCM) [40], central nervous system (CNS) [41] and breast cancer [42] studies and two sets of the Rough Sets and Current Trends in Computing (RSCTC) 2010 challenge [43]. Seven come from the scikit-feature repository [44], including the prostate [45] and lung [46] sets. The selection rule was fixed in advance: discrete labels, at least 2000 attributes, at most 300 samples, one dataset per source, no discretised versions. Constant attributes were removed, and the missing values of the lymphoma set were replaced by attribute means over all its samples. Computing these means within each training fold instead changed the accuracy gap between MHMG and NRS* on this dataset by only $+0.007$.

Protocol

We split each dataset by stratified five-fold cross-validation with three shuffling seeds, which gives fifteen train/test pairs. Attributes were min-max scaled on the training part, and every method saw the training part only [47]. A 3-NN classifier and an SVM with a radial basis function (RBF) kernel and default settings were then evaluated on the held-out part. 3-NN is the main evaluator, because all neighbourhood criteria are distance-based. It is also close to what MHMG optimises (Theorem 1) and may therefore favour it, so the main comparisons are also reported with the SVM. All settings were fixed before the runs. The methods were implemented in Python 3.9 with NumPy and scikit-learn and run on one computer with an Apple M4 Pro processor and 48 GB of memory.

Statistical analysis

Following Demšar [48], we compare several methods over the datasets with the Friedman test [49] and the Nemenyi post-hoc test, shown as critical-difference diagrams. We also compare each competitor with MHMG by a two-sided Wilcoxon signed-rank test [50] on the dataset means, with Holm correction [51] within each table. In the

tables, Δ is the mean 3-NN accuracy of MHMG minus that of the competitor, with a 95 percent bootstrap confidence interval (CI) (10 000 resamples of the datasets) that is not adjusted for multiplicity. W/T/L counts the datasets on which MHMG is more accurate, tied or less accurate. The p-value is the Holm-corrected Wilcoxon p-value unless the caption states otherwise.

Radius-based models

MHMG is Algorithm 1 with the published settings [5]. NRS adds the attribute that most enlarges the positive region and stops when the dependency of the full set is reached or no attribute enlarges it; it then removes attributes that are not needed. In the main comparison, ε is 0.1, 0.2 or 0.3. To examine the role of the radius, we add $\varepsilon = 0.4$, 0.5, 0.75 and 1.0 and a data-scaled radius of 0.3 times the median pairwise distance, which gives eight radii. NRS* takes, for each dataset, the radius with the best test accuracy among the first three, and NRS$^*_8$ among all eight; both favour NRS. NRS-cv and NRS-cv$_8$ tune the radius over the same sets by inner three-fold cross-validation. An empty reduct counts as a failure and is never chosen.

FNRS uses the data-scaled radius and stops at 95 percent of the full dependency. We report it with a cap of 50 attributes (FNRS-50) and stopped at the size of MHMG on the same fold (FNRS-q). A separate comparison stops it at 10, 20 and 50 attributes and continues the MHMG path to the same sizes.

Models that adapt the neighbourhood

KNN-NRS [10] uses the three nearest samples under the Chebyshev distance; we also ran it with the Euclidean distance. WKNRS [11] uses the five nearest samples with $\alpha = 0.01$, and GBNRS [7] uses purity 0.9. Adaptive NRS is a greedy re-implementation of the adaptive-radius criterion of Qiu et al. [13], without their pied kingfisher optimiser. Adaptive NRS and GBNRS were too slow for thousands of candidate attributes, so both ran on the 80 attributes of largest training variance, together with MHMG on the same attributes (MHMG-80). All four models are our re-implementations of the published criteria and stopping rules. Each run had a limit of 30 minutes per fold, which no run reached.

Controls

Two controls separate the effect of the criterion from that of the reduct size. First, we cut the greedy path of MHMG to the size of the NRS reduct on the same fold. Second, we drew ten random subsets of the MHMG size on each fold and averaged their accuracy; on the 80 prefiltered attributes, random subsets of the MHMG-80 size play the same role.

External selectors

Five selectors receive the size q of the MHMG reduct on the same fold: the ANOVA F filter, FCQ (F-test correlation quotient) [30], ReliefF [27], L1-regularised logistic regression (L1) and SVM-RFE [33]. The same five also chose their own size by inner three-fold cross-validation within ten attributes (suffix -cv10), which is the regime of MHMG.

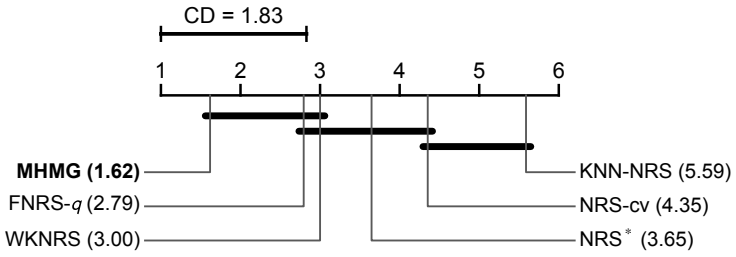


Fig. 2 Critical-difference (CD) diagram of 3-NN accuracy ranks of the six models with small reducts (Nemenyi test, $\alpha = 0.05$); FNRS-50 is excluded. Bars join models that do not differ significantly

4.2 Comparison within the neighbourhood rough-set family

Tables 6 and 7 give the accuracy of every neighbourhood model on every dataset, Table 8 gives the sizes and running times, and Table 9 summarises the paired comparisons with MHMG. NRS, KNN-NRS, WKNRS and adaptive NRS stopped after a few attributes, and MHMG was more accurate than each of them. FNRS and GBNRS did not stop early, but they kept many more attributes, and at the same or a smaller size MHMG was at least as accurate as both.

MHMG kept a median of 7.4 attributes and reached a mean 3-NN accuracy of 0.771. NRS was less accurate for each of the three radii. The gap was largest at $\varepsilon = 0.1$ (+0.106), and it remained +0.054 when the best of these radii was chosen on the test folds; Section 4.3 widens the grid. KNN-NRS returned a single attribute in 254 of 255 folds, because its positive region with three neighbours already reached the value of the full attribute set, and it was +0.146 less accurate. The Chebyshev distance is not the cause: with the Euclidean distance KNN-NRS still kept at most two attributes in 228 of 255 folds. WKNRS was the strongest of the models that stop early. It kept 4.3 attributes and was +0.027 less accurate than MHMG, which was better on 14 of the 17 datasets. It was also the slowest model in the family.

Fig. 2 compares the ranks of the six models that select from all attributes and return small reducts, with a median size of at most 7.4 attributes (Table 9). FNRS-50 is left out of this comparison, because it kept 50 attributes in every fold; it is compared with MHMG at equal sizes below. A Friedman test rejected equal performance ($\chi^2 =$ 45.9, p ¡0.001). MHMG had the best average rank (1.62). The Nemenyi test separated it from NRS*, NRS-cv and KNN-NRS, but not from FNRS-q or WKNRS, which the paired Wilcoxon tests of Table 9 do separate. With the SVM, MHMG and FNRS-q again held the first two ranks.

FNRS with 50 attributes was the most accurate model in Table 6, and it would rank first if it were added to Fig. 2. This came from its size. FNRS never reached its own stopping target and stopped at the cap of 50 in every fold, almost seven times the size of MHMG. At the size of MHMG it was +0.030 less accurate. Under the SVM, however, FNRS-50 was the more accurate of the two, by +0.039 (confidence interval

Table 6 3-NN accuracy (mean±std over 15 folds) of the neighbourhood rough-set models. Bold: best value in each block; ranks are computed within each block.

	All attributes							80 attributes of largest variance		
Dataset	MHMG	NRS*	NRS-cv	FNRS-50	FNRS-q	KNN-NRS	WKNRS	MHMG-80	ANRS-80	GBNRS-80
Lymphoma	**0.919±0.078**	0.822±0.101	0.830±0.102	**0.919±0.089**	0.904±0.102	0.785±0.170	0.859±0.107	**0.874±0.110**	0.526±0.170	0.778±0.173
Leukemia	0.908±0.073	0.870±0.085	0.849±0.072	**0.948±0.082**	0.916±0.085	0.889±0.052	0.857±0.060	**0.903±0.066**	0.622±0.108	0.871±0.097
Burkitt	**0.871±0.039**	0.830±0.048	0.841±0.051	0.826±0.032	0.839±0.059	0.779±0.053	0.833±0.056	**0.856±0.037**	0.591±0.090	0.809±0.050
GCM	0.565±0.064	0.326±0.077	0.328±0.081	0.526±0.049	0.456±0.068	0.284±0.059	**0.596±0.077**	0.428±0.058	0.405±0.066	**0.433±0.079**
Tumors-C	0.578±0.107	0.622±0.099	0.589±0.128	**0.694±0.108**	0.656±0.137	0.600±0.085	0.606±0.153	0.572±0.172	0.522±0.156	**0.622±0.113**
RSCTC-1	0.492±0.101	0.489±0.091	0.460±0.106	**0.559±0.129**	0.489±0.089	0.486±0.065	0.502±0.114	0.514±0.090	0.492±0.061	**0.565±0.057**
RSCTC-3	**0.807±0.093**	0.691±0.082	0.670±0.113	0.695±0.102	0.667±0.157	0.502±0.095	0.765±0.093	**0.607±0.139**	0.312±0.161	0.600±0.071
AP-OmU	0.867±0.054	0.824±0.041	0.827±0.054	**0.904±0.049**	0.834±0.058	0.816±0.111	0.866±0.042	0.813±0.073	0.584±0.059	**0.839±0.060**
Colon	0.812±0.110	0.850±0.098	0.780±0.139	**0.873±0.070**	0.834±0.106	0.721±0.097	0.786±0.106	0.774±0.122	0.591±0.131	**0.797±0.112**
Duke	0.895±0.060	0.891±0.063	0.849±0.106	**0.953±0.086**	0.760±0.095	0.818±0.076	0.895±0.075	**0.891±0.091**	0.631±0.169	0.654±0.159
CLL-SUB	**0.672±0.109**	0.574±0.107	0.604±0.114	0.657±0.091	0.649±0.078	0.504±0.094	0.637±0.071	**0.550±0.119**	0.430±0.107	0.540±0.113
Glioma	**0.660±0.091**	0.600±0.120	0.567±0.140	0.647±0.151	0.640±0.106	0.287±0.083	0.600±0.141	**0.633±0.111**	0.480±0.174	0.607±0.205
GLI-85	0.851±0.094	0.820±0.085	0.839±0.075	**0.875±0.062**	0.847±0.062	0.765±0.097	0.792±0.091	**0.757±0.109**	0.627±0.106	0.718±0.139
Prostate	**0.905±0.065**	0.896±0.074	0.890±0.071	**0.905±0.066**	0.875±0.062	0.690±0.149	0.840±0.086	**0.756±0.077**	0.618±0.106	0.690±0.097
SMK-CAN	0.637±0.081	0.627±0.087	0.620±0.079	**0.669±0.063**	0.626±0.066	0.554±0.081	0.632±0.103	0.592±0.081	0.521±0.089	**0.624±0.081**
TOX-171	0.751±0.080	0.579±0.112	0.577±0.105	**0.811±0.091**	0.704±0.124	0.401±0.113	0.698±0.083	**0.634±0.162**	0.268±0.093	0.619±0.105
Lung	0.916±0.033	0.867±0.055	0.864±0.040	**0.931±0.025**	0.892±0.038	0.737±0.036	0.880±0.041	**0.905±0.041**	0.733±0.078	0.890±0.048
Average	0.771	0.716	0.705	0.788	0.740	0.625	0.744	0.709	0.527	0.686
Average rank	2.26	4.59	5.29	1.71	3.74	6.59	3.82	1.35	3.00	1.65

Table 7 SVM accuracy (mean±std over 15 folds) of the neighbourhood rough-set models. Bold: best value in each block; ranks are computed within each block.

	All attributes							80 attributes of largest variance		
Dataset	MHMG	NRS*	NRS-cv	FNRS-50	FNRS-q	KNN-NRS	WKNRS	MHMG-80	ANRS-80	GBNRS-80
Lymphoma	**0.919±0.098**	0.837±0.125	0.830±0.125	**0.919±0.089**	0.889±0.111	0.807±0.142	0.844±0.092	**0.881±0.098**	0.548±0.176	0.844±0.125
Leukemia	0.912±0.072	0.866±0.090	0.871±0.067	**0.977±0.034**	0.902±0.079	0.879±0.055	0.871±0.061	**0.885±0.065**	0.657±0.020	0.875±0.130
Burkitt	**0.897±0.053**	0.850±0.043	0.855±0.040	**0.897±0.034**	0.862±0.043	0.829±0.052	0.839±0.070	**0.874±0.043**	0.621±0.081	0.858±0.053
GCM	0.511±0.057	0.432±0.053	0.407±0.064	**0.582±0.066**	0.507±0.066	0.293±0.037	0.468±0.069	**0.453±0.045**	0.414±0.062	0.435±0.060
Tumors-C	0.650±0.114	0.633±0.121	0.611±0.112	**0.739±0.094**	0.683±0.127	0.656±0.109	0.644±0.092	0.644±0.080	**0.650±0.047**	0.628±0.043
RSCTC-1	0.524±0.137	0.527±0.071	0.511±0.112	**0.603±0.118**	0.517±0.082	0.521±0.079	0.527±0.079	0.546±0.067	0.556±0.067	**0.581±0.090**
RSCTC-3	**0.814±0.069**	0.740±0.054	0.712±0.122	0.768±0.121	0.723±0.104	0.558±0.089	0.779±0.087	0.635±0.127	0.330±0.145	**0.646±0.141**
AP-OmU	0.872±0.059	0.859±0.053	0.851±0.060	**0.907±0.044**	0.851±0.060	0.837±0.073	0.846±0.043	**0.842±0.064**	0.612±0.028	**0.842±0.077**
Colon	0.824±0.115	0.834±0.090	0.764±0.119	**0.883±0.057**	0.856±0.091	0.725±0.099	0.775±0.096	0.800±0.096	0.663±0.075	**0.862±0.097**
Duke	0.837±0.082	0.899±0.075	0.868±0.105	**0.969±0.053**	0.780±0.052	0.794±0.109	0.895±0.092	**0.895±0.108**	0.607±0.135	0.673±0.213
CLL-SUB	**0.714±0.096**	0.577±0.099	0.622±0.098	0.645±0.111	0.667±0.092	0.550±0.090	0.661±0.102	**0.595±0.094**	0.445±0.114	0.541±0.055
Glioma	0.613±0.119	0.627±0.167	0.647±0.141	**0.720±0.121**	0.667±0.111	0.353±0.106	0.620±0.182	**0.707±0.128**	0.520±0.121	0.627±0.194
GLI-85	0.859±0.088	0.824±0.094	0.839±0.087	**0.894±0.055**	0.875±0.073	0.792±0.099	0.796±0.077	0.753±0.074	0.690±0.041	**0.761±0.075**
Prostate	0.915±0.060	**0.919±0.076**	0.899±0.078	0.918±0.060	0.898±0.057	0.740±0.106	0.834±0.104	**0.788±0.105**	0.566±0.100	0.716±0.107
SMK-CAN	0.651±0.104	0.666±0.090	0.675±0.080	**0.693±0.076**	0.635±0.052	0.581±0.103	0.642±0.104	**0.679±0.069**	0.583±0.088	0.665±0.109
TOX-171	0.747±0.087	0.614±0.104	0.618±0.099	**0.805±0.089**	0.718±0.088	0.460±0.089	0.708±0.099	0.630±0.145	0.275±0.096	**0.698±0.129**
Lung	0.926±0.037	0.872±0.053	0.879±0.029	**0.928±0.029**	0.900±0.037	0.752±0.033	0.897±0.040	**0.911±0.032**	0.764±0.067	0.893±0.043
Average	0.775	0.740	0.733	0.815	0.761	0.655	0.744	0.736	0.559	0.714
Average rank	2.82	4.32	4.94	1.41	3.62	6.41	4.47	1.44	2.82	1.74

Table 8 Mean number of selected attributes and, in parentheses, mean running time in seconds per fold. Settings as in Table 6.

	All attributes							80 attributes of largest variance		
Dataset	MHMG	NRS*	NRS-cv	FNRS-50	FNRS-q	KNN-NRS	WKNRS	MHMG-80	ANRS-80	GBNRS-80
Lymphoma	6.9 (1.2)	3.7 (0.1)	3.3 (0.5)	50.0 (1.8)	6.9 (0.3)	1.0 (0.7)	2.6 (11.0)	3.5 (<0.1)	2.2 (0.2)	76.2 (23.1)
Leukemia	3.3 (1.4)	2.4 (0.2)	3.9 (1.4)	50.0 (5.1)	3.3 (0.4)	1.0 (2.0)	3.0 (33.7)	3.7 (<0.1)	2.9 (0.4)	53.9 (18.8)
Burkitt	11.4 (41.3)	6.1 (6.6)	6.3 (35.2)	50.0 (81.6)	11.4 (18.9)	1.0 (21.9)	5.5 (556.1)	6.6 (0.1)	4.8 (2.1)	62.5 (71.0)
GCM	11.3 (42.1)	15.1 (12.2)	12.9 (44.6)	50.0 (45.6)	11.3 (10.6)	1.0 (13.2)	15.2 (862.2)	6.1 (0.1)	9.1 (3.6)	45.1 (93.4)
Tumors-C	4.7 (1.6)	6.6 (0.3)	5.8 (1.6)	50.0 (4.1)	4.7 (0.4)	1.0 (1.6)	2.9 (30.5)	2.4 (<0.1)	2.6 (0.3)	73.9 (41.8)
RSCTC-1	5.4 (9.4)	8.4 (2.8)	5.9 (10.2)	50.0 (26.3)	5.4 (3.0)	1.0 (9.2)	3.2 (185.6)	2.9 (<0.1)	3.0 (0.6)	58.4 (66.0)
RSCTC-3	12.7 (24.1)	6.5 (2.0)	5.7 (8.2)	50.0 (23.1)	12.7 (5.9)	1.0 (8.3)	8.1 (338.7)	4.5 (<0.1)	4.1 (0.7)	60.3 (55.4)
AP-OmU	8.2 (11.9)	9.9 (4.7)	7.7 (17.5)	50.0 (34.9)	8.2 (5.9)	1.0 (9.7)	6.4 (270.4)	2.5 (<0.1)	3.3 (1.2)	53.9 (55.3)
Colon	6.5 (0.6)	4.7 (0.1)	3.7 (0.3)	50.0 (1.2)	6.5 (0.2)	1.0 (0.5)	3.1 (8.4)	4.9 (<0.1)	2.4 (0.3)	73.1 (24.3)
Duke	1.6 (0.9)	5.0 (0.4)	4.4 (1.8)	50.0 (6.5)	1.6 (0.3)	1.0 (2.4)	2.8 (40.0)	3.6 (<0.1)	2.0 (0.3)	16.7 (7.8)
CLL-SUB	7.8 (7.0)	8.5 (1.6)	6.8 (5.9)	50.0 (14.9)	7.8 (2.4)	1.0 (5.0)	6.2 (163.9)	3.7 (<0.1)	3.8 (0.7)	54.1 (63.3)
Glioma	7.4 (2.0)	4.1 (0.1)	4.5 (0.8)	50.0 (2.2)	7.4 (0.3)	1.0 (0.8)	4.3 (21.5)	5.3 (<0.1)	4.0 (0.4)	43.3 (22.5)
GLI-85	6.4 (8.0)	3.9 (1.0)	4.1 (5.3)	50.0 (19.9)	6.4 (2.6)	1.0 (7.4)	3.6 (144.3)	2.5 (<0.1)	2.9 (0.5)	61.0 (33.0)
Prostate	7.4 (2.8)	2.9 (0.3)	3.5 (1.8)	50.0 (7.0)	7.4 (1.1)	1.0 (2.3)	3.7 (46.4)	4.4 (<0.1)	5.3 (1.0)	47.7 (43.1)
SMK-CAN	6.1 (15.3)	9.3 (6.6)	7.4 (26.5)	50.0 (55.9)	6.1 (7.2)	1.0 (16.3)	4.5 (380.4)	3.7 (<0.1)	3.5 (1.2)	47.7 (83.9)
TOX-171	10.5 (8.3)	9.7 (1.8)	8.7 (7.1)	50.0 (14.0)	10.5 (3.0)	1.1 (4.5)	9.5 (193.1)	6.7 (0.1)	3.3 (1.1)	69.2 (111.7)
Lung	10.7 (5.5)	5.6 (0.8)	6.4 (4.3)	50.0 (10.7)	10.7 (2.3)	1.0 (2.9)	6.5 (81.8)	8.9 (0.1)	6.5 (2.5)	46.1 (36.2)
Average	7.5 (10.8)	6.6 (2.4)	5.9 (10.2)	50.0 (20.9)	7.5 (3.8)	1.0 (6.4)	5.4 (198.1)	4.5 (<0.1)	3.9 (1.0)	55.5 (50.0)

Table 9 Summary of the comparison within the neighbourhood rough-set family (17 datasets). Size: median over the datasets of the mean size; time: median seconds per fold; D: datasets with all 15 folds. NRS*: best of the three radii on the test folds (Section 4.3 gives the eight-radius version). In the lower block, Δ is relative to MHMG-80.

Model	Size	3-NN	SVM	Time	Δ [95% CI]	W/T/L	p	D
MHMG	7.4	0.771	0.775	7.0	–	–	–	17
NRS ($\varepsilon = 0.1$)	3.8	0.665	0.697	0.8	+0.106 [+0.063, +0.151]	16/0/1	<0.001	17
NRS ($\varepsilon = 0.2$)	5.1	0.684	0.717	1.0	+0.087 [+0.055, +0.125]	17/0/0	<0.001	17
NRS ($\varepsilon = 0.3$)	6.6	0.704	0.730	1.3	+0.067 [+0.034, +0.101]	15/0/2	0.003	17
NRS* (best ε on test)	6.1	0.716	0.740	1.0	+0.054 [+0.023, +0.090]	15/0/2	0.013	17
NRS-cv	5.8	0.705	0.733	5.3	+0.066 [+0.040, +0.098]	16/0/1	<0.001	17
FNRS (50 attributes)	50.0	0.788	0.815	14.0	−0.017 [−0.040, +0.009]	6/1/10	0.156	17
FNRS at the MHMG size	7.4	0.740	0.761	2.4	+0.030 [+0.006, +0.056]	14/0/3	0.022	17
KNN-NRS	1.0	0.625	0.655	4.5	+0.146 [+0.092, +0.203]	16/0/1	<0.001	17
WKNRS	4.3	0.744	0.744	144.3	+0.027 [+0.012, +0.041]	14/0/3	0.013	17
MHMG-80	3.7	0.709	0.736	0.0	–	–	–	17
Adaptive NRS-80	3.3	0.527	0.559	0.7	+0.183 [+0.135, +0.233]	17/0/0	<0.001	17
GBNRS-80	54.1	0.686	0.714	43.1	+0.024 [−0.004, +0.059]	11/0/6	0.225	17

Table 10 MHMG and FNRS continued to the same number of attributes (mean accuracy over 17 datasets; Δ and W/L for 3-NN).

	3-NN				SVM	
Size	MHMG	FNRS	Δ [95% CI]	W/L	MHMG	FNRS
10	0.775	0.752	+0.023 [−0.002, +0.047]	13/4	0.782	0.780
20	0.787	0.763	+0.024 [−0.006, +0.055]	11/5	0.793	0.798
50	0.799	0.788	+0.011 [−0.015, +0.043]	7/9	0.814	0.815

+0.014 to +0.063). It was better on 13 of the 17 datasets ($p = 0.009$, uncorrected), and its accuracy was even above that of the full attribute set (Table 7). Table 10 continues both methods to 10, 20 and 50 attributes. With 3-NN, MHMG was at least as accurate on average at every size, and the lower confidence limit of the difference stayed above −0.02. At 50 attributes the means were 0.799 for MHMG and 0.788 for FNRS, although FNRS was better on 9 of the 17 datasets. With the SVM the mean difference was below 0.01 at every size (0.814 against 0.815 at 50 attributes).

Adaptive NRS and GBNRS were compared with MHMG on the 80 attributes of largest variance. Adaptive NRS stopped after 3.3 attributes and was +0.183 less accurate than MHMG on the same attributes; MHMG was better on 17 of the 17 datasets. It was also less accurate than random subsets of the same 80 attributes (−0.049, confidence interval −0.070 to −0.028; better on only 2 datasets). Our greedy re-implementation may therefore understate the original method, which searches with a swarm optimiser. GBNRS behaved in the opposite way. Following its reference loop, it kept adding attributes until its dependency reached the value of all 80, and it kept a median of 79 of them per fold, so it hardly reduced the set. MHMG reached the same accuracy with about four attributes (+0.024).

The early stop of NRS is the saturation of Proposition 3. When the full attribute set puts every sample in the positive region, the stopping rule of NRS ends the search exactly when every safety radius exceeds ε. In all 85 training folds of the first repetition, this happened after a median of 4 to 7 attributes, depending on the radius. How early this happens does not depend on well-chosen attributes. Random attribute subsets

Table 11 NRS over a wide grid of radii (17 datasets, 255 folds). Empty: folds with an empty reduct; size: mean number of attributes of the non-empty reducts; D: datasets with a non-empty reduct in all 15 folds; 3-NN: mean over these datasets. †Non-empty reducts at $\varepsilon = 1.0$ arise only from rounding errors of the scaled values.

ε	Empty	Size	D	3-NN
0.1	0	3.7	17	0.665
0.2	0	5.4	17	0.684
0.3	0	7.4	17	0.704
0.4	0	9.9	17	0.703
0.5	0	12.8	17	0.712
0.75	28	21.5	13	0.708
1.0†	226	12.5	1	–
data-scaled	255	–	0	–
NRS_8^* (best on test)	–	–	17	0.743
$\mathrm{NRS\text{-}cv}_8$	–	–	17	0.709
MHMG	0	7.5	17	0.771

saturated NRS after a median of 6 attributes at $\varepsilon = 0.1$ and 14 at $\varepsilon = 0.3$, and this number showed no clear relation to the number of attributes of the dataset (Spearman $\rho = 0.31$, $p = 0.22$). When $m \gg n$, almost any handful of attributes pushes every radius past a fixed threshold. MHMG has no threshold to pass (Proposition 5). In the median fold, at least 70 percent of the radii of the MHMG reducts exceeded each of the radii 0.1, 0.2 and 0.3. The energy still grew after this point, because it counts the samples inside the radii instead of the radii above a value.

4.3 Sensitivity to the radius

MHMG has no radius to tune, so the sensitivity question concerns NRS. Table 11 and Fig. 3 show NRS over the grid of eight radii. A larger radius delayed saturation, and NRS kept more attributes, from a mean of 3.7 at $\varepsilon = 0.1$ to 12.8 at $\varepsilon = 0.5$. Its accuracy rose only to 0.712, below the 0.771 of MHMG. At every radius up to 0.75, MHMG had the higher mean accuracy on the datasets where NRS returned a reduct in all folds.

Beyond $\varepsilon = 0.5$ the greedy search began to fail at the start. NRS returned an empty reduct in 28 of 255 folds at $\varepsilon = 0.75$, in 226 at $\varepsilon = 1.0$, and in every fold with the data-scaled radius. The full attribute set was not the cause: it put every training sample in the positive region at $\varepsilon = 0.75$ and $\varepsilon = 1.0$, and at least 0.91 of them with the data-scaled radius. In every empty fold, no single attribute could raise the dependency above zero, so the search never took its first step; for $\varepsilon \geq 1$ this is what Proposition 4 predicts. The few non-empty reducts at $\varepsilon = 1.0$ come from rounding errors of the scaled values. A small radius therefore saturates after a few attributes, and a large one gives the search no first step.

Choosing the best of the eight radii on the test folds narrowed the gap to +0.028 (confidence interval +0.002 to +0.054), with MHMG better on 12 of the 17 datasets. This margin is small, and the Wilcoxon test did not reject equality ($p = 0.098$). The chosen radius was 0.4 or larger on 10 datasets, where NRS also kept more attributes

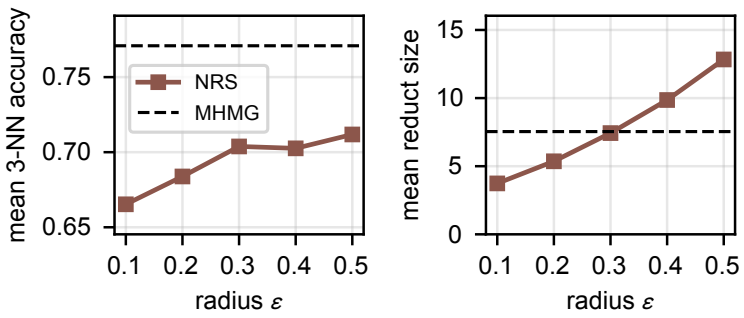


Fig. 3 Mean 3-NN accuracy (left) and reduct size (right) of NRS for the radii at which it returned a reduct on every fold. Dashed line: MHMG, which has no radius

Table 12 Control comparisons (3-NN, 17 datasets). Δ: MHMG minus the control; equal size: MHMG cut to the size of the NRS reduct; p: Holm-corrected over the first six rows, uncorrected in the last two.

Control	Δ [95% CI]	W/T/L	p
NRS_8^* (best of 8 radii)	+0.028 [+0.002, +0.054]	12/0/5	0.098
NRS-cv_8	+0.062 [+0.037, +0.087]	15/0/2	<0.001
$NRS_{0.1}$, equal size	+0.063 [+0.029, +0.098]	15/0/2	0.003
$NRS_{0.2}$, equal size	+0.077 [+0.050, +0.109]	17/0/0	<0.001
$NRS_{0.3}$, equal size	+0.063 [+0.033, +0.097]	13/0/4	0.003
Random, MHMG size	+0.196 [+0.138, +0.255]	16/0/1	<0.001
Random, 80 attr. (MHMG-80)	+0.134 [+0.092, +0.177]	15/0/2	<0.001
KNN-NRS, Euclidean	+0.117 [+0.073, +0.165]	16/0/1	<0.001

than MHMG. Tuning the radius over the same grid by cross-validation, which is the choice available in practice, left NRS +0.062 below MHMG (+0.037 to +0.087).

4.4 Controls for size and chance

Two controls test whether the advantage of MHMG comes from its criterion or only from the number of attributes it keeps (Table 12). When the greedy path of MHMG was cut to the size of the NRS reduct on the same fold, MHMG was still more accurate than NRS at each of the three radii, by +0.063 to +0.077. The gap to NRS is therefore not only a matter of size. MHMG was also far more accurate than random subsets of its own size (+0.196, better on 16 of the 17 datasets). On the 80 prefiltered attributes, MHMG-80 was more accurate than random subsets (+0.134). Both controls gave the same result with the SVM.

4.5 Comparison with standard selectors

We compare MHMG with five standard selectors, first at the size of MHMG and then with sizes tuned by cross-validation. MHMG was as accurate as the best standard

Table 13 3-NN accuracy (mean±std over 15 folds) of MHMG, the five same-size selectors, NRS*, FNRS-50 and all attributes. Bold: best selector on each dataset.

Dataset	MHMG	F	FCQ	ReliefF	L1	RFE	NRS*	FNRS	All attributes
Lymphoma	0.919±0.078	0.889±0.103	0.896±0.107	0.904±0.093	**0.933±0.082**	0.837±0.110	0.822±0.101	0.919±0.089	0.800±0.120
Leukemia	0.908±0.073	0.912±0.066	0.875±0.088	0.870±0.100	0.889±0.089	0.880±0.080	0.870±0.085	**0.948±0.082**	0.861±0.065
Burkitt	**0.871±0.039**	0.852±0.031	0.852±0.041	0.859±0.052	0.823±0.055	0.777±0.114	0.830±0.048	0.826±0.032	0.839±0.041
GCM	**0.565±0.064**	0.414±0.056	0.516±0.069	0.402±0.052	0.500±0.042	0.275±0.070	0.326±0.077	0.526±0.049	0.530±0.074
Tumors-C	0.578±0.107	0.500±0.161	0.583±0.104	0.650±0.170	0.611±0.132	0.611±0.129	0.622±0.099	**0.694±0.108**	0.600±0.090
RSCTC-1	0.492±0.101	0.533±0.110	0.559±0.096	0.562±0.101	**0.565±0.065**	0.543±0.112	0.489±0.091	0.559±0.129	0.524±0.097
RSCTC-3	**0.807±0.093**	0.561±0.086	0.695±0.132	0.723±0.078	0.796±0.119	0.793±0.078	0.691±0.082	0.695±0.102	0.747±0.085
AP-OmU	0.867±0.054	0.846±0.066	0.856±0.052	0.834±0.086	0.879±0.028	0.796±0.095	0.824±0.041	**0.904±0.049**	0.861±0.059
Colon	0.812±0.110	0.855±0.083	0.828±0.120	0.827±0.099	0.829±0.090	0.752±0.141	0.850±0.098	**0.873±0.070**	0.777±0.157
Duke	0.895±0.060	0.802±0.085	0.779±0.074	0.753±0.133	0.764±0.107	0.694±0.118	0.891±0.063	**0.953±0.086**	0.778±0.137
CLL-SUB	0.672±0.109	0.522±0.094	0.579±0.104	0.583±0.089	**0.678±0.120**	0.603±0.109	0.574±0.107	0.657±0.091	0.549±0.104
Glioma	0.660±0.091	0.640±0.150	0.660±0.130	**0.713±0.160**	0.660±0.135	0.560±0.091	0.600±0.120	0.647±0.151	0.780±0.121
GLI-85	0.851±0.094	0.851±0.077	0.847±0.073	0.871±0.078	0.827±0.061	0.773±0.080	0.820±0.085	**0.875±0.062**	0.867±0.065
Prostate	0.905±0.065	0.918±0.048	**0.925±0.048**	0.895±0.045	**0.925±0.052**	0.879±0.059	0.896±0.074	0.905±0.066	0.823±0.080
SMK-CAN	0.637±0.081	0.637±0.091	0.629±0.063	0.610±0.069	0.632±0.097	0.611±0.079	0.627±0.087	**0.669±0.063**	0.683±0.074
TOX-171	0.751±0.080	0.589±0.088	0.657±0.070	0.634±0.092	0.770±0.060	0.634±0.103	0.579±0.112	**0.811±0.091**	0.796±0.062
Lung	0.916±0.033	0.811±0.046	0.911±0.029	0.872±0.043	0.921±0.033	0.857±0.044	0.867±0.055	**0.931±0.025**	0.962±0.028
Mean	0.771	0.714	0.744	0.739	0.765	0.699	0.716	0.788	0.752

Table 14 MHMG against the five same-size selectors (3-NN, 17 datasets). p: uncorrected; p_{Holm}: Holm-corrected.

vs.	Median Δ	Mean Δ [95% CI]	W/T/L	p	p_{Holm}
F	+0.021	+0.057 [+0.021, +0.097]	11/2/4	0.018	0.072
FCQ	+0.011	+0.027 [+0.004, +0.051]	12/1/4	0.042	0.125
ReliefF	+0.027	+0.032 [+0.001, +0.064]	12/0/5	0.109	0.218
L1	−0.005	+0.006 [−0.013, +0.028]	7/1/9	0.925	0.925
RFE	+0.069	+0.072 [+0.039, +0.111]	15/0/2	0.001	0.004

Table 15 MHMG against selectors that tune their size within ten attributes (17 datasets). Size: median number of attributes; Φ: median Nogueira index; time: median seconds per fold, tuning included.

Method	Size	3-NN	SVM	Φ	Time	Δ [95% CI]	W/T/L	p
MHMG	7.4	0.771	0.775	0.13	6.98	–	–	–
L1-cv10	7.4	0.778	0.786	0.39	0.18	−0.007 [−0.028, +0.015]	5/0/12	0.706
FCQ-cv10	7.4	0.754	0.760	0.33	0.04	+0.017 [−0.006, +0.043]	8/0/9	0.706
ReliefF-cv10	6.5	0.740	0.748	0.58	0.31	+0.031 [−0.000, +0.063]	10/1/6	0.293
F-cv10	6.3	0.716	0.737	0.59	0.04	+0.055 [+0.015, +0.097]	12/0/5	0.081
RFE-cv10	6.8	0.698	0.721	0.33	3.80	+0.073 [+0.034, +0.116]	14/0/3	0.005

selector of the same size and close to the best tuned selectors. Table 13 gives the 3-NN accuracy on every dataset. At its own size, MHMG had the highest mean 3-NN accuracy of the six same-size selectors (0.771). L1 was second with 0.765. Only the difference against RFE was significant after Holm correction (Table 14).

Table 15 compares MHMG with the same five selectors when each chose its own size within ten attributes. MHMG has no tuning loop, yet for 4 of the five the confidence interval of the difference stayed above −0.02. Tuned L1 was the exception. It was slightly more accurate on average (difference -0.007, confidence interval -0.028 to +0.015), and it was faster and more stable. With an SVM, the full attribute set was more accurate than every subset of at most ten attributes (0.790 against 0.775 for MHMG).

4.6 Costs of the criterion

This subsection reports what the MHMG criterion costs in stability, robustness to label noise and running time. The main price of the criterion is stability. The median Nogueira index [35] of the MHMG reducts over the fifteen folds was 0.133, against 0.595 for F and 0.351 for L1 (Table 16). Within the family, NRS* was even less stable (0.075), while FNRS with its 50 attributes was more stable (0.281). Proposition 6 points to one mechanism. Each radius depends on one or a few nearest enemies, and a different fold can remove them. Under the MHMG reducts, a median of 0.433 of the training radii changed when the test fold was removed, about twice the value 0.202 that Corollary 1 gives for subsets chosen independently of the folds (Fig. 4). The excess comes from selecting the attributes on the same training fold, and it matches the low stability of the reducts in Table 16.

Table 16 Selection stability over the 15 training folds (Nogueira index Φ, median over datasets) and median selection time per fold.

Method	Φ	Time (s)
MHMG	0.133	6.98
F	0.595	0.00
FCQ	0.316	0.00
ReliefF	0.595	0.12
L1	0.351	0.04
RFE	0.191	1.26
NRS*	0.075	1.03
FNRS	0.281	14.00

Table 17 Mean 3-NN accuracy when a fraction ρ of the training labels is flipped, and drop from $\rho = 0$. p: Wilcoxon test of the drop at $\rho = 0.2$ against MHMG (uncorrected); last column: datasets on which MHMG is more accurate at $\rho = 0.2$.

Method	$\rho = 0$	0.1	0.2	$\Delta_{0.1}$	$\Delta_{0.2}$	p	MHMG better
MHMG	0.771	0.699	0.633	−0.072	−0.138	–	–
F	0.714	0.695	0.646	−0.018	−0.067	0.002	5/17
FCQ	0.744	0.705	0.639	−0.039	−0.105	0.020	6/17
ReliefF	0.739	0.721	0.661	−0.018	−0.078	0.003	8/17
L1	0.765	0.686	0.602	−0.079	−0.162	0.080	11/17
NRS*	0.716	0.625	0.566	−0.091	−0.151	0.431	16/17
FNRS	0.788	0.717	0.656	−0.071	−0.131	0.678	3/17
All attributes	0.752	0.724	0.673	−0.027	−0.079	0.001	4/17

Label noise also lowered the accuracy of the whole radius-based family. When 20 percent of the training labels were flipped, MHMG lost 0.138 in mean 3-NN accuracy, about twice as much as the F filter (0.067). NRS* and FNRS lost about as much as MHMG, and L1 lost the most (Table 17). A flipped sample becomes the nearest enemy of its former neighbours and shrinks their radii, which changes all three radius-based criteria directly. Finally, MHMG took a median of 7.0 seconds per fold. This is about 158 times the time of L1 and 6.2 times that of NRS, but within the range of the other neighbourhood models (Table 9). Running time rose with data size, correlating with both the number of attributes and the number of samples (Spearman $\rho = 0.72$ and 0.74, both $p < 0.01$), from 0.6 to 42.1 seconds (Colon and GCM).

5 Discussion

5.1 Saturation and the role of the radius

The NRS dependency fails when $m \gg n$ because it counts the radii above a fixed threshold, while distances grow with the number of attributes. With thousands of candidate attributes and a few dozen samples per class, a handful of attributes already places every sample farther than ε from its nearest enemy. From then on no attribute can change the NRS dependency (Proposition 3), and the search has nothing left to

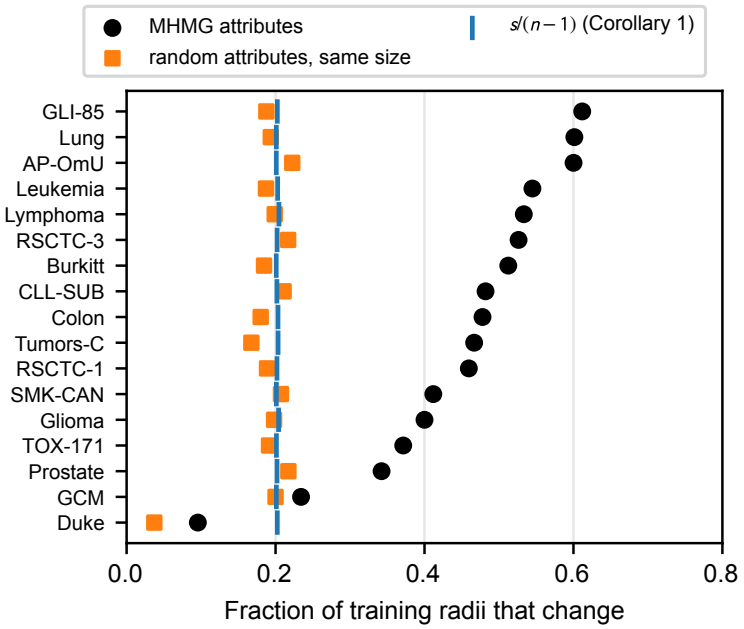


Fig. 4 Fraction of training samples whose safety radius changes when the test fold is removed (first repetition), for MHMG attributes, random attributes of the same number, and the value of Corollary 1

compare. A larger radius postpones this point, but then no single attribute raises the dependency above zero and the greedy search cannot start (Proposition 4, Table 11). Even the best of eight radii, chosen on the test folds, stayed slightly below MHMG on average. The models that adapt the neighbourhood stop early for a related reason. As Remark 2 explains, they count the samples whose neighbourhood is pure, and on these data the count reaches its full-set value after a few attributes. GBNRS avoids the early stop in the opposite way, by keeping almost every attribute.

MHMG uses the same safety radii, but it compares them with distances inside the class rather than with a constant. An added attribute changes both sides of this comparison. The energy can therefore still rise after every radius has passed any fixed value, as Example 1 shows on six samples. Theorem 1 links this rise to accuracy. Once the granule of a sample holds $k+1$ samples, the leave-one-out k-NN rule classifies that sample correctly on the training data, so a growing granule moves its sample towards this guarantee. The bound is loose: on the MHMG reducts its median for $k = 3$ was 0.535, against a training leave-one-out accuracy of 0.925. It explains the direction of the search rather than the accuracy reached. MHMG has its own stopping point, though, set by the rounding of γ, which we return to below.

FNRS lies between the two. Its dependency is a mean of truncated radii, and it keeps rising until the smallest radius reaches ε. On our data FNRS never reached its stopping target and ended at the cap of 50 attributes in every fold. In practice it needs

a second parameter, the cap or the target, to return a small subset, and at the size of MHMG it was less accurate.

5.2 Instability and label noise

Each safety radius rests on one or a few nearest enemies. Proposition 6 makes this precise, since a radius changes on held-out data exactly when all its nearest enemies are removed. With five folds, about a fifth of the radii change even for random attributes, and the attributes chosen by MHMG roughly doubled this fraction (Fig. 4). Once the radii move, a different attribute can win a close comparison in the greedy search, and the reduct changes with it. A mislabelled sample acts in a similar way. It becomes the nearest enemy of its former neighbours and shrinks their radii. Every criterion built on safety radii is exposed to the same mechanism. NRS* was even less stable than MHMG, and both NRS* and FNRS lost about as much accuracy under label noise. Aggregating reducts over resamples [52, 53] is the natural remedy, and we leave it to future work.

5.3 Recommendations and limitations

Within the neighbourhood rough-set family and with a nearest-neighbour classifier, our results favour MHMG over NRS with a fixed radius. MHMG avoids saturation and needs neither a radius nor a target size. The rounding precision of its stopping test, however, acts as a parameter. If a model with a fixed radius is required, FNRS is a better choice than NRS. With an SVM as the final classifier, FNRS with 50 attributes was more accurate than MHMG, and the full attribute set was more accurate than every subset of at most ten attributes. For a small attribute subset that need not be a rough-set reduct, tuned L1 matched MHMG in accuracy at a fraction of its running time. The F filter and ReliefF gave the most reproducible subsets and were the least affected by label noise.

All seventeen datasets are gene-expression data, so other data with $m \gg n$, such as spectra or text, remain untested. NDER, neighbourhood entropy and FSSNC also use a global radius, but they depend on more than the safety radii, so Propositions 3 and 4 do not cover them, and we did not test them. Proposition 4 also assumes the Euclidean distance on attributes scaled to $[0, 1]$.

Each competitor was run with fixed settings taken from its original paper or reference code (Section 4.1), and other settings might change their results. Adaptive NRS and GBNRS were compared only on the 80 attributes of largest variance, where MHMG itself reached 0.709 against 0.771 on all attributes. Our main evaluator, 3-NN, is related to what MHMG optimises (Theorem 1), so the 3-NN comparisons favour MHMG. With an SVM, MHMG again ranked second in the family after FNRS with 50 attributes.

The running time of MHMG, NRS and FNRS grows as $O((q+1)mn^2)$ (Section 3.6). The tables here have at most a few hundred samples, and the mean time per fold of MHMG did not exceed 42.1 seconds on any of them. It is quadratic in the number of samples, so much larger sample sizes would need a faster implementation, which we did not develop.

MHMG has one hidden parameter, the rounding of γ in its stopping test. With two decimals, the published setting, the search stops when γ no longer increases. Rounding to three decimals roughly doubled the median size, from 7.4 to 15.3 attributes. It also raised the mean 3-NN accuracy by +0.022 (confidence interval +0.008 to +0.036), and it was better on 13 of the 17 datasets. We kept the published setting and did not tune it.

6 Conclusion

This paper analysed why neighbourhood rough-set attribute reduction with a fixed radius fails when the attributes far outnumber the samples. The NRS dependency sees the data only through the safety radii. It saturates once every radius exceeds the threshold, and with a radius large enough to avoid this, every small subset has an empty positive region, so the greedy search cannot start. On seventeen gene-expression datasets, even random attribute subsets saturated NRS after a few attributes.

The published MHMG criterion, which has no distance radius, compares distances only with each other, so no fixed threshold can saturate it, and its energy bounds the training leave-one-out nearest-neighbour accuracy. With 3-NN, which favours it, MHMG was more accurate than NRS at every radius at which NRS returned a reduct, also at equal size. Against the best of eight radii chosen on the test folds, its margin was small. Granular-ball NRS and FNRS avoided early stopping only by keeping many more attributes. With an SVM, FNRS with 50 attributes was more accurate than MHMG, and tuned L1 was as accurate as MHMG, more stable and faster.

The criterion keeps the weaknesses of nearest enemies. Its reducts are unstable and sensitive to label noise, as those of NRS and FNRS are, and the rounding in its stopping rule acts as a hidden parameter. Future work should aggregate reducts over resamples to stabilise them [52, 53]. A stopping rule based on the training k-NN accuracy that the energy bounds could also replace the rounding of γ.

Author contributions
Thanh Dai Tran designed the study with Thi Thuy Tram Nguyen, wrote the code, ran the experiments, analysed the results and wrote the manuscript. Thi Thuy Tram Nguyen contributed to the design of the study and to the review of the manuscript. Long Giang Nguyen and Manh Hung Nguyen supervised the work and reviewed and edited the manuscript. All authors read and approved the final manuscript.

Funding
This research received no external funding.

Data and code availability
All seventeen datasets are public (OpenML and the scikit-feature repository). The repository at https://github.com/ttdaiuneti/mhmg-gene-selection contains the fold-level results (one row per dataset, repetition, fold and method), the selected attributes of every run, the implementations of MHMG, NRS and FNRS, the re-implementations of the other neighbourhood models, the run scripts with all settings, and the scripts that generate every table, figure and number in this article.

Declarations

Ethics approval and consent to participate
Not applicable. The study uses public, de-identified datasets.

Consent to publish
Not applicable.

Clinical trial number
Not applicable.

Competing interests
The authors declare no conflict of interest.

References

[1] Saeys, Y., Inza, I., Larrañaga, P.: A review of feature selection techniques in bioinformatics. Bioinformatics **23**(19), 2507–2517 (2007) https://doi.org/10.1093/bioinformatics/btm344

[2] Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A., Benítez, J.M., Herrera, F.: A review of microarray datasets and applied feature selection methods. Information Sciences **282**, 111–135 (2014) https://doi.org/10.1016/j.ins.2014.05.042

[3] Pawlak, Z.: Rough sets. International Journal of Computer & Information Sciences **11**(5), 341–356 (1982) https://doi.org/10.1007/BF01001956

[4] Hu, Q., Yu, D., Liu, J., Wu, C.: Neighborhood rough set based heterogeneous feature subset selection. Information Sciences **178**(18), 3577–3594 (2008) https://doi.org/10.1016/j.ins.2008.05.024

[5] Nguyen, T.T.T., Nguyen, L.G., Nguyen, M.H., Tran, T.D.: Attribute reduction via knowledge energy and maximal homogeneous metric granules in neighborhood rough sets. International Journal of Approximate Reasoning **198**, 109787 (2026) https://doi.org/10.1016/j.ijar.2026.109787

[6] Wang, C., Shao, M., He, Q., Qian, Y., Qi, Y.: Feature subset selection based on fuzzy neighborhood rough sets. Knowledge-Based Systems **111**, 173–179 (2016) https://doi.org/10.1016/j.knosys.2016.08.009

[7] Xia, S., Zhang, H., Li, W., Wang, G., Giem, E., Chen, Z.: GBNRS: a novel rough set algorithm for fast adaptive attribute reduction in classification. IEEE Transactions on Knowledge and Data Engineering **34**(3), 1231–1242 (2022) https://doi.org/10.1109/tkde.2020.2997039

[8] Sun, L., Liang, H., Ding, W., Xu, J.: Granular ball fuzzy neighborhood rough sets-based feature selection via multiobjective mayfly optimization. IEEE Transactions on Fuzzy Systems **32**(11), 6112–6124 (2024) https://doi.org/10.1109/TFUZZ.2024.3440575

[9] Sun, L., Du, W., Ding, W., Long, Q., Xu, J.: Granular ball-based fuzzy multi-neighborhood rough set for feature selection via label enhancement. Engineering Applications of Artificial Intelligence **145**, 110191 (2025) https://doi.org/10.1016/j.engappai.2025.110191

[10] Wang, C., Shi, Y., Fan, X., Shao, M.: Attribute reduction based on k-nearest neighborhood rough sets. International Journal of Approximate Reasoning **106**, 18–31 (2019) https://doi.org/10.1016/j.ijar.2018.12.013

[11] Wang, N., Zhao, E.: A new method for feature selection based on weighted k-nearest neighborhood rough set. Expert Systems with Applications **238**, 122324 (2024) https://doi.org/10.1016/j.eswa.2023.122324

[12] Zhang, X., Wang, J.: A noise-robust feature selection using KNN and weighted fuzzy rough sets for imbalanced multi-scale data. Applied Soft Computing **194**, 114964 (2026) https://doi.org/10.1016/j.asoc.2026.114964

[13] Qiu, W., Liu, C., Lin, B., Chen, X., Miao, D.: Attribute reduction based on adaptive neighborhood rough sets and three-way pied kingfisher optimizer. Expert Systems with Applications **271**, 126618 (2025) https://doi.org/10.1016/j.eswa.2025.126618

[14] Ju, H., Ding, W., Shi, Z., Huang, J., Yang, J., Yang, X.: Attribute reduction with personalized information granularity of nearest mutual neighbors. Information Sciences **613**, 114–138 (2022) https://doi.org/10.1016/j.ins.2022.09.006

[15] Yang, D., Wu, J., Yu, Y., Huang, J., Lan, J.: Accelerated attribute reduction based on weighted discriminative heterogeneous neighborhood relationships for neighborhood rough sets. Neurocomputing **701**, 134648 (2026) https://doi.org/10.1016/j.neucom.2026.134648

[16] Thuy, N.N., Wongthanavasu, S.: A novel feature selection method for high-dimensional mixed decision tables. IEEE Transactions on Neural Networks and Learning Systems **33**(7), 3024–3037 (2022) https://doi.org/10.1109/TNNLS.2020.3048080

[17] Thuy, N.N., Wongthanavasu, S.: Attribute reduction with fuzzy divergence-based weighted neighborhood rough sets. International Journal of Approximate Reasoning **173**, 109256 (2024) https://doi.org/10.1016/j.ijar.2024.109256

[18] Dai, J., Zhu, Z., Li, M., Zou, X., Zhang, C.: Attribute reduction for heterogeneous data based on monotonic relative neighborhood granularity. International

Journal of Approximate Reasoning **170**, 109210 (2024) https://doi.org/10.1016/j.ijar.2024.109210

[19] Hu, M., Guo, Y., Chen, D., Tsang, E.C.C., Zhang, Q.: Attribute reduction based on neighborhood constrained fuzzy rough sets. Knowledge-Based Systems **274**, 110632 (2023) https://doi.org/10.1016/j.knosys.2023.110632

[20] Wang, Y., Sun, M., Long, L., Liu, J., Ren, Y.: Feature gene selection based on fuzzy neighborhood joint entropy. Complex & Intelligent Systems **10**(1), 129–144 (2024) https://doi.org/10.1007/s40747-023-01138-9

[21] Liu, K., Li, T., Yang, X., Ju, H., Yang, X., Liu, D.: Hierarchical neighborhood entropy based multi-granularity attribute reduction with application to gene prioritization. International Journal of Approximate Reasoning **148**, 57–67 (2022) https://doi.org/10.1016/j.ijar.2022.05.011

[22] Hu, Q., Pedrycz, W., Yu, D., Lang, J.: Selecting discrete and continuous features based on neighborhood decision error minimization. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **40**(1), 137–150 (2010) https://doi.org/10.1109/TSMCB.2009.2024166

[23] Brighton, H., Mellish, C.: Advances in Instance Selection for Instance-Based Learning Algorithms. Data Mining and Knowledge Discovery **6**(2), 153–172 (2002) https://doi.org/10.1023/A:1014043630878

[24] Leyva, E., González, A., Pérez, R.: Three new instance selection methods based on local sets: A comparative study with several approaches from a bi-objective perspective. Pattern Recognition **48**(4), 1523–1537 (2015) https://doi.org/10.1016/j.patcog.2014.10.001

[25] Gilad-Bachrach, R., Navot, A., Tishby, N.: Margin based feature selection - theory and algorithms. In: Twenty-first International Conference on Machine Learning - ICML '04, p. 43 (2004). https://doi.org/10.1145/1015330.1015352

[26] Kononenko, I.: Estimating attributes: Analysis and extensions of RELIEF. In: Lecture Notes in Computer Science, pp. 171–182 (1994). https://doi.org/10.1007/3-540-57868-4_57

[27] Robnik-Šikonja, M., Kononenko, I.: Theoretical and Empirical Analysis of ReliefF and RReliefF. Machine Learning **53**(1-2), 23–69 (2003) https://doi.org/10.1023/A:1025667309714

[28] Ho, T.K., Basu, M.: Complexity measures of supervised classification problems. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(3), 289–300 (2002) https://doi.org/10.1109/34.990132

[29] Lorena, A.C., Garcia, L.P.F., Lehmann, J., Souto, M.C.P., Ho, T.K.: How complex

is your classification problem? A survey on measuring classification complexity. ACM Computing Surveys **52**(5), 1–34 (2019) https://doi.org/10.1145/3347711

[30] Ding, C., Peng, H.: Minimum redundancy feature selection from microarray gene expression data. Journal of Bioinformatics and Computational Biology **03**(02), 185–205 (2005) https://doi.org/10.1142/S0219720005001004

[31] Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(8), 1226–1238 (2005) https://doi.org/10.1109/TPAMI.2005.159

[32] Tibshirani, R.: Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology **58**(1), 267–288 (1996) https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[33] Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning **46**(1-3), 389–422 (2002) https://doi.org/10.1023/A:1012487302797

[34] Kalousis, A., Prados, J., Hilario, M.: Stability of feature selection algorithms: a study on high-dimensional spaces. Knowledge and Information Systems **12**(1), 95–116 (2007) https://doi.org/10.1007/s10115-006-0040-8

[35] Nogueira, S., Sechidis, K., Brown, G.: On the Stability of Feature Selection Algorithms. Journal of Machine Learning Research **18**(174), 1–54 (2018)

[36] Vanschoren, J., Rijn, J.N., Bischl, B., Torgo, L.: OpenML. ACM SIGKDD Explorations Newsletter **15**(2), 49–60 (2014) https://doi.org/10.1145/2641190.2641198

[37] Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S.: Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. Science **286**(5439), 531–537 (1999) https://doi.org/10.1126/science.286.5439.531

[38] Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mack, D., Levine, A.J.: Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proceedings of the National Academy of Sciences **96**(12), 6745–6750 (1999) https://doi.org/10.1073/pnas.96.12.6745

[39] Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L., Marti, G.E., Moore, T., Hudson, J., Lu, L., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Levy,

R., Wilson, W., Grever, M.R., Byrd, J.C., Botstein, D., Brown, P.O., Staudt, L.M.: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature **403**(6769), 503–511 (2000) https://doi.org/10.1038/35000501

[40] Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C.-H., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J.P., Poggio, T., Gerald, W., Loda, M., Lander, E.S., Golub, T.R.: Multiclass cancer diagnosis using tumor gene expression signatures. Proceedings of the National Academy of Sciences **98**(26), 15149–15154 (2001) https://doi.org/10.1073/pnas.211566398

[41] Pomeroy, S.L., Tamayo, P., Gaasenbeek, M., Sturla, L.M., Angelo, M., McLaughlin, M.E., Kim, J.Y.H., Goumnerova, L.C., Black, P.M., Lau, C., Allen, J.C., Zagzag, D., Olson, J.M., Curran, T., Wetmore, C., Biegel, J.A., Poggio, T., Mukherjee, S., Rifkin, R., Califano, A., Stolovitzky, G., Louis, D.N., Mesirov, J.P., Lander, E.S., Golub, T.R.: Prediction of central nervous system embryonal tumour outcome based on gene expression. Nature **415**(6870), 436–442 (2002) https://doi.org/10.1038/415436a

[42] West, M., Blanchette, C., Dressman, H., Huang, E., Ishida, S., Spang, R., Zuzan, H., Olson, J.A., Marks, J.R., Nevins, J.R.: Predicting the clinical status of human breast cancer by using gene expression profiles. Proceedings of the National Academy of Sciences **98**(20), 11462–11467 (2001) https://doi.org/10.1073/pnas.201162998

[43] Wojnarski, M., Janusz, A., Nguyen, H.S., Bazan, J., Luo, C., Chen, Z., Hu, F., Wang, G., Guan, L., Luo, H., Gao, J., Shen, Y., Nikulin, V., Huang, T.-H., McLachlan, G.J., Bošnjak, M., Gamberger, D.: RSCTC'2010 discovery challenge: mining DNA microarray data for medical diagnosis and treatment. In: Lecture Notes in Computer Science, pp. 4–19 (2010). https://doi.org/10.1007/978-3-642-13529-3_3

[44] Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R.P., Tang, J., Liu, H.: Feature Selection. ACM Computing Surveys **50**(6), 1–45 (2018) https://doi.org/10.1145/3136625

[45] Singh, D., Febbo, P.G., Ross, K., Jackson, D.G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A.A., D'Amico, A.V., Richie, J.P., Lander, E.S., Loda, M., Kantoff, P.W., Golub, T.R., Sellers, W.R.: Gene expression correlates of clinical prostate cancer behavior. Cancer Cell **1**(2), 203–209 (2002) https://doi.org/10.1016/S1535-6108(02)00030-2

[46] Bhattacharjee, A., Richards, W.G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., Loda, M., Weber, G., Mark, E.J., Lander, E.S., Wong, W., Johnson, B.E., Golub, T.R., Sugarbaker, D.J., Meyerson, M.: Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proceedings of the National Academy of Sciences **98**(24), 13790–13795 (2001) https://doi.org/10.1073/pnas.

191502998

[47] Ambroise, C., McLachlan, G.J.: Selection bias in gene extraction on the basis of microarray gene-expression data. Proceedings of the National Academy of Sciences **99**(10), 6562–6566 (2002) https://doi.org/10.1073/pnas.102102699

[48] Demšar, J.: Statistical Comparisons of Classifiers over Multiple Data Sets. Journal of Machine Learning Research **7**, 1–30 (2006)

[49] Friedman, M.: The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. Journal of the American Statistical Association **32**(200), 675–701 (1937) https://doi.org/10.1080/01621459.1937.10503522

[50] Wilcoxon, F.: Individual Comparisons by Ranking Methods. Biometrics Bulletin **1**(6), 80 (1945) https://doi.org/10.2307/3001968

[51] Holm, S.: A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics **6**(2), 65–70 (1979)

[52] Meinshausen, N., Bühlmann, P.: Stability Selection. Journal of the Royal Statistical Society Series B: Statistical Methodology **72**(4), 417–473 (2010) https://doi.org/10.1111/j.1467-9868.2010.00740.x

[53] Wang, A., Liu, H., Yang, J., Chen, G.: Ensemble feature selection for stable biomarker identification and cancer classification from microarray expression data. Computers in Biology and Medicine **142**, 105208 (2022) https://doi.org/10.1016/j.compbiomed.2021.105208